Self–Observer Rating Discrepancies of Managers in Asia: A study of derailment characteristics and behaviors in Southern and Confucian Asia

William A. Gentry*, Jeffrey Yip*, and Kelly M. Hannum***

*Research, Innovation, & Product Development, Center for Creative Leadership, Greensboro, NC, USA.
gentryb@ccl.org
**Jeffrey Yip is now at the School of Management, Boston University

Antecedents to self–observer rating discrepancies in multisource instruments have been established at the individual and organizational level. However, research examining cultural antecedents is limited, which is particularly relevant as multisource instruments gain popularity around the world. We investigated multisource ratings of 860 Asian managers from the regions of Southern Asia ($n = 261$) and Confucian Asia ($n = 599$) and analyzed cultural differences in self–observer rating discrepancies. Multivariate regression procedures revealed that the self–observer rating discrepancy was wider for managers from Southern Asia as compared with Confucian Asia. The reason for the discrepancy was driven by managers' self-ratings being different across cultures than by observer ratings from managers' bosses, direct reports, or peers; the predictor is related to self-ratings not observer ratings, producing differential self–observer ratings due to self-ratings. We discuss cultural differences in self- and observer ratings within Asia and provide implications for the practice of multisource assessments.

1. Introduction

Multisource (i.e., multirater, 360°) instruments provide a wealth of information and data about managers and organizations by gathering ratings from the self and multiple 'observer' perspectives (e.g., peer, subordinate, supervisor, client, customer). Many Fortune 1,000 and 500 companies administer multisource instruments for employee development, feedback, and other purposes (Atwater & Waldman, 1998; Conway, Lombardo, & Sanders, 2001). These data can be useful indicators of managerial and organizational performance (e.g., Atwater, Ostroff, Yammarino, & Fleenor, 1998; Atwater, Roush, & Fischthal, 1995; Church, 1997, 2000; Conway *et al*., 2001; Dalessio, 1998; London & Smither, 1995; Morgeson, Mumford, & Campion, 2005; Smither & Walker, 2001; Tornow, 1993; Yammarino, 2003). However, a manager's self-rating on a construct is frequently discrepant (i.e., different, dissimilar, incongruous, or in disagreement) from observers' ratings of that manager on that same construct (Brutus, Fleenor, & McCauley, 1999; Gentry, Hannum, Ekelund, & de Jong, 2007; Morgeson *et al*., 2005; Mount, Judge, Scullen, Sytsma, & Hezlett, 1998; Sala, 2003). Reasons why such self–observer rating discrepancies exist is a major topic of multisource research.

This study extends previous research on self–observer rating discrepancies (e.g., Atwater *et al*., 1998, 1995; Church, 1997; Ostroff, Atwater, & Feinberg, 2004; Sala, 2003) by first, specifically examining ratings of the characteristics and behaviors associated with derailment (being fired, demoted, or plateauing in a job). Given the costly effects of derailment to managers, their coworkers, and their organization (Finkelstein, 2004; Finkin, 1991; Gillespie, Walsh, Winefield, Dua, & Stough, 2001; Lombardo & McCauley, 1988; Smart, 1999; Wells, 2005), studying self–observer rating discrepancies in derailment may help inform and prevent managerial derailment and its negative impact on organizational outcomes, as self–observer rating discrepancies may be a key indicator of leadership and managerial outcomes (Atwater &

Yammarino, 1992; Atwater *et al*., 1998, 1995; Church, 1997; Johnson & Ferstl, 1999).

Secondly, we attempt to extend previous research concerning derailment and self–observer rating discrepancies by focusing on managers from Asian countries. The antecedents to self–observer rating discrepancies have been established at the individual level (Ostroff *et al*., 2004) and the organizational level (Gentry *et al*., 2007; Sala, 2003). However, research examining cultural-level antecedents, though growing, is still rather limited. As multisource instruments gain popularity around the world (Atwater, Brett, & Charles, 2007; Atwater, Waldman, Ostroff, Robie, & Johnson, 2005; Brutus, Leslie, & McDonald-Mann, 2001), the resulting need to conduct multisource research across and within different countries intensifies. The need to conduct this type of research specifically on Asian managers is increasingly important as Asia gains prominence in international business (Mathews, 2006) and studying an Asian population brings great opportunity to contribute to existing knowledge around the world (Lau, 2002; Meyer, 2006).

This article will first detail why studying managerial derailment is a worthwhile endeavor to pursue and provide background of the construct. Next, the article will provide a theoretical background to the existence of self–observer rating discrepancies, and will offer theoretical reasoning behind the study of Confucian Asia and Southern Asia managers, and why these two groups may differ in self–observer rating discrepancies. We then provide the method and results to our study of 860 Asian managers and their multisource ratings of derailment, and finish with a discussion and implications for those working with assessments.

1.1. Managerial derailment

Lombardo and McCauley (1988) described that derailment 'occurs when a manager who was expected to go higher in the organization and who was judged to have the ability to do so is fired, demoted, or plateaued below expected levels of achievement' (p. 1). Managers derail as a result of an apparent lack of fit between personal characteristics or skills, and job demands (Leslie & Van Velsor, 1996). Managers derail because their preference to do things their own way eventually led to ineffectiveness in their job (Kovach, 1986). Similarly, managers derail because they could not adjust their behavior to different changes and demands in their managerial role (Lombardo & Eichinger, 1989/2005).

McCall and Lombardo (1983) were among the first to study managerial derailment by interviewing both successful executives and executives dealing with derailment. Those interviews and subsequent research (e.g., Leslie & Van Velsor, 1996; Lombardo & McCauley, 1988; Lombardo, McCauley, McDonald-Mann, & Leslie, 1999; Morrison, White, & Van Velsor, 1987) revealed characteristics and behaviors of previously derailed managers, classified under five clusters: (a) 'Problems with Interpersonal Relationships' – organizational isolation and describing a manager as authoritarian, cold, aloof, arrogant, and insensitive; (b) 'Difficulty Leading a Team' – a manager failing to staff effectively, failing to build and lead a team, and the inability to handle conflict; (c) 'Difficulty Changing or Adapting' – a manager's inability to adapt to a boss with a different managerial or interpersonal style, and the inability of a manager to grow, learn, develop, and think strategically; (d) 'Failure to Meet Business Objectives' – describing a manager as overly ambitious, lacking follow-through, or as a poor performer; (e) 'Too Narrow Functional Orientation' – a manager's ill-preparedness for promotion, and inability to manage outside of his or her current function.

The consequences of managerial derailment for the derailed manager include loss in money or pay, reputation, or sense of well being. The effects of derailment are not isolated to just the derailed manager; coworkers and the organization can also experience financial and emotional repercussions (Finkelstein, 2004; Lombardo & McCauley, 1988; Smart, 1999; Wells, 2005). As a result, studying whether or not managers display the characteristics and behaviors associated with derailment is crucial and represents an important aspect of managerial and multisource assessment research. Previously, managerial derailment research has covered how derailment takes place, the precursors to derailment (Hogan & Hogan, 2001; McCall & Lombardo, 1983), and derailment's negative outcomes (Finkelstein, 2004; Smart, 1999; Wells, 2005). The present research will expand on previous research by examining the antecedents of the self–observer rating discrepancy of the characteristics and behaviors of derailment among managers in Asia.

1.2. Self–observer rating discrepancy: theoretical background and hypothesis development

Atwater and Yammarino's (1997) model of self–observer rating discrepancy provides the theoretical framework of the present study. Multisource instruments may identify whether or not managers display the characteristics and behaviors of derailment. But frequently, self-ratings do not always agree with observer ratings (Harris & Schaubroeck, 1988; Tsui & Ohlott, 1988). Atwater and Yammarino (1997) argued that self-ratings and observer ratings provide different perspectives on the same phenomenon. By asking observers (e.g., boss, direct reports, peers) about a person, and comparing their aggregated responses to that of self-ratings, agreement (or disagreement) between self- and observer ratings becomes valuable both as information in and of itself, and as determinants of performance outcomes and success. For instance, congruence (i.e., agreement) between self- and observer ratings is associated with managerial and organizational outcomes (Atwater & Yammarino, 1992;

Atwater *et al*., 1998, 1995; Church, 1997; Johnson & Ferstl, 1999; Yammarino & Atwater, 1997). As managers whose ratings are in agreement tend to be better performers, managers whose ratings are in disagreement (i.e., have a discrepancy between self- and observer ratings) may well be on a path toward ineffectiveness. Self–observer discrepancies in rating the characteristics and behaviors of derailment can therefore become a crucial signal for managers to react, change, and develop before actual derailment occurs, thereby averting the costly effects of actual derailment previously mentioned.

Many wonder why such a discrepancy would occur. As a result, many have weighed in with their opinions. One reason could be that a 'disconnect' between what a manager thinks of himself or herself, and what observers think of the manager may exist (Goleman, Boyatzis, & McKee, 2001). Also, managers may be arrogant and insensitive, unwilling to perform self-assessments, may be unwilling to accept input and truthful feedback from others, or may have reluctant coworkers who do not offer straightforward assessments of managers (Conger & Nadler, 2004; Dotlich & Cairo, 2003; Kaplan, Drath, & Kofodimos, 1987; Kovach, 1986; Kramer, 2003; Levinson, 1994). Furthermore, self–observer discrepancies may exist because self-ratings may be 'more favorable' than observer ratings due to leniency bias (Church, 1997; Podsakoff & Organ, 1986; Van Velsor, Taylor, & Leslie, 1993).

Early multisource research examining self–observer rating discrepancies predominantly examined American managers, or did not specifically intend to study managers outside the United States (Atwater & Yammarino, 1992; Atwater *et al*., 1998, 1995; Brutus *et al*., 1999; Ostroff *et al*., 2004; Sala, 2003). Given the popularity of multisource instruments around the world, recent multisource research has specifically focused on samples outside the United States, such as in Europe (e.g., Atwater *et al*., 2005; Gentry *et al*., 2007) and Latin America (Varela & Premeaux, 2008). One may wonder whether similar findings of self–observer rating discrepancies exist within Asia. Such a research endeavor is relevant as Asia gains prominence in international business (Mathews, 2006) and some would consider conducting research on Asian populations to greatly contribute to and extend research knowledge (Lau, 2002; Meyer, 2006). Given the preponderance of evidence just discussed from other cultures that self-ratings are different from observer ratings, the following hypotheses are given concerning the Asian context of the present study:

Hypothesis 1 (a–c): Rating discrepancies on derailment characteristics and behaviors will exist for Asian managers when examining (a) self–boss ratings; (b) self–direct report ratings; and (c) self–peer ratings.

Knowing self–observer rating discrepancies exist is one issue. However, knowing why such discrepancies exist is also imperative to study (Brutus *et al*., 1999; Ostroff *et al*., 2004). Previous multisource research has shown that certain individual-level characteristics such as gender, age, and education (Ostroff *et al*., 2004) and personality (Brutus *et al*., 1999) are antecedents to self–observer rating discrepancies. Examining another level of analysis, organizational (i.e., managerial) level is also an antecedent, such that managers at higher organizational levels have bigger/wider self–observer rating discrepancies than managers at lower levels (Gentry *et al*., 2007; Sala, 2003). As multisource instruments gain prominence worldwide, investigating cultural-level antecedents to self–observer rating discrepancies seems a reasonable and logical next step in research, particularly, as London (2003) notes, perceptions of self and observers 'may be influenced by cultural factors and [such] factors should be examined' (p. 212). In an exploratory analysis, Gentry *et al*. (2007) showed that US managers have a bigger/wider self–observer rating discrepancy than European managers, but gave no theoretical reasoning behind this finding. More recent, Varela and Premeaux (2008) found that cultural values played a part in self- and observer ratings in the highly collectivistic and power distance cultures of Venezuela and Colombia.

The managers in the present study are all practicing Asian managers. It is necessary to note however, that Asia is a geographical, not cultural entity (Nandy, 1998). Although the concept of 'Asia' is commonly recognized as a geopolitical cluster of countries, its internal cultural dynamics are hugely diverse. Variations in economic development, political systems, cultural values and practices are important in considering the differences in Asia. That said, despite cultural differences across Asia, researchers have noted commonalities among distinctive clusters of countries within Asia (Cattell, 1950; Hofstede, 1980; House, Hanges, Javidan, Dorfman, & Gupta, 2004; Ronen & Shenkar, 1985). Data collected by the GLOBE study (House *et al*., 2004) and Hofstede (1980) demonstrate a number of cultural commonalities among clusters within Asia. Both studies have made arguments for grouping countries into clusters based on similar cultural dimensions, shared histories, languages, and ideologies.

Within Asia, House *et al*. (2004) identified two distinct cultural clusters – the Confucian Asia cluster (China, Hong Kong, Japan, Singapore, South Korea, and Taiwan) and the Southern Asia cluster (India, Indonesia, Iran, Malaysia, Philippines, and Thailand). The clustering of Confucian and Southern Asia countries has been validated through discriminant analysis conducted by House at al. (2004) and elaborated on by Gupta, Hanges, and Dorfman (2002) and Chhokar, Brodbeck, and House (2007). For consistency, we also use these clusters in the present study. According to House *et al*. (2004), the teachings and works of Confucius have a distinct historical influence on the Confucian cluster, whereas those in

the Southern Asian cluster share Brahminism and Buddhist, Islamic, and Christian influences.

Among the cultural differences between Confucian and Southern Asia identified by House *et al.* (2004), two organizational indicators in particular are worth noting – that of organizational power distance and institutional collectivism. The findings by House *et al.* (2004) point to higher organizational power distance practices in the Southern Asia cluster than the Confucian cluster. In addition, the Confucian cluster is higher on institutional collectivism practices than the Southern Asia cluster. This is reinforced by higher societal collectivism scores for the Confucian Asian cluster (House *et al.*, 2004). Taken together, such differences reflect a core differentiation in terms of cultural practices. The higher power distance scores are reflective of a more status conscious culture in Southern Asia as evident in the hierarchical organization of societies within the cluster (Gupta, Surie, Javidan, & Chhokar, 2002). The higher institutional and societal collectivism scores are reflective of an emphasis on relationships and obligations in the Confucian Asia cluster, as evident in the Confucian concepts of *guanxi* (relationships) and *renqing* (obligation; Chen, 2004). Markus and Kitayama (1991) posited that because cultures differ from each other in fundamental ways, so too will the self-concepts of people from those cultures.

Based on the cultural differences identified between the Confucian and Southern Asia cluster of countries, we believe this has implications on the self- and observer ratings between the two clusters. We hypothesize that the discrepancy between self- and observer ratings of the extent to which managers display the behaviors and characteristics of derailment will be 'bigger/wider' (i.e., more discrepant or different) in the Southern Asia cluster than the Confucian Asia cluster of countries. With higher organizational power distance among the Southern Asia cluster of countries, wider discrepancies may exist between self- and observer ratings as compared with the Confucian Asia cluster. In addition, with the higher institutional collectivism in the Confucian Asia cluster, more modest ratings and narrower discrepancies may be found. While previous research was more exploratory, and did not give theoretical reasoning why some cultures have bigger/wider self–observer rating discrepancies than others (e.g., Gentry *et al.*, 2007), we provided a theoretical basis for the following hypothesis:

Hypothesis 2(a–c): Rating discrepancies on derailment characteristics and behaviors will be 'bigger/wider' in Southern Asia than in Confucian Asia when examining (a) self–boss ratings; (b) self–direct report ratings; and (c) self–peer ratings.

Because the outcomes of derailment may be detrimental to a manager, coworkers, and the organization (Finkelstein, 2004; Lombardo & McCauley, 1988; Smart, 1999; Wells, 2005), the present research is important. Knowledge that a discrepancy exists on the characteristics and behaviors of derailment, and the antecedents of such discrepancies, can extend previous derailment research. Furthermore, given that multisource instruments are used around the world (Atwater *et al.*, 2007, 2005; Brutus *et al.*, 2001) more research is needed to examine cultural level antecedents to self–observer rating discrepancies, in particular, research with a theoretical foundation such as the present research. By focusing on Asian managers, this research attends to cultural-level antecedents to self–observer rating discrepancies and also answers the call of others (e.g., Lau, 2002; Mathews, 2006; Meyer, 2006) to examine Asian managers to contribute to and extend research and knowledge.

2. Method

2.1. Participants and procedures

Data came from a larger archival database of managers with multisource ratings between 2001 and 2006. Data for the analysis of this study only included managers from Asian countries that contained at least 40 managers in the database who were native to and currently working in the same Asian country. As a result, we used complete data from 860 Asian managers from over 140 different companies (63% from the private sector) covering 16 different industries (e.g., health, manufacturing, transportation, finance, education, etc.) from the countries of China, Hong Kong, India, Indonesia, Japan, Singapore, South Korea, and Thailand. The managers' mean age was approximately 41 years (range = 22–85), 71% of the managers were male, and 95% had at a minimum a bachelors degrees. Finally, 20.6% were 'Middle-level' managers, 43.6% were 'Upper-Middle-level' managers, and 35.8% were from the 'High-level' managers from the 'Top' or 'Executive' ranks.

Consistent with previous research such as the GLOBE study (House *et al.*, 2004), we grouped the data into two Asian clusters (origin): Confucian Asia and Southern Asia. Confucian Asia ($n = 599$) consisted of the countries China, Hong Kong, Japan, Singapore, and South Korea. Southern Asia consisted of India, Indonesia, and Thailand ($n = 261$). We dummy-coded the origin variable, such that Confucian Asia = 0 and Southern Asia = 1. There were no statistically significant differences in age, organizational level, job tenure, and organization tenure of the managers in the two Asian culture clusters.

2.2. Measures

A multisource developmental feedback instrument called BENCHMARKS® (Lombardo & McCauley, 1994; Lombardo *et al.*, 1999; McCauley & Lombardo, 1990) gathered

ratings from the self, direct report, peer, and boss perspectives. The multisource instrument is based on research about how successful managers learn, grow, and change (Lindsey, Homes, & McCall, 1987; Morrison *et al*., 1987), and about how they derail (Leslie & Van Velsor, 1996; Lombardo & McCauley, 1988). In addition, it is a well-validated and reviewed multisource feedback instrument (Carty, 2003; Center for Creative Leadership, 2004; Douglas, 2003; Lombardo & McCauley, 1994; McCauley, Lombardo, & Usher, 1989; Spangler, 2003; Zedeck, 1995). Furthermore, the multisource instrument has also been used with different data in prior research including self–observer agreement and correlates or predictors of agreement and effectiveness (e.g., Atwater *et al*., 1998; Brutus, Fleenor, & London, 1998; Brutus *et al*., 1999; Conway, 2000; Fleenor, McCauley, & Brutus, 1996; Gentry *et al*., 2007). Consistent with past research (e.g., Gentry, Katz, & McFeeters, 2009; Graves, Ohlott, & Ruderman, 2007; Lyness & Judiesch, 2008), scores from 40 items used to measure the five weaknesses associated with previously derailed managers were averaged together to form a composite of derailment. Raters used a 5-point Likert-type scale for the 40 questions that covered the derailment clusters, with 1 = *Strongly Disagree* (that the manager displays the following) to 5 = *Strongly Agree* (that the manager displays the following). Higher scores (closer to '5' in magnitude) implied that raters more strongly agreed that the manager displayed the characteristic or behavior of derailment. The questions were the same to the manager and his/her observers, only the frame of reference was different (rate yourself vs. rate the person). Coefficient alpha for the derailment measure for each of the four rater sources was above .96.

In the current study, complete data were obtained from one boss, an average of 3.31 direct reports (range 1–10) and 3.18 peers (range 1–8) per manager. To justify aggregating across direct reports and peers for each manager that had more than one rater, intraclass correlation coefficient [ICC(1); Bliese, 2000] statistics were computed. The ICC(1) for peers of Southern Asia managers was .22 ($F = 2.07$, $p < .001$) and for Confucian Asia was .21 ($F = 2.03$, $p < .001$), while the ICC(1) for direct reports of Southern Asia managers was .17 ($F = 1.77$, $p < .001$) and for Confucian Asia was .19 ($F = 1.88$, $p < .001$). These scores usually indicate suitable requirements to aggregate multisource feedback ratings of observers into one score.

2.3. Analytic approach

Multivariate regression procedures were used to examine discrepancy scores. Multivariate regression procedures are a more suitable approach to analyze self–observer discrepancy (difference) scores as outcome variables (see Edwards, 1995). Edwards (1994, 1995) and Edwards and Parry (1993) alluded to many studies that used algebraic, absolute, squared difference, or the sum of absolute or squared difference between self- and observer-ratings. But these approaches are flawed for reasons including unambiguous interpretation, decreased reliability, confounding of the effects of components, and using a univariate framework for what is meant to be measured with a multivariate model, all with the consequence of misleading results (Cronbach, 1958; Cronbach & Furby, 1970; Edwards, 1994, 1995; Edwards & Cooper, 1990; Edwards & Parry, 1993). By combining the self- and observer score into one index, researchers can no longer examine the independent effects from each rating source, and would never discover the underlying explanation for the discrepancy (Edwards, 1995; Ostroff *et al*., 2004). As a result, our research appropriately retains both the self-rating and observer rating separately and tests them jointly for the derailment outcome. Self- and observer ratings were centered based on the midpoint of their shared scale (Edwards, 1994). Therefore, scores ranged from '-2' to '$+2$,' with ratings closer to '-2' in magnitude meaning that managers were less likely to display the characteristics and behaviors of derailment, and scores closer to '$+2$' in magnitude meaning that managers were more likely to display the characteristics and behaviors of derailment.

The hypotheses tested in this study used a multivariate framework with self- and observer ratings considered jointly as outcome variables, and determined whether origin of the manager (Confucian Asia or Southern Asia) predicted a discrepancy between self- and observer ratings of the extent to which a target-manager displays the characteristics and behaviors of derailment. We examined whether the relationship between origin and self–observer ratings for each derailment cluster was significant overall (i.e., an omnibus multivariate test based on Wilks' Λ, a test for self-ratings and for observer ratings jointly). If the Wilks' Λ was statistically significant, then origin of the manager is statistically significantly related to self- and observer ratings considered jointly. The next step would be to inspect the source of the discrepancy. We would then treat the self- and observer ratings as distinct (separate) outcome variables. As a result, we could therefore conclude whether (a) origin of the manager is related to self-ratings, thereby demonstrating that the self–observer discrepancy is due to inflated self-ratings; (b) origin of the manager is related to observer ratings, thereby demonstrating the self–observer discrepancy is due to inflated observer ratings; or (c) a combination of both.

3. Results

The means, standard deviations, and intercorrelations for the study's variables are found in Table 1. The lower-half of the correlation matrix is for Confucian Asia managers, the upper half is for Southern Asia managers.

Table 1. Means, standard deviaions, correlations, and alpha reliabilities among variables

Variable	M_c	SD_c	M_s	SD_s	1	2	3	4
1. Self derailment	−1.28	.49	−1.39	.42	–	.12*	.00	.29**
2. Boss derailment	−1.25	.55	−1.17	.53	.12**	–	.36**	.25**
3. Peer derailment	−1.09	.42	−1.03	.45	.11**	.38**	–	.25**
4. Direct report derailment	−1.13	.43	−1.13	.49	.11**	.27**	.39**	–

Note. Subscript c is for Confucian Asian countries. Subscript s is for Southern Asian Countries. Lower diagonal is intercorrelations between Confucian Asian countries. Upper diagonal is intercorrelations between Southern Asian countries. Derailment scores are centered around midpoint of scale. *$p < .05$; **$p < .01$.

Before comparing data from Confucian and Southern Asia cultural groups, measurement equivalence across these groups was established. Multigroup Structural Equation Modeling (SEM) was used to test the measurement equivalence. The correlation matrix was analyzed and the same reference items were used across cultural groups. The multigroup SEM model tested was among the most constrictive, a parallel-test model. The parallel-test model is defined as correlated measures with equal true-score variances and equal error variances. This model was tested in LISREL by constraining the elements of the Phi, Lambda, and Theta–delta matrices to be equal across groups ($\Phi_{\text{Confucian Asia}} = \Phi_{\text{Southern Asia}}$ and $\Lambda_{\text{Confucian Asia}} = \Lambda_{\text{Southern Asia}}$ and $\Phi_{\delta_{\text{Confucian Asia}}} = \Phi_{\delta_{\text{Southern Asia}}}$). Reliance on a single indicator of model fit can be misleading because fit indicators can lead to differing conclusions (Bollen, 1989; Hu & Bentler, 1995; Kline, 1998). Therefore, the model fit indices reported include the Root Mean Square of Approximation (RMSEA), comparative fit index (CFI), nonnormed fit index (NNFI), and standardized root mean square residual (SRMR). RMSEA is also known as RMS, RMSE or discrepancy per degree of freedom and is an absolute fit index that adjusts for the number of free parameters in the model. RMSEA values of $<.10$ are considered marginal (Steigler, 1990). Values of $<.08$ are generally considered reasonable (Browne & Cudeck, 1993). RMSEA values of $<.05$ are typically indicative of a close fit of the model (Browne & Cudeck, 1993). CFI, proposed by Bentler (1990), is often used to evaluate competing models. NNFI is a generalized version of the Tucker–Lewis Index (Hu & Bentler, 1995). CFI and NNFI are incremental fit indices ranging from 0 to 1, with values of .90 or greater generally considered acceptable (Bentler, 1990; Bentler & Bonett, 1980; Hu & Bentler, 1995). Marsh, Balla, and Hau (1996) advise researchers to provide both indices. SRMR is the mean difference between the predicted and observed variances, based on standardized residuals. The smaller the SRMR, the better the fit. When model fit is perfect SRMR is equal to zero. The selection of fit criteria was informed by recommended practice (Bollen & Long, 1993; Hu & Bentler, 1995; Marsh *et al*., 1996). Adequate model fit across groups with the above constraints in place provides evidence that the measure is parallel and thus equivalent. A comparison of model fit indices for this model and the other models investigated appears as Table 2. The four selected model fit indicators for this model suggest adequate measurement equivalence (RMSEA = .07, CFI = .98, NNFI = .98, SRMR = .06).

Table 2. Tests of model fit for multigroup measurement models

Model	RMSEA	CFI	NNFI	SRMR
Parallel	.07	.98	.98	.06

Note. RMSEA = root mean square error of approximation; CFI = comparative fit index (Bentler, 1990); NNFI = non-normed fit index (Bentler & Bonett, 1980); SRMR = standardized root mean square residual.

Given the cultural differences in Asia (Nandy, 1998) and the subsequent directional nature of the hypothesis (the discrepancy for derailment ratings will be bigger/wider in the Southern Asia cluster than the Confucian Asia cluster of countries for self–boss, self–direct report, and self–peer), the results from the multivariate test between self- and each observer ratings are found in Table 3. Each multivariate analysis shows that origin was significantly related to the set of self- and observer ratings considered jointly as can be seen by the statistically significant Wilks' Λ. Specifically, the Wilks' Λ for the Self–Boss analysis was .98 ($F = 6.85$, $p < .001$) with an effect size (partial eta-squared, η^2) of .02. The Wilks' Λ for the Self–Direct Report analysis was .99 ($F = 4.91$, $p < .01$) with an effect size (η^2) of .01 and the Wilks' Λ for the Self–Peer analysis was .99 ($F = 6.72$, $p < .001$) with an effect size (η^2) of .02. Hypotheses 1a–c are supported, such that rating discrepancies on derailment characteristics and behaviors existed when examining (1a) self–boss ratings; (1b) self–direct report ratings; and (1c) self–peer ratings.

With each of the omnibus tests revealing statistical significance, we then conducted regression analyses treating each self- and observer rating of derailment separately to examine the nature of the relationship between origin and ratings of derailment. These results can also be found at the bottom of Table 3. For each of the three analyses, the regression coefficient for self-ratings was statistically significant ($b = -.11$, $p < .01$). Managers from different origins (Confucian and Southern Asia) rated themselves differently, such that managers from Southern Asia rated themselves as less likely to

Table 3. Multivariate and separate regression analysis results with origin predicting self–observer ratings

Multivariate regression									
Variable	Boss			Direct report			Peer		
	Wilks	ΛF	η^2	Wilks	ΛF	η^2	Wilks	ΛF	η^2
Self–observer	.98	6.85***	.02	.99	4.91**	.01	.99	6.72***	.02

Separate regression analysis						
Variable	*b*	R^2	*b*	R^2	*b*	R^2
Self	−.11**	.011	−.11**	.011	−.11**	.011
Observer	.07	.003	.00	.000	.06	.003

*$p < .05$; **$p < .01$; ***$p < .001$.

display the characteristics and behaviors of derailment (scores were more negative in magnitude, or closer to '−2') than managers from Confucian Asia. In addition, the regression coefficient for each of the observer ratings (boss, direct reports, and peers) was not statistically significant; observers rated the target-managers from both Confucian Asia and Southern Asia statistically similarly on the characteristics and behaviors of derailment. This is visually displayed in Figure 1a (Self–Boss), Figure 1b (Self–Direct Report), and Figure 1c (Self–Peer). As one can see, the discrepancy between the self- and observer rating is wider for the Southern Asia culture than the Confucian Asia culture. Further, according to regression results, self-ratings were the source of the rating discrepancy. Thus, Hypotheses 2a–c were supported, such that the rating discrepancies on derailment characteristics and behaviors were 'bigger/wider' in Southern Asia than in Confucian Asia when examining (2a) self–boss ratings, (2b) self–direct report ratings, and (2c) self–peer ratings.

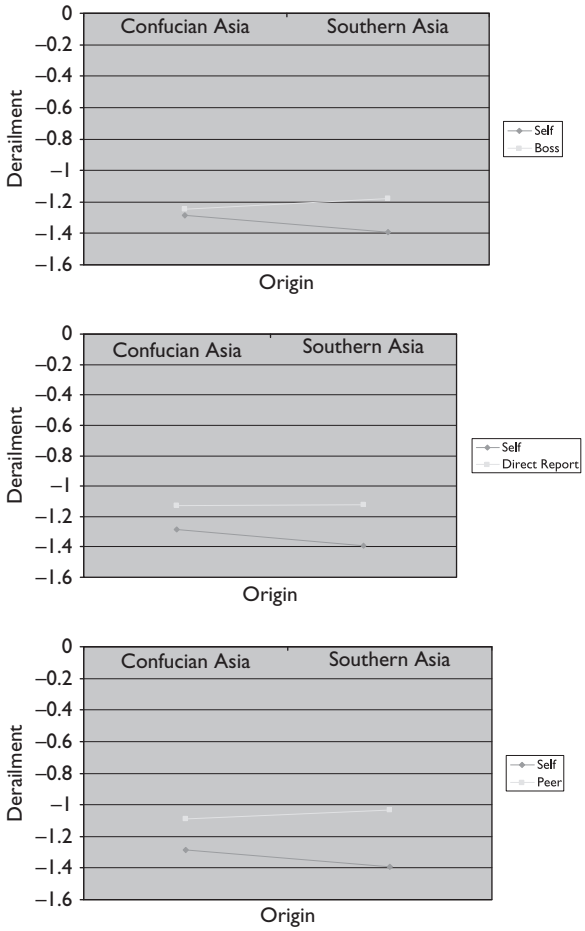

Figure 1. (a) Self–boss discrepancies of derailment behaviors as a function of origin. (b) Self–direct report discrepancies of derailment behaviors as a function of origin. (c) Self–peer discrepancies of derailment behaviors as a function of origin.

4. Discussion

Results from our study revealed via the omnibus multivariate test that a discrepancy between how managers in Asia rated themselves and how their bosses, direct reports, and peers rated them on characteristics and behaviors of derailment existed, such that managers themselves believe they are less likely to display those characteristics and behaviors than each of the three observer rater sources, supporting Hypothesis 1a–c. Furthermore, by separating managers into the cultural clusters of Southern Asia and Confucian Asia, our findings reveal that managers in the Southern Asia cluster have a 'bigger/wider' self–boss, self–direct report, and self–peer rating discrepancy than those in the Confucian Asia countries supporting Hypothesis 2a–c (see Figure 1a–c). Our analysis further revealed that the source of the discrepancy is driven by self-ratings, such that managers in Southern Asia rate themselves differently, seeing themselves as less likely to show derailment signs (scores closer to −2 in magnitude) than managers in Confucian

Asia. Furthermore, the observers (bosses, direct reports, and peers) of these managers in each of the Asian clusters rated the managers the same statistically.

Managers' self-ratings are consistently discrepant from ratings from their boss, direct reports, and peers. In other words, managers in both Confucian and Southern Asia consistently perceive that they are less likely to show derailment signs than the ratings from these three rater sources. Research has attempted to examine what are the antecedents to such discrepancies. Previous self–observer rating discrepancy research has focused on particular levels. For instance, Brutus *et al*. (1999) and Ostroff *et al*. (2004) examined individual-level antecedents, and both Sala (2003) and Gentry *et al*. (2007) focused on organizational-level antecedents to self–observer rating discrepancies. Our research considers cultural antecedents in understanding self–observer discrepancies of derailment potential within Asia.

Through the use of multivariate analyses (Edwards, 1995), one can see that the self–boss, self–direct report, and self–peer rating discrepancy is consistently driven by self-ratings, with a wider discrepancy for managers in Southern Asia than Confucian Asia. Managers are consistently overrating their derailment potential as compared with their observers (i.e., managers see themselves as less likely to show derailment signs than their bosses, peers, or direct reports see them). Self–observer discrepancies may be a key sign of managerial outcomes (Atwater & Yammarino, 1992; Atwater *et al*., 1998, 1995; Church, 1997; Johnson & Ferstl, 1999) and when managers' ratings are discrepant from observer ratings, this can serve as a serious warning sign to managers that they may be on a track toward derailment. This research provides another instance but from a different (cultural) context, that managers' self-ratings are discrepant (different) from the ratings of their observers.

The findings of consistent overrating of derailment potential by Asian managers reflect a 'leniency bias' where self-ratings are more favorable than ratings by other sources (Church, 1997; Nilsen & Campbell, 1993; Podsakoff & Organ, 1986; Van Velsor *et al*., 1993). Social psychologists such as Brown (2003) suggest that leniency and self-enhancement in ratings is a phenomenon across cultures. More specifically, a meta-analysis by Harris and Schaubroeck (1988) found self-ratings to average over half a standard deviation higher than supervisor ratings and approximately one-quarter of a standard deviation higher than peer ratings. However, others argue that this differs across cultures, particularly between Euro-American and Asian cultures (Farh, Dobbins, & Cheng, 1991; Xie, Roy, & Chen, 2006).

The Cultural Relativity Hypothesis by Farh *et al*. (1991) argues that the leniency bias identified in the self-rating literature is primarily a result of the individualistic orientation inherent in Euro-American cultures and proposes that self-raters in Confucian cultures would demonstrate a 'modesty bias' as a result of their collectivistic orientation. While our results are in line with the leniency bias, it also extends the Cultural Relativity Hypothesis, where managers in the more collectivistic Confucian Asia cluster, consistently rate themselves less favorably than their counterparts in the Southern Asia cluster.

The comparatively modest self ratings in the Confucian Asia cluster gives evidence toward the individualist-collectivist interpretation of the Cultural Relativity Hypothesis, where managers in Confucian Asia, with higher institutional collectivism, rate themselves less favorably than their counterparts in Southern Asia. This echoes Markus and Kitayama's (1991) theory that the self-concepts of people differ across cultures and that the self is construed to be more interdependent in collectivist cultures. The more modest ratings in the Confucian Asia cluster can be understood by a culture where people are discouraged to think highly of themselves, in large part because positive self-views conflict with fulfillment of interdependent cultural goals.

As Hui and Triandis (1989) observed, collectivist cultures tend to rate closer to the mean and avoid the use of extreme ratings on scales. Similarly, Hofstede (1980) described collectivist cultures as those with a stronger 'we' consciousness. Evidence of such behavior was provided by Yang and Chiu's (1987) observation that in Confucian cultures, individuals are expected to exercise modesty in describing one's achievements and constantly strive to maintain self-control. As a manifestation of this collectivist culture, individual self-enhancement is seen as a threat to the collective good of society; members are expected to be more occupied with the performance of the group than of one's self.

In contrast to the higher collectivist orientations of the Confucian Asia cluster, the Southern Asia cluster's long exposure to Western influence, with a shared colonial history and practice of Western type of democracy, has inculcated more individualist orientations (Roland, 1988). While the Southern Asia Cluster is more collectivistic than Anglo-European clusters, it is not as high on institutional collectivism as the Confucian Asia cluster (House *et al*., 2004). Managers in the Southern Asia cluster reflect individualist values in their high need for achievement, striving for excellence, competitiveness, and individual freedom (Kumar, 1996; Sinha, 1990).

A further contrast between Confucian and Southern Asia is the higher power distance score in the Southern Asia cluster. The high organizational power distance score in the Southern Asia cluster is reflective of a culture of hierarchical power and control and a high status orientation (Singh & Bhandarker, 1990; Virmani & Gupta, 1981), making feedback difficult. We propose that this further exacerbates the self–observer discrepancy gap. As Hofstede (1980) observed, managers from larger power distance cultures tend to share beliefs that

authority is not to be questioned and that the use of referent power is accepted.

The high power distance scores in the Southern Asia cluster can be attributed to a history and tradition of organizing of society into various socioeconomic class structures reinforced by spiritual beliefs. Colonialism and foreign invasions in the past may also account for submission to power and an institutionalization of power distance within the workplace. While these norms are valued as a sign of respect for authority, they create barriers for feedback and may widen the gap between perceptions of effectiveness between managers and their observers. For example, Shipper, Kincaid, Rotondo, and Hoffman (2003) found that self-awareness of interactive and controlling skills is lower among managers in high power distance cultures.

Given the reluctance of people from high power distance cultures to criticize or evaluate those in positions of authority (Hofstede, 2001), we would expect self-ratings and their distance with observer ratings to be impacted by cultural differences in power distance. In high power distance cultures such as the Southern Asia cluster, managers might be less inclined to elicit feedback and receive feedback from others. People in high power distance cultures may also be less inclined to provide regular feedback and coaching to their managers, due to cultural norms around social distance and boundaries across hierarchies in the organization (Kakar, Kakar, Kets de Vries, & Vrignaud, 2002). We propose this as one explanation for the larger discrepancy between self–observer ratings of managers and their observers in the Southern Asia cluster.

In summary, we proposed that higher organizational power distance in the Southern Asia cluster and the higher institutional collectivism in the Confucian Asia cluster are possible explanations to understand the wider self–observer discrepancy in the Southern Asia cluster. In particular, we identify that self-ratings are the source of the discrepancy for the characteristics and behaviors of derailment. That said, limitations in drawing on an external study to explain cultural variance in our sample are acknowledged, and further research to validate the relationship between cultural variables and rating discrepancies is suggested.

4.1. Limitations and future research

Every research endeavor is faced with challenges and limitations resulting in a less than perfect situation. The present research is no exception. In order to allow readers to weigh the evidence presented, we provide our perspective on the shortcomings of this research as well as ideas for future research.

The observed discrepancies in this study are admittedly small in magnitude. This is certainly a reasonable criticism. However, it is worth noting that the managers providing data for this research tended to be rated as those being less likely to display the characteristics and behaviors of derailment (scores closer to '−2' in magnitude). Thus the available variance was somewhat truncated resulting in smaller discrepancies than might be found in a general population of managers, which leads to the next possible criticism.

A second criticism could be the sample. Practicing Asian managers participating in a multisource feedback process may not reflect the general population of practicing managers in terms of performance. The managers may represent high performers or high potentials in their respective organization. Future research should use a more 'general' population of managers.

Our research suggests that there are cultural variances in self- and observer ratings within Asia. However, the design of our study does not allow us to specify the link between cultural differences and rating discrepancies. There is a need for more research on one's own individual cultural perspective and its impact on self- and observer ratings. In the absence of cultural data at the individual level within our study, we based our interpretations on cluster level data from the GLOBE study (House *et al*., 2004) and Hofstede (1980) in explaining the differences between the Confucian and Southern Asia clusters. We also recognize that the Southern Asia cluster is a constellation of diverse cultures, unlike the Confucian Asia cluster which is fairly homogeneous, with a shared collectivist identity. We have identified organizational power distance as an explanatory factor for the Southern Asia cluster due to its high score from the GLOBE study and in reference to findings that identify power distance as a cross-cutting cultural attribute within this cluster (Hofstede & Bond, 1988; House *et al*., 2004). We propose that further research be conducted at the individual level with cultural variables to examine the effects of power distance and collectivism on self- and observer ratings. Individual cognitive processes for instance need to be considered as potential moderators between the effects of culture on self- and observer ratings (Atwater & Yammarino, 1997; Markus & Kitayama, 1991). To date, there has been little research that considers cultural variations within Asia and its concomitant effects on discrepancies in self- and observer ratings. Future research can also be used to identify the degree to which the effects of cultural variables on ratings (as opposed to individual traits) are significant. Experimental designs can further elucidate the underlying mechanisms of cultural effects on self–observer ratings.

A final limitation could be focused on the derailment variable. More specifically, ratings of a target-manager having or displaying the characteristics and behaviors of previously derailed managers does not equate to actual derailment of the target-manager. Future research should examine actual derailed managers or managers who

recovered from managerial derailment (Kovach, 1989; Shipper & Dillard, 2000). In addition, future research could also examine outcome variables that are not derailment-related, such as leadership-related variables including initiating structure or consideration.

4.2. Implications for the practice of assessment

It is important to note that an accurate and honest feedback system will be of little value if managers ignore or reject information from multisource instruments. Some may believe that self–observer discrepancies are not valuable information to have. However, ratings from sources other than self-ratings can be predictive of objective performance and other outcomes (Atwater & Yammarino, 1992; Atwater *et al*., 1998, 1995; Church, 1997; Johnson & Ferstl, 1999; Sala & Dwight, 2002) and discrepancies in ratings could therefore signal performance outcomes. If 'merely a difference of perception' could impact outcomes, more consideration may be given to self–observer discrepancies, especially when considering the detrimental effects of managerial derailment.

It is possible that managers may not obtain honest feedback (Goleman *et al*., 2001). This could lead to managerial self-perceptions that are not the same as observer perceptions. If true, using a multisource instrument may be one of the few opportunities for managers to obtain honest feedback revealing potential problem areas for managers to address. In addition, fear of retribution could be a concern (Dalton, 1997). As a result, confidentiality is imperative since multisource instruments may be the only opportunity for managers to obtain honest feedback (Dalessio, 1998; Fleenor & Brutus, 2001; Lepsinger & Lucia, 2001; McEvoy & Buller, 1987). Furthermore, managers may be unrealistic or inaccurate in their self-perception because they have never received feedback or never received negative feedback (Yammarino & Atwater, 2001). Rater training may therefore be of particular use both for the manager and observers, covering such topics as reducing leniency or central tendency bias, or more macrolevel issues such as the purpose of the multisource process (Fleenor & Brutus, 2001).

For the practicing manager, derailment ratings should be seen as a means to development and not an end in itself. It is important to recognize that variation in ratings should not be treated as 'true scores,' but rather feedback from different perspectives. Noting that cultural variance plays a significant role in ratings, determining how cultural notions of power and social distance affect a manager's self-rating, and ratings from observers needs to be considered during the assessment process, and during multisource feedback.

Drawing on the findings of this study, those working in assessment and those providing feedback to practicing managers could consider how the cultural framework of different regions in Asia might affect discrepancies in self–observer ratings. As previously discussed, it is important to note that Asia as a geographical entity should not be treated as a homogeneous cultural one. An understanding into the cultural differences between the Southern Asia and Confucian Asia cluster of countries would be helpful in interpreting variance among self–observer ratings. Similarly, cross-cultural implications of interpreting 360-feedback in and across Asia should be considered. As Leslie, Gryskiewicz, and Dalton (1998) observed, in cultures with high organizational power distance, the shift in the evaluation role from the supervisor to other coworkers may run counter to the power structure of organizations, which have implications for multisource assessments and their interpretation. Highlighting cultural effects on self- and observer ratings may prove to be an important source for assessment, training, development, and the use of multisource feedback in Asia.

The effects of managerial derailment can be extensive (Finkelstein, 2004; Smart, 1999; Wells, 2005). The results of this study broadened and expanded upon previous multisource research by focusing on derailment and examining cultural antecedents of self–observer rating discrepancies on Asian managers. In this study, Asian managers believed they were less likely to display the characteristics and behaviors associated with derailment than their observers believed. There were also noted differences between different cultures, such that those in the Southern Asia cluster had a 'bigger/wider' self–observer rating discrepancy in displaying the characteristics and behaviors associated with derailment than those of the Confucian Asia cluster, with self-ratings being a major contributor to the discrepancy. The discrepancy between self- and observer ratings may specify an issue that managers, organizations, and those in the service of assessments, should address around the world, particularly in an Asian context.

References

Atwater, L. and Waldman, D. (1998). Accountability in 360 degree feedback. *HR Magazine*, *43*, 96–104.

Atwater, L., Waldman, D., Ostroff, C., Robie, C. and 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., Brett, J.F. and Charles, A.C. (2007). Multisource feedback: Lessons learned and implications for practice. *Human Resource Management*, *46*, 285–307.

Atwater, L.E., Ostroff, C., Yammarino, F.J. and Fleenor, J.W. (1998). Self–other agreement: Does it really matter? *Personnel Psychology*, *51*, 577–598.

Atwater, L.E., Roush, P. and Fischthal, A. (1995). The influence of upward feedback on self and follower ratings of leadership. *Personnel Psychology, 48*, 35–59.

Atwater, L.E. and Yammarino, F.J. (1992). Does self–other agreement on leadership perceptions moderate the validity of leadership and performance predictions? *Personnel Psychology, 45*, 141–164.

Atwater, L.E. and Yammarino, F. (1997). Self–other rating agreement: A review and model. In Ferris, G.R. (Ed.), *Research in personnel and human resources management* (*Vol. 15*, pp. 141–164). Greenwich, CN: JAI Press Inc.

Bentler, P.M. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, *107*, 238–246.

Bentler, P.M. and Bonett, D.G. (1980). Significance tests and goodness of fit in the analysis ofcovariance structures. *Psychological Bulletin*, *88*, 588–606.

Bliese, P.D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analyses. In Klein, K.J. and Kozlowski, S.W.J. (Eds.), *Multilevel theory, research, and methods in organizations: foundations, extensions, and new directions* (pp. 349–381). San Francisco, CA: Jossey-Bass.

Bollen, K.A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Bollen, K.A. and Long, J.S. (Eds.). (1993). *Testing structural equation models*. Newbury Park, CA: Sage.

Brown, J.D. (2003). The self-enhancement motive in collectivistic cultures: The rumors of my death have been greatly exaggerated. *Journal of Cross-Cultural Psychology, 34*, 603–605.

Browne, M.W. and Cudeck, R. (1993). Alternative ways of assessing model fit. In Bollen, K.A. and Long, J.S. (Eds.), *Testing structural equation models*. Newbury Park, CA: Sage.

Brutus, S., Fleenor, J.W. and 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.

Brutus, S., Fleenor, J.W. and McCauley, C.D. (1999). Demographic and personality predictors of congruence in multi-source ratings. *Journal of Management Development*, *18*, 417–435.

Brutus, S., Leslie, J.B. and McDonald-Mann, D. (2001). Cross-cultural issues in multisource feedback. In Bracken, D., Timmreck, C. and Church, A. (Eds.), *The handbook of multisource feedback: the comprehensive resource for designing and implementing MSF processes* (pp. 433–446). San Francisco, CA: Jossey-Bass.

Carty, H.M. (2003). Review of BENCHMARKS® [revised]. In Plake, B.S., Impara, J. and Spies, R.A. (Eds.), *The fifteenth mental measurements yearbook* (pp. 123–124). Lincoln, NE: Buros Institute of Mental Measurements.

Cattell, R. (1950). The principal culture patterns discoverable in the syntax dimensions of existing nations. *Journal of Social Psychology, 32*, 215–253.

Center for Creative Leadership (2004). *BENCHMARKS® facilitator's guide*. Greensboro, NC: Author.

Chen, M. (2004). *Asian management systems*. London, UK: Thomson.

Chhokar, J.S., Brodbeck, F.C. and House, R.J. (Eds.). (2007). *Culture and leadership across the world: the globe book of in-depth studies of 25 societies*. Mahwah, NJ: Erlbaum.

Church, A.H. (1997). Managerial self-awareness in high-performing individuals in organizations. *Journal of Applied Psychology, 82*, 281–292.

Church, A.H. (2000). Do higher performing managers actually receive better ratings? A validation of multirater assessment methodology. *Consulting Psychology Journal: Practice and Research*, *52*, 99–116.

Conger, J.A. and Nadler, D.A. (2004). When CEOs step up to fail. *MIT Sloan Management Review, 45*, 50–56.

Conway, J.M. (2000). Managerial performance development constructs and personality correlates. *Human Performance*, *13*, 23–46.

Conway, J.M., Lombardo, K. and Sanders, K.C. (2001). A meta-analysis of incremental validity and nomological networks for subordinate and peer ratings. *Human Performance*, *14*, 267–303.

Cronbach, L.J. (1958). Proposals leading to analytic treatment of social perceptions scores. In Tagiuri, R. and Petrulo, L. (Eds.), *Person perception and interpersonal behavior* (pp. 353–379). Stanford, CA: Stanford University Press.

Cronbach, L.J. and Furby, L. (1970). How should we measure "change" – or should we? *Psychological Bulletin*, *74*, 68–80.

Dalessio, A.T. (1998). Using multi-source feedback for employee development and personnel decisions. In Smither, J.W. (Ed.), *Performance appraisal: state-of-the-art in practice* (pp. 278–330). San Francisco, CA: Jossey-Bass.

Dalton, M.A. (1997). When the purpose of using mutli-rater feedback is behavior change. In Bracken, D.W., Dalton, M.A., Jako, R.A., McCauley, C.D. and Pollman, V.A. (Eds.), *Should 360-degree feedback be used only for developmental purposes?* (pp. 1–6). Greensboro, NC: Center for Creative Leadership.

Dotlich, D.L. and Cairo, P.C. (2003). *Why CEOs fail: the 11 behaviors that can derail your climb to the top – and how to manage them*. San Francisco, CA: Jossey-Bass.

Douglas, C.A. (2003). *Key events and lessons for managers in a diverse workforce: A report on research and findings*. Greensboro, NC: Center for Creative Leadership.

Edwards, J.R. (1994). The study of congruence in organizational behavior research: Critique and proposed alternative. *Organizational Behavior and Human Decision Processes*, *58*, 51–100.

Edwards, J.R. (1995). Alternatives to difference scores as dependent variables in the study of congruence in organizational research. *Organizational Behavior and Human Decision Processes*, *64*, 307–324.

Edwards, J.R. and Cooper, C.L. (1990). The person-environment fit approach to stress: Recurring problems and some suggested solutions. *Journal of Organizational Behavior, 10*, 293–307.

Edwards, J.R. and Parry, M.E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal*, *36*, 1577–1613.

Farh, J.L., Dobbins, G.H. and Cheng, B.S. (1991). Cultural relativity in action: A comparison of self-ratings made by Chinese and U.S. Workers. *Personnel Psychology, 44*, 129–147.

Finkelstein, S. (2004). *Why smart executives fail*. New York, NY: Portfolio.

Finkin, E.F. (1991, March/April). Techniques for making people more productive. *The Journal of Business Strategy, 12*, 53–56.

Fleenor, J.W. and Brutus, S. (2001). Multisource feedback for personnel decisions. In Bracken, D.W., Timmreck, C.W. and Church, A.H. (Eds.), *The handbook of multisource feedback: the comprehensive resource for designing and implementing MSF processes* (pp. 335–351). San Francisco, CA: Jossey-Bass Inc.

Fleenor, J.W., McCauley, C.D. and Brutus, S. (1996). Self–other rating agreement and leader effectiveness. *Leadership Quarterly*, *7*, 487–506.

Gentry, W.A., Hannum, K.M., Ekelund, B.Z. and de Jong, A. (2007). A study of the discrepancy between self- and observer-ratings on managerial derailment characteristics of European managers. *European Journal of Work and Organizational Psychology*, *16*, 295–325.

Gentry, W.A., Katz, R.B. and McFeeters, B. (2009). The continual need for improvement to avoid derailment: A study of college and university administrators. *Higher Education Research and Development*, *28*, 335–348.

Gillespie, N.A., Walsh, M., Winefield, A.H., Dua, J. and Stough, C. (2001). Occupational stress in universities: Staff perceptions of the causes, consequences, and moderators of stress. *Work and Stress*, *15*, 53–72.

Goleman, D., Boyatzis, R. and McKee, A. (2001). Primal leadership: The hidden driver of great performance. *Harvard Business Review*, *79*, 42–51.

Graves, L.M., Ohlott, P.J. and Ruderman, M.N. (2007). Commitment to family roles: Effects on managers' attitudes and performance. *Journal of Applied Psychology*, *92*, 44–56.

Gupta, V., Hanges, P.J. and Dorfman, P.W. (2002). Cultural clusters: Methodology and findings. *Journal of World Business*, *37*, 3–10.

Gupta, V., Surie, G., Javidan, M. and Chhokar, J. (2002). Southern Asia cluster: Where the old meets the new? *Journal of World Business*, *37*, 16–27.

Harris, M.M. and Schaubroeck, J. (1988). A meta-analysis of self–supervisor, self–peer, and peer–supervisor ratings. *Personnel Psychology*, *41*, 43–61.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Newbury Park, CA: Sage.

Hofstede, G. (2001). *Culture's consequences* (2nd ed.). Beverly Hills, CA: Sage.

Hofstede, G. and Bond, M.H. (1988). The Confucius connection: From cultural roots to economic growth. *Organizational Dynamics*, *16*, 5–21.

Hogan, R. and Hogan, J. (2001). Assessing leadership: A view from the dark side. *International Journal of Selection and Assessment*, *9*, 40–51.

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

Hu, L.T. and Bentler, P.M. (1995). Evaluating model fit. In Hoyle, R.H. (Ed.), *Structural equation modeling: concepts, issues, and applications* (pp. 76–99). Thousand Oaks, CA: Sage.

Hui, C.H. and Triandis, H.C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology*, *20*, 296–309.

Johnson, J.W. and Ferstl, K.L. (1999). The effects of inter-rater and self–other agreement on performance improvement following upward feedback. *Personnel Psychology*, *52*, 272–303.

Kakar, S., Kakar, S., Kets de Vries, M.F.R. and Vrignaud, P. (2002). Leadership in Indian organizations from a comparative perspective. *International Journal of Cross Cultural Management*, *2*, 239–250.

Kaplan, R.E., Drath, W.H. and Kofodimos, J.R. (1987). High hurdles: The challenge of executive self-development. *Academy of Management Executive*, *1*, 195–205.

Kline, R.B. (1998). *Principles and practice of structural equation modeling*. New York, NY: Guilford Press.

Kovach, B.E. (1986). The derailment of fast-track managers. *Organizational Dynamics*, *15*, 41–48.

Kovach, B.E. (1989). Successful derailment: What fast-trackers can learn while they're off the track. *Organizational Dynamics*, *18*, 33–47.

Kramer, R. (2003). Recognizing the symptoms of reckless leadership. *Harvard Business Review*, *81*, 65.

Kumar, S. (1996, Summer). Towards evolving an Indian style of management based on Indian values and work ideals. *Abhigyan*, 25–34.

Lau, C.M. (2002). Asian management research: Frontiers and challenges. *Asia Pacific Journal of Management*, *19*, 171–178.

Lepsinger, R. and Lucia, A.D. (2001). Performance management and decision making. In Bracken, D., Timmreck, C. and Church, A. (Eds.), *The handbook of multisource feedback: the comprehensive resource for designing and implementing MSF processes* (pp. 318–334). San Francisco, CA: Jossey-Bass.

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

Leslie, J.B. and Van Velsor, E. (1996). *A look at derailment today: North American and Europe*. Greensboro, NC: Center for Creative Leadership.

Levinson, H. (1994). Beyond the selection failures. *Consulting Psychology Journal*, *46*, 3–8.

Lindsey, E.H., Homes, V. and McCall, M.W. Jr. (1987). *Key events in executives' lives*. Greensboro, NC: Center for Creative Leadership.

Lombardo, M.M. and Eichinger, R.W. (1989/2005). *Preventing derailment: what to do before it's too late*. Greensboro, NC: Center for Creative Leadership.

Lombardo, M.M. and McCauley, C.D. (1988). *The dynamics of management derailment. (Tech. Rep. No. 34)*. Greensboro, NC: Center for Creative Leadership.

Lombardo, M.M. and McCauley, C.D. (1994). *BENCHMARKS®: a manual and trainer's guide*. Greensboro, NC: Center for Creative Leadership.

Lombardo, M.M., McCauley, C.D., McDonald-Mann, D. and Leslie, J.B. (1999). *BENCHMARKS®: developmental reference points*. Greensboro, NC: Center for Creative Leadership.

London, M. (2003). *Job feedback*. Mahwah, NJ: LEA.

London, M. and Smither, J.W. (1995). Can multi-source feedback change self-evaluations, skill development, and performance? Theory-based applications and directions for research. *Personnel Psychology*, *48*, 375–390.

Lyness, K.S. and Judiesch, M.K. (2008). Can a manager have a life and a career? International and multisource perspectives on

work-life balance and career advancement potential. *Journal of Applied Psychology*, *93*, 789–805.

Markus, H.R. and Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion and motivation. *Psychological Review*, *98*, 224–253.

Marsh, H.W., Balla, J.R. and Hau, K.T. (1996). An evaluation of incremental fit indices: A clarification of mathematical and empirical properties. In Marcoulides, G.A. and Schmacker, R.E. (Eds.), *Advanced structural equation modeling: issues and techniques* (pp. 315–353). Mahwah, NJ: Lawrence Erlbaum.

Mathews, J.A. (2006). Dragon multinationals: New players in 21st century globalization. *Asia Pacific Journal of Management*, *23*, 5–27.

McCall, M. and Lombardo, M. (1983). *Off the track: why and how successful executives get derailed*. Greensboro, NC: Center for Creative Leadership.

McCauley, C. and Lombardo, M. (1990). BENCHMARKS®: An instrument for diagnosing managerial strengths and weaknesses. In Clark, K.E. and Clark, M.B. (Eds.), *Measures of leadership* (pp. 535–545). West Orange, NJ: Leadership Library of America.

McCauley, C., Lombardo, M. and Usher, C. (1989). Diagnosing management development needs: An instrument based on how managers develop. *Journal of Management*, *15*, 389–403.

McEvoy, G.M. and Buller, P.F. (1987). User acceptance of peer appraisals in an industrial setting. *Personnel Psychology*, *40*, 785–797.

Meyer, K.E. (2006). Asian management research needs more self-confidence. *Asia Pacific Journal of Management*, *23*, 119–137.

Morgeson, F.P., Mumford, T.V. and Campion, C.A. (2005). Coming full circle: Using research and practice to address 27 questions about 360-degree feedback programs. *Consulting Psychology Journal: Practice and Research*, *57*, 196–209.

Morrison, A.M., White, R.P. and Van Velsor, E. (1987). *Breaking the glass ceiling: can women reach the top of America's largest corporations?* Reading, MA: Addison-Wesley.

Mount, M.K., Judge, T.A., Scullen, T.E., Sytsma, M.R. and Hezlett, S.A. (1998). Trait, rater, and level effects in 360-degree performance ratings. *Personnel Psychology*, *51*, 557–576.

Nandy, A. (1998). Defining a new cosmopolitanism: Towards a dialogue of Asian civilizations. In Chen, K.-H. (Ed.), *Trajectories: Inter-Asia cultural studies* (pp. 142–149). London, UK: Routledge.

Nilsen, D. and Campbell, D.P. (1993). Self–observer rating discrepancies: Once an overrater, always an overrater? *Human Resource Management*, *32*, 265–281.

Ostroff, C., Atwater, L.E. and Feinberg, B.J. (2004). Understanding self–other agreement: A look at rater and ratee characteristics, context, and outcomes. *Personnel Psychology*, *51*, 333–375.

Podsakoff, P. and Organ, D. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management*, *12*, 86–94.

Roland, A. (1988). *In search of self in India and Japan: towards a cross-cultural psychology*. Princeton, NJ: Princeton University Press.

Ronen, S. and Shenkar, O. (1985). Clustering countries on attitudinal dimensions: A review and synthesis. *Academy of Management Review*, *10*, 435–454.

Sala, F. (2003). Executive blind spots: Discrepancies between self- and other-ratings. *Consulting Psychology Journal*, *55*, 222–229.

Sala, F. and Dwight, S.A. (2002). Predicting executive performance with multi-rater surveys: Whom you ask makes a difference. *Consulting Psychology Journal*, *54*, 166–172.

Shipper, F. and Dillard, J.E. Jr. (2000). A study of impending derailment and recovery of middle managers across career stages. *Human Resource Management*, *39*, 331–345.

Shipper, F., Kincaid, J.F., Rotondo, D.M. and Hoffman, R.C. IV (2003). A cross cultural exploratory study of the linkage between emotional intelligence and managerial effectiveness. *The International Journal of Organizational Analysis*, *11*, 171–191.

Singh, P. and Bhandarker, A. (1990). *Corporate success and transformational leadership*. New Delhi, India: Wiley Eastern.

Sinha, J.B.P. (1990). *Work culture in the Indian context*. New Delhi, India: Sage.

Smart, B. (1999). *Topgrading: How leading companies win by hiring, coaching, and keeping the best people*. Upper Saddle River, NJ: Prentice Hall.

Smither, J.W. and Walker, A.G. (2001). Measuring the impact of multisource feedback. In Bracken, D., Timmreck, C. and Church, A. (Eds.), *The handbook of multisource feedback: the comprehensive resource for designing and implementing MSF processes* (pp. 256–271). San Francisco, CA: Jossey-Bass.

Spangler, M. (2003). Review of BENCHMARKS® [revised]. In Plake, B.S., Impara, J. and Spies, R.A. (Eds.), *The fifteenth mental measurements yearbook* (pp. 124–126). Lincoln, NE: Buros Institute of Mental Measurements.

Steigler, J. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*, 173–180.

Tornow, W.W. (1993). Perceptions or reality: Is multi-perspective measurement a means or an end? *Human Resource Management*, *32*, 246–264.

Tsui, A.S. and Ohlott, P.J. (1988). Multiple assessment of managerial effectiveness: Interrater agreement and consensus in effectiveness models. *Personnel Psychology*, *41*, 779–803.

Van Velsor, E., Taylor, S. and Leslie, J.B. (1993). An examination of the relationships among self-perception accuracy, self-awareness, gender, and leader effectiveness. *Human Resource Management*, *32*, 249–263.

Varela, O.E. and 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.

Virmani, B.R. and Gupta, S.U. (1981). *Indian management*. New Delhi, India: Vision.

Wells, S.J. (2005). Diving in. *HR Magazine*, *50*, 54–59.

Xie, J.L., Roy, J-P. and Chen, Z.G (2006). Cultural and individual differences in self-rating behavior: An extension and refinement of the cultural relativity hypothesis. *Journal of Organizational Behavior*, *27*, 341–364.

Yammarino, F. (2003). Modern data analytic techniques for multisource feedback. *Organizational Research Methods*, *6*, 6–14.

Yammarino, F. and Atwater, L. (1997). Do managers see themselves as others see them? Implications of self–other rating agreement for human resource management. *Organizational Dynamics*, *25*, 35–44.

Yammarino, F. and Atwater, L. (2001). Understanding agreement in multisource feedback. In Bracken, D., Timmreck, C. and Church, A. (Eds.), *The handbook of multisource feedback: the comprehensive resource for designing and implementing MSF processes* (pp. 204–220). San Francisco, CA: Jossey-Bass.

Yang, C.F. and Chiu, C.Y. (1987). The dilemmas experienced by chinese subjects: A critical review of their influence on the use of Western originated rating scales. *Acta Psychologica Taiwanica*, *29*, 113–132 (in Chinese).

Zedeck, S. (1995). Review of BENCHMARKS®. In Conoley, J. and Impara, J. (Eds.), *The twelfth mental measurements yearbook* (*Vol. 1*, pp. 128–129). Lincoln, NE: Buros Institute of Mental Measurements.