Finding Value in 360-Feedback Rater Disagreements

Renee Yang
&
Karen M. Fuhrmeister

Hogan Assessment Systems

Session Abstract

360-degree feedback has long been used by organizations. However, there is a lack of consensus on how to compile feedback from different sources. The goal of this symposium is to discuss rater disagreements from both theoretical and empirical perspectives, and demonstrate the value of understanding unique inputs from various sources.

Session Summary

The literature on multisource performance appraisal, or "360 ratings," is well established. Organizations use 360's for multiple purposes such as predicting performance (Atwater & Yammarino, 1992), developing leadership competencies (Dai, De Meuse, & Peterson, 2010), and assessing team performance (Reilly & McGourty, 1998). However, despite their popularity in both applied and academic circles, some still question the utility of 360's because of issues surrounding potential rater error and bias (Hauenstein, 1998; Kane & Lawler, 1978; Pulakos, 1997). Of the errors and biases that may influence multisource performance appraisal results, researchers have focused most on observed discrepancies in ratings provided by different groups or rater sources (i.e., supervisors versus subordinates).

With a few exceptions (e.g., Bliese, 2000; LeBreton et al., 2003), most researchers find limited convergence in ratings provided by different parties. Correlations among ratings provided by different groups tend to be modest at best (e.g., Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Mount, 1984). Average ratings also differ by source. Numerous researchers (e.g., Harris & Schaubroeck, 1988; Mount, 1984; Thornton, 1980) report significant differences between self ratings and ratings provided by others, such as supervisors, peers, and direct reports.

Historically, researchers dismissed observed discrepancies in between-source ratings as error (Viswesvaran, Schmidt, & Ones, 2005). More recently, however, professionals have re-examined the meaning of these differences, reconsidering discrepancies as a function of differing conceptualizations of effective performance held by various rater groups (Campbell & Lee, 1988) or differing opportunities to observe the behavior of a target (Lance, Teachout, & Donnelly, 1992; Murphy & Cleveland, 1995). This ecological perspective suggests that we should not expect convergence of between-source ratings (Lance & Woehr, 1989), and that observed variance offers a more complete account of employee performance (Hoffman & Woehr, 2009). Examining rating discrepancies can lead to richer and more tailored feedback discussions, and may lead managers to identify areas in which performance improvements are necessary (Vecchio & Anderson, 2009). The goal of this symposium is to bring together a variety of perspectives to gain an understanding of how to leverage 360 feedback from multiple rater sources. Specifically, we seek to answer two questions: (a) what causes rater disagreement and (b) what is the value of understanding rater disagreement?

The Dai paper presents a theoretical perspective on why between-source discrepancies occur. Drawing from past research on rating behaviors, the author proposes that raters form different frame of reference when rating difference constructs, which affects the extent of inter-rater agreements. Using archival data, Dai compared inter-rater agreement patterns on two different

constructs (skills vs. importance) and discussed the relationship between the scope of those constructs and inter-rater discrepancy.

To explain self-other disagreements from an individual differences perspective, Yang, Fuhrmeister, and Do examine 360-degree feedback data from over 200 Australian leaders in a variety of industries to explore the association between ratee personality and self-other discrepancy. Specifically, they demonstrate the value of understanding ratee characteristics when designing appropriate interventions for performance improvements, which carries implications for 360 research and practice.

As a step further, the McCook study uses 360-degree feedback from an unusually complete data set (with self ratings, other ratings, as well as personality assessment data for both the ratees and raters) to examine the interaction between rater and ratee personality, and its impact on self-other disagreements. Specifically, the author looks at how differences between a leader and their raters can affect the gap between self-other ratings.

Finally, David Peterson examines issues related to self-other disagreements. He will discuss three issues: how to get a better sense of the appropriate criterion in 360 assessments, how we can improve our understanding of bias or error in both self and other ratings, and how we can apply these insights to individual 360-feedback.

Our discussant, Dr. Anna Brown is a psychometrician with an established reputation and extensive industry experience. Her research has been focusing on the measurement of non-cognitive domains, particularly workplace competencies and personal styles. Dr. Brown's main areas of expertise and research interests include modelling of comparative judgments, modelling response processes contributing to common biases, and impression management in questionnaire data, test optimisation and computerised adaptive testing (CAT). During the symposium, she will share her extensive international experience with 360-degree feedback research and practice, as well as her evaluation of the research presented.

Summary of Hogan's Contribution

Multisource feedback (hereafter "360") has been increasingly popular since its emergence (Church, 1995; Hazucha, Hezlett, & Schneider, 1993; London & Beatty, 1993). Because single-source supervisor ratings often lead to criterion deficiency (e.g., Cooper, 1981; Oh & Berry, 2009), researchers and practitioners often turn to 360 to compile additional inputs from self, subordinates, peers, and other resources (e.g., clients) to more comprehensively capture job performance (Conway & Huffcutt, 1997; Craig & Hannum, 2006). However, despite the popularity of 360 practices, there is a lack of agreement upon the value of multisource feedback.

Often, arguments against 360 focus on the relationships between ratings provided by different sources. Particularly, while numerous researchers (e.g., Harris & Schaubroeck, 1988; Mount, 1984; Thornton, 1980) find significant differences between self-ratings and ratings provided by others (e.g., supervisors, peers, and direct reports), a few others challenge the existence of true rater disagreements (e.g., Bliese, 2000; LeBreton et al., 2003) and advocate aggregating inputs from multiple sources. The purpose of the present study is to provide empirical evidence for the value of examining disagreements between self-ratings and those provided by other sources. Specifically, we explore the relationship between ratee personality and self-other disagreements.

According to Bracken et al. (2001), the primary goal of 360 is to motivate behavior change through feedback. Therefore, it is critical to understand the personal attributes of ratees to develop tailored interventions that ensure performance improvements. One such personal attribute worth investigating is self-awareness. Past research on rating biases suggest that self-ratings can be unreliable due to factors such as self-serving attribution bias, actor-observer effect, and self-enhancement mechanisms (Farh & Dobbins, 1989a; 1989b). Yet, being aware of ones strengths and limitations is important for leadership effectiveness.

Early research on managerial self-awareness using limited samples of managers from specific industries (e.g., navel officers, hospital administrators) show significant relationship between self-awareness and leadership effectiveness (e.g., Ashford & Tsui, 1991; Atwater and Yammarino, 1992; Van Velsor et al., 1993). As an extension of these findings, Church (1997) used several different organizational samples representative of the general management populations and a variety of managerial performance measures to explore the link between self-awareness and managerial excellence. According to Church (1997), high-performing managers show higher congruence between self-ratings and ratings from direct reports. Besides implications for positive leadership behaviors, a recent study by Tang, Dai, and De Meuse (2011) also reveals a close relationship between self-other disagreements and leadership derailment behaviors. For example, self under-ratings of leadership derailment factors are related to lower leadership effectiveness.

Despite the close relationship between self-awareness and leadership effectiveness, questions remain concerning the mechanisms by which individuals differ in self-awareness. To fill this gap in existing research, we examined data from a sample of employees in various managerial jobs across Australia to examine the relationship between ratee personality and self-other disagreements in 360 evaluations.

4

Our sample included 233 managers and executives who completed the Hogan Personality Inventory (HPI; R. Hogan & Hogan, 2007), the Hogan Development Survey (HDS; R. Hogan & Hogan, 2009), the Motives, Values, Preferences Inventory (MVPI; R. Hogan & J. Hogan, 2010), and a 360-degree leadership performance measurement (Peter Berry Consultancy, 2009). The HPI is the first Five-Factor Model personality measure designed for use in business settings within a normal population; the HDS is the most validated measure of personality developed to assess personality characteristics associated with job derailment; and the MVPI provides insights into person-organization fit. The 360 tool used in the present study measures leadership behaviors as defined by a four-domain model, which includes Self Management, Relationship Management, Business Skills, and Strategic Skills. The feedback process involved the target manager making self-ratings of performance. Other sources, such as direct reports, peers, managers, and others (e.g., clients), also rated the target manager's performance using the same items.

Our first step was to compare mean rating differences by rater group. We conducted analyses of variance (ANOVA) and Tukey post-hoc comparisons for scores in each performance domain. Results indicated significant differences between self-ratings and ratings from other sources across the four domains. As an example (Table 1), on Self Management, we found significant difference between self-ratings and inputs from direct reports. Next, to examine the relationship between ratee personality and rater disagreements, we correlated ratee personality scale scores with self-other discrepancies. We computed self-other discrepancies for each ratee-rater pair by subtracting ratings from each source from self-ratings. This approach allows us to account for both the magnitude and direction of rater disagreements. Results indicated significant correlations between a number of personality scales and rater disagreements for each domain. For example, as shown in Table 2, HPI Ambition scores show significant and positive correlations with the discrepancy between self-ratings and direct report ratings across four domains. This indicates that, across a variety of performance areas, individuals who are ambitious tend to view themselves as more competent than their subordinates view them.

These findings facilitate understanding of individual differences in ratee self-awareness. Moreover, the significant relationship between ratee personality and self-other discrepancy suggests that certain individual characteristics, such as ambition, may contribute to these differences and a tendency to lack self-awareness. These findings may not only help explain self-other discrepancies on 360 measures, but also provide further information to use in the development process. During the symposium, we will present our methodology, share and interpret our findings, and discuss their implications for 360 research and practice.

References

Ashford, S. J. & Tsui, A. S. (1991). Self-regulation for managerial effectiveness: The role of active feedback seeking. *Academy of Management Journal*, *34*, 251-280. doi: 10.2307/256442

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

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

Bracken, D. W., Timmreck, C. W., & Church. A. H. (2001). *The Handbook of Multi-source Feedback: The Comprehensive Resource for Designing and Implementing MSF Processes*. San Francisco, CA: Jossey-Bass.

Campbell, D. J., & Lee, C. (1988). Self-appraisal in performance evaluation: Development versus evaluation. *Academy of Management Review, 13*, 302 – 314.

Church, A. H. (1995). First-rate multirater feedback. *Training & Development*, *49*, 42-43.

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

Conway, J. M., & Huffcutt, A. I. (1997). Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance*, *10*, 331-360.

Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin*, *90*, 218-244.

Craig, S. B., & Hannum, K. (2006). Research update: 360-degree performance assessment. *Consulting Psychology Journal: Practice and Research*, *58*, 117.

Dai, G., De Meuse, K.P., & Peterson, C. (2010). Impact of Multi-Source Feedback on Leadership Competency Development: A Longitudinal Field Study. *Journal of Managerial Issues, 22*(2), 197-219.

Farh, J. L., & Dobbins, G. H. (1989a). Effects of comparative performance information on the accuracy of self-ratings and agreement between self and supervisor ratings. *Journal of Applied Psychology, 74*, 606-610.

Farh, J. L., & Dobbins, G. H. (1989b). Effects of self-esteem on leniency bias in self-reports of performance: A structural equation model analysis. *Personnel Psychology, 42*, 835-850.

Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology*, *41*, 43-62.

Hauenstein, N. M. A. (1998). Training raters to increase the accuracy of appraisals and the usefulness of feedback. In J. W. Smither (Ed.), *Performance Appraisal: State of the Art in Practice* (pp. 404 – 444). San Francisco, CA: Jossey-Bass Publishers.

Hazucha, J. F., Hezlett, S. A., & Schneider, R. J. (1993). The impact of 360-degree feedback on management skills development. *Human Resource Management*, *32*, 325-351.

Hoffman, B. J., & Woehr, D. J. (2009). Disentagling the meaning of multisource performance rating source and dimension factors. *Personnel Psychology, 62,* 735 – 765.

Hogan, R., & Hogan, J. (2007). *Hogan Personality Inventory manual* (3rd ed.). Tulsa, OK: Hogan Assessment Systems.

Hogan, R., & Hogan, J. (2009). *Hogan Development Survey manual* (2nd ed.). Tulsa, OK: Hogan Press.

Hogan, J., & Hogan, R. (2010). *Motives, Values, Preferences Inventory manual: 2010 administrative and norming updates*. Tulsa, OK: Hogan Press.

Kane, J. S., & Lawler, E. E. (1978). Methods of peer assessment. *Psychological Bulletin, 85*, 555 – 586.

Lance, C. E., & Woehr, D. J. (1989). The validity of performance judgments: Normative accuracy model versus ecological perspectives. In D. F. Ray (Ed.), *Proceedings of the Southern Management Association* (pp. 115 – 117). Starkville, MS: Southern Management Association.

Lance, C. E., Teachout, M. S., & Donnelly, T. M. (1992). Specification of the criterion construct space: An application of hierarchical confirmatory factor analysis. *Journal of Applied Psychology, 77*, 437 – 452.

LeBreton, J. M., Burgess, J. R., 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.

London, M., & Beatty, R. W. (1993). 360-degree feedback as a competitive advantage. *Human Resource Management*, *32*, 353-372.

Mount, M. K. (1984). Psychometric properties of subordinate ratings of managerial performance. *Personnel Psychology*, *37*, 687-702.

Murphy, K. R., & Cleveland, J. N. (1995). *Understanding Performance Appraisal: Social, Organizational, and Goal Based Perspectives.* Thousand Oaks, CA: Sage.

Oh, I. S., & Berry, C. M. (2009). The five-factor model of personality and managerial performance: validity gains through the use of 360 degree performance ratings. J*ournal of Applied Psychology*, *94*, 1498.

Peter Berry Consultancy (2009). T*he 360 review and information*. Retrieved September 1, 2014 from http://www.peterberry.com.au/files/pbc_instruments_/360_fact_sheet_2009.pdf.

Pulakos, E. D. (1997). Ratings of job performance. In D. Whetzel and G. R. Wheaton (Eds.), *Applied Measurement Methods in Industrial Psychology* (pp. 291 – 318). Palo Alto, CA: Davies-Black Publishing.

Reilly, R. R., & McGourty, J. (1998). Performance appraisal in team settings. In J. W. Smither (Ed.), *Performance Appraisal: State of the Art in Practice* (pp. 244 – 277). San Francisco, CA: Jossey-Bass Publishers.

Tang, K. Y., Dai, G., & De Meuse, K. P. (2013). Assessing leadership derailment factors in 360° feedback: Differences across position levels and self-other agreement. *Leadership & Organization Development Journal*, *34*, 326-343.

Thornton, G. C., III (1980). Psychometric properties of self-appraisals of job performance. *Personnel Psychology, 33*, 263-271.

Van Velsor, E., Taylor, S., & 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.

Vecchio, R. P., & Anderson, R. J. (2009). Agreement in self-other ratings of leader effectiveness: The role of demographics and personality. *International Journal of Selection and Assessment, 17,* 165 – 179.

Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2005). Is there a general factor in ratings of performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology, 90,* 108 – 131.

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

Participant List
(in alphabetical order)

Anna Brown
University of Kent
Discussant

Guangrong Dai
Korn/Ferry International
Presenter

Karen Fuhrmeister
Hogan Assessment Systems
Chair, Co-Author

Keith McCook
Assess Systems
Presenter

Renee Yang
Hogan Assessment Systems
Presenter