

### **360-Degree Feedback: Is It Valid?**

Ben Roselle MA, MBA and Colleen McGinnis, PsyD Candidate

One of the most popular recent advances in leadership development is the 360-degree feedback system. Many organizations use this type of assessment to collect information from various people who can accurately rate the performance of a specific manager with whom they work. In the eyes of most users, the strength in such instruments is their capacity to capture multiple perspectives that most often include manager, self, peers, and direct reports. If your organization has purchased and used a 360-degree instrument, do you know the validity of the feedback?

The results of these surveys provide the basis for important human resource decisions—such as individual development goals, promotion, and training emphases—that, cumulatively, can make or break the success of an organization over time. Most organizations do not use multi-rater feedback for selection or promotion decisions, in large part because the results rely on subjective perspectives, and the instruments are not designed for these purposes. Despite the popularity of these instruments, however, the majority do not report validity data to confirm that they actually measure the underlying leadership factors they purport to assess.

This is the first of a series of three white papers, and it presents an introduction to multi-rater feedback. The next two papers highlight key construct validity themes from a recent factor analysis of the FULLVIEW Feedback Inventory™, developed by Roselle Leadership Strategies, and help you understand and apply these findings to your organization's leadership development efforts.

**What is Multi-rater/360-degree Evaluation?** Although leadership performance historically has been measured through performance appraisals delivered solely by an individual's supervisor, the last three decades have seen the emergence of multi-rater, 360-degree feedback systems. Effective leadership is a complex construct, requiring leaders to master a host of sophisticated cognitive, strategic, and interpersonal skills. Starting in 1967, researchers began to note that using only a single rating source to evaluate leadership might not provide all of the information necessary to evaluate a leader's performance properly.

Since then, relevant research has convincingly demonstrated that a single assessment of a leader, either by self-evaluation or by a supervisor, is inadequate to capture that leader's performance fully. First, individuals are not always the most astute evaluators of their own performance. Self-ratings of any behavior are often widely different in comparison to ratings of that same behavior when completed by another observer (Atwater & Yammarino, 1992). Second, various rating perspectives (i.e., supervisor versus peer, manager versus direct report) actually assess different underlying performance constructs (Turler, 2008). That is, individuals in differing organizational roles have limited opportunities to observe a specific individual's behaviors, so we need multiple perspectives to measure performance accurately. However, this leaves the question of how to interpret the variation in ratings between raters.

**Measuring Validity and Reliability.** Concern about inter-rater agreement focuses on the meaning of low agreement across organizational perspectives. If two perspectives disagree substantially in their ratings, the meaning of that discrepancy remains unclear. Theories range from those claiming that the data are inaccurate or meaningless, to those concluding that differing perspectives supply equally valid data. Tornow (1993) suggested that, "the very differences in perspectives among those who provide feedback can enhance the personal learning that takes place." Therefore, the differences in rater perspectives are *not* treated as error variance (variation that needs to be reduced), but rather as critical additional information that makes the findings more reliable and gives them deeper perspective.

Further, Scullen et al., (2000), hypothesized that observed variations in ratings might reflect actual differences in performance, because a manager is likely to perform differentially in front of diverse groups of people. Specifically, they found that both supervisor and subordinate perspectives capture something unique to those perspectives, but peers do not. They suggest that these rating differences are more a function of true differences in the observed performance than of variations in the observers themselves (bias). Despite the fact that differing perspectives exist on each individual leader, Scullen, Mount and Judge (2003) also have shown that raters across various perspectives share a common conceptualization of a specific leader's overall performance.

Knowing that it is crucial to gather multiple perspectives when attempting to create the most accurate possible picture of performance, and with so many instruments from which to choose, how can you know where to start? According to VanVelsor et al. (1997), authors of these instruments must meet the guidelines of a comprehensive process for evaluating 360-degree instruments. According to them, an author of this type of instrument must:

1. Attempt to identify the full range of behaviors or skills believed to represent leadership competencies.
2. Provide reliability information regarding whether the instrument items cluster in behavioral competencies that are internally consistent, distinct from each other and useful for feedback.
3. Provide validity information about whether the scales actually measure the behavioral dimensions they purport to measure (construct validity).

If your company is currently using some form of multi-rater feedback, did someone scrutinize it for these three features? The majority of multi-rater feedback providers designed and implemented their 360-degree feedback tools with the assumption that they accurately measure the leadership skills necessary for success in a particular organization. They picked items that logically seemed important to leader success, or they evaluated data they collected on competencies that support or undermine leader success. Some used a combination of logic and data collection, but most did not assess them for validity.

In 2008, Roselle Leadership Strategies, Inc. decided to embark on a rigorous validation study of our own 360-degree instrument, the FULLVIEW Feedback Inventory™. Part two of this three-part series on multi-rater feedback tools describes the outcome of our empirical study. The results illustrate why we are so confident when using the and core competencies of the FULLVIEW Feedback Inventory™ in our Development Assessments, our Good Managers to Great Leaders™ workshop series, and our executive coaching engagements.

## Works Cited

- 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.
- Scullen, S.E., Mount, M.K., & Goff, M. (2000). Understanding the Latent Structure of Job Performance Ratings. *Journal of Applied Psychology*, 85, 956-970.
- Scullen, S.E., Mount, M.K. & Judge, T.A. (2003). Evidence of the construct validity of developmental ratings of managerial performance. *Journal of Applied Psychology*, 88, 50 – 66.
- Tornow, W. (1993). Perception or reality: Is multi-perspective measurement a means or an end? *Human Resource Management*, 32, 221-229.
- Turkel, C.C. (2008). Female Leaders' 360-degree self-perception accuracy for leadership competencies and skills. *Dissertation Abstracts*.
- VanVelsor, E., Jean-Brittain, L. & Fleenor, J.W. (1997). *Choosing 360: A Guide to Evaluating Multi-Rater Feedback Instruments for Management Development*. Greensboro, N.C.: Center for Creative Leadership.