

Perspectives

Patterns of Rater Accuracy in 360-Degree Feedback

Robert W. Eichinger, CEO, and Michael M. Lombardo, Director of Research, Lominger Limited, Inc.

For over a decade, we have collected 360 data on men, women, individual contributors, managers, and executives. We have looked at age, how long the rater has known the learner, how often the rater has worked with the learner, confidentiality, the various rater groups (self, boss, direct reports, peers, and customers). Additionally, we have examined how sure raters are of their ratings, what competencies are most overused, and what they find difficult to rate. During this same period we have collected criterion ratings: independent ratings of current and long-term performance (two years out), ratings of potential, stock options, profit measures, and promotion (also two years out). Here, we present our findings about which factors seem to affect rater accuracy in relating to various criterion variables of importance in organizations.

What Has the Least Effect on Rater Accuracy? Gender

Gender research has typically found few significant differences in performance. Women may have to perform slightly better to get to the same level as men and are more participative and attuned to others. Men do slightly better at common business problem-solving skills. Rating biases (women rating women higher and vice-versa) are small to nonexistent. This is not to say that there are no differences or no biases. (For a review of the preceding research, see Lombardo & Eichinger, 2002.)

Many of the so-called differences come from popular surveys and analyses that fail to take into account the type of job and function the male or female is performing. For example, HR jobs typically call for higher interpersonal skills, and more women are in these roles; many line jobs are male-dominated and allow for lower levels of interpersonal skill; therefore, men and women will exhibit different levels of interpersonal skill. Studies that look at men and women in the same companies, in the same jobs, at the same level, with the same amount of managerial experience, find that effectiveness is about equal. Our findings (typically found in scientific research as well) are:

1. Gender is not related to promotion or performance overall or by level.
2. Men and women are not rated differently.
3. Males and females do not rate their gender higher (or lower) than the other.

4. Rating agreement between genders is high: 84% of the competencies were rated the same.
5. Female and male raters agree far more with each other than with the person rated, regardless of gender. They are rating the person, not the person's gender. Both women and men are rated higher by both genders on certain competencies. Women are rated higher on interpersonal skills such as compassion and patience, and on operating skills such as planning. Men are rated higher on some problem-solving, business skills such as strategic agility and technical learning, and on command (crisis) skills. Much of this reflects job differences, with far more women in staff jobs and men in line jobs.
6. There are a few differences, with women rating women higher on team building and motivating others, and on some of the other operating skills like organizing and managing and measuring work. Men rate men higher on competencies such as organizational agility and ambition.
7. Thematically, women get somewhat higher ratings on many interpersonal and operating skills; men on problem solving, business and organizational savvy.
8. The differences are small, and tell us nothing about an individual based on gender.
9. The magnitude of the correlations with performance is the same for both genders (For detail, see Lombardo & Eichinger, 2003).

What Raters Say They Do Not Know or Cannot Rate Clearly

These two responses are identical: The top 10 and the bottom 10 are exactly the same. The top 10 are mostly competencies that indicate the learner is not a manager. Many organizations use 360s with high-potential individual contributors, and measure things like developing direct reports or hiring skills to get a gauge on the person's basic preferences and skills, proclivity to help others, and to size up people's strengths and weaknesses. While this is probably a good practice, it drives up the do-not-know/cannot-rate responses.

Additionally, a few competencies toward the top show what we would expect: The rater does not know the person well enough to rate him or her. Boss relationships, career ambition, and work/life balance are examples of competencies that are harder for many people to see directly. The do-not-know/cannot-rate list is not particularly related to how sure the raters are of their ratings. Only three of the top 10 are the same, so it is not a matter of people checking "do not know" to what they would otherwise give a low rating. The bottom 10 (the least do-not-knows or cannot-rate-clearly: what people are most comfortable rating) are visible or apparent behaviors, such as action orientation or listening.

The use of “do not know or cannot rate” clearly appears to be useful to identify competencies that are not relevant or that are hard to rate because of a lack of familiarity. Neither the hard-to-rate nor the easy-to-rate items are especially related to current or long-term performance. “Do not know or cannot rate” responses provide useful, face-value information. But high or low totals appear to have nothing to do with the accuracy of ratings. The ratings of competencies people rate 99 percent of the time are no more related to performance than those rated much less often.

Age

Older raters (50+) rate learners slightly higher on straightforward operating skills like planning or directing others, but essentially we found few rating tendencies based on age.

Work History

We divided our data into raters who had worked with the person (1) once, (2) more than once, and (3) more than once who had a relationship with the learner off-work. To our surprise, all groups were equally accurate. Learners do get higher ratings on some competencies when raters have an off-work relationship with them. They are seen as higher in integrity, good with customers, smart problem-solvers, better with conflict, and fairer, but this did not affect the overall accuracy of this group. Although this is the only study of this type we are aware of, the initial finding is that one reasonable exposure to a learner is enough to produce accurate ratings. Exposure across settings does not seem to add to accuracy.

Self-Other Agreement by Itself

As we point out later, self-other agreement without additional information is not meaningful or accurate.

What Has the Most Effect on Rater Accuracy?

How Long the Rater Has Known the Person

The “known for one to three years” group is the most accurate in our research ($r=.51$ for both groups). Those who had known the person less than one year just missed significance in the correlation of their ratings with independent measures of performance. The “three-to-five-year” group was significant with performance, but much lower ($r=.31$); and the “known beyond five years” ratings go up, show little differentiation between high and low performers, and are not related to performance. The groups that have known the person one to five years agree most with each other and least with the other groups (long- and short-timers). Knowing the person long enough to get past first impressions, but not so long as to begin to generalize favorably, seems to produce the most accurate ratings.

Confidentiality

As we reported in an earlier issue (see “Should 360 Feedback Be Confidential?”), when raters think their ratings will be or could be revealed, two-thirds of the ratings go up significantly, and, more importantly, become less correlated to performance.

Antonioni’s study (1994) of upward feedback (ratings of the supervisor by direct reports only) found that direct reports whose ratings were not anonymous rated their managers significantly higher than direct reports whose ratings were anonymous. In a study of 58,000 performance appraisals, scores went up significantly in public appraisal processes (Jawahar & Williams, 1997). The title of the article was “When All the Children Are Above Average.” The main reason 360 feedback arose is because it is difficult for peers, direct reports, managers, and executives to engage in straight talk about weaknesses. Giving critical feedback to direct reports face-to-face is ranked 63rd out of 67 (fifth from the bottom) for the typical supervisor, manager, and executive (Lombardo & Eichinger, 2002). That is why most performance appraisals and even anonymous 360s are inflated.

How Sure Raters Are of Their Ratings

On the surface, sureness ratings look transparent. When raters rate higher, they report being more sure. When they rate a competency lower, they report being less sure. The 10 competencies on which raters are surest of their ratings average 12th of 67 in rank order in our norm base (they are sure of their high ratings of those competencies). The 10 on which they are least sure average 54th in rank order out of 67. But when we divide raters into three levels of sureness, and then correlate their competency ratings with performance, sureness has an obvious connection to performance ratings: The surer the rater, the more likely his or her competency ratings correlate with independent performance rating. The relationships with performance are stronger (higher correlations), and there are more significant results on the more complex and harder to-rate-competencies; however, all sureness groups provide some value. Sureness ratings appear to be a valuable piece of information in 360s. Displaying results by sureness level makes sense, and more attention should be paid to raters who are very sure of their ratings.

Who Provides the Ratings

Boss is the most accurate rater, followed by peers and direct reports (who tend to rate somewhat high and undifferentiated between lower and higher performers). Other studies reached similar conclusions (see Atkins & Wood, 2002; Kaplan & Kaiser, 2003). Self is the least accurate (Harris & Schaubroeck, 1988; Conway & Huffcutt, 1997; Lombardo & Eichinger, 2003) along with customers (Lombardo & Eichinger, 2003). They can fairly be called inaccurate, as their ratings did not exceed chance in our studies. In recent studies, peers have been found to overestimate the performance of poor performers (Atkins & Wood, 2002). Antonioni and Park (2001) found that liking someone significantly affected peer and direct-report ratings, but not those of bosses. The effect was stronger the longer the rater had observed the person. Feedback that includes only upward and peer feedback is less accurate. Too much credence is given to these ratings: While they are hardly useless, accuracy can be spotty, and much more attention should be paid to boss ratings.

Self-Other Agreement Coupled with Performance and/or Promotion Data

Without external measures of performance and actual promotion data, it is hard to make sense of rating agreement patterns. Those who are low performers or fired rate themselves higher than others do (Fleenor, et al., 1996; Atwater, et al., 1998; Shipper & Dillard, 2000; Lombardo & Eichinger, 2003); those promoted rate lower than others (Lombardo & Eichinger, 2003), and high performers are either in agreement (Atkins & Wood, 2002; Church, 1997) or lower (Goleman, 1998; Shipper & Dillard, 2000; Lombardo & Eichinger, 2003). One implication of these findings is to *watch out for blind spots*. The fatal pattern in lack of self-awareness is relatively high self-ratings compared with those of others, especially boss. On most 360s with a five-point scale, one point or more should be highly significant. Look for patterns of inflated strength assessment, especially in conflict, perspective, honor, and hands-on management skills. Another implication is to *treat under-rating differently*. In the past and currently, when one underrates, this has often been called a hidden strength or labeled as a lack of self-confidence, self-esteem, or lack of self-knowledge. While any of those is possible, research also suggests that the person may be highly self-critical, have high standards, and be a high performer. A lower self-rating only has meaning once we add the information of how well the person performs. A third implication is that *if the person's ratings are in agreement* with those of others, the person is more likely to be a high performer.

Overusing Competencies

As we argued in a previous issue, overusing competencies can matter greatly. The average overuse of a competency rated as a strength for the learner is 17 percent across our sample of nearly 100,000 raters. (See Exhibit 1 for detail.) Few 360s collect this information, with Kaplan and Kaiser (2003) the only other source we could locate. The authors found even higher levels of overuse than we did. Here we argue that learners are missing a critical piece of data by not being told what they overdo. Their feedback would be more accurate if overuse

ratings were included. Knowing yourself is still the best option—strengths, averages, weaknesses, and untested areas.

Successful people do not have all possible skills but they tend not to have blind spots either. Their edge is in fully knowing themselves: developing where they can, enhancing average areas that become more important, and neutralizing weak areas where they will never be strong. Focus on strengths, but do it with some caution and perspective. Is it a good idea to focus on and leverage strengths? Of course. It accounts for much of our success. Each person should make a list of the corollary weaknesses that typically accompany these specific strengths (Lombardo & Eichinger, 2002). Make it an objective to get feedback on those specific weaknesses over time and attempt to neutralize the ones that begin to matter. In times of transition, overdone strengths can lead to derailment. At any time they can hamper performance (Lombardo & Eichinger, 2003; McCall, et al., 1988; Morrison, et al., 1987; Shipper & Dillard, 2000).

Managing diversity is the least overused and the most overused strength. Seventy-eight percent of raters respond "Not at All" to the overuse question and 13 percent respond "Constantly." Their answer is seldom in the middle. Over 50 percent of raters say these three competencies are overused at least occasionally: action-oriented, command skills, managerial courage.

References

- Antonioni, D. (1994). "The Effects of Feedback Accountability on Upward Appraisal Ratings." *Personnel Psychology*, 47: 349-356.
- Antonioni, D., & Park, H. (2001). "The Relationship Between Rater Affect and Three Sources of 360-Degree Feedback Ratings." *Journal of Management*, 27: 479-495.
- Atkins, P.W.B., & Wood, R.E. (2002). "Self Versus Others' Ratings as Predictors of Assessment Center Ratings." *Personnel Psychology*, 55.
- Atwater, L.E., Ostroff, C., Yammarino, F.J., & Fleenor, J.W. (1998). "Self-Other Rating Agreement: Does It Really Matter." *Personnel Psychology*, 51: 576-597.
- Church, A.H. (1997). "Managerial Self-Awareness in High Performing Individuals in Organizations." *Journal of Applied Psychology*, 82: 281-292.
- 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.
- Fleenor, J.W., McCauley, C.D., & Brutus, S. (1996). "Self Other Agreement and Leader Effectiveness." *The Leadership Quarterly*, 7(4): 487-506.
- Goleman, D. (1998). "What Makes a Leader?" *Harvard Business Review* (November-December).
- Harris, M.M. & Schaubroeck, J. (1988). "A Meta-Analysis of Self-Supervisor, Self-Peer and Peer-Supervisor Ratings." *Personnel Psychology*, 41: 43-62.
- Jawahar, I.M., & Williams, C.R. (1997). "When All the Children Are Above Average: The Performance Appraisal Purpose Effect." *Personnel Psychology*, 50(4): 905-925.
- Kaplan, R.E., & Kaiser, R.B. (2003). "Rethinking a Classic Distinction in Leadership: Implications for the Assessment and Development of Executives." *Consulting Psychology Journal: Research and Practice*, 55: 15-25.
- Lombardo, M. & Eichinger, R. (2003). *Leadership Architect Norms and Validity Report*. Minneapolis: Lominger.
- Lombardo, M. & Eichinger, R. (2002). *The Leadership Machine*. Minneapolis: Lominger.
- McCall, M., Lombardo, M. & Morrison, A. (1988). *The Lessons of Experience*. Lexington, MA: Lexington Books.
- Morrison, A., White, R., & Van Velsor, E. (1992). *Breaking the Glass Ceiling*. Reading, Mass: Addison Wesley.
- Shipper, F., & Dillard, J. (2000). "A Study of Impending Derailment and Recovery of Middle Managers Across Career Stages." *Human Resource Management*, 39(4): 331-347.

EXHIBIT 1

Overused Competencies

Most Overused Competencies

- Personal Learning
- Career Ambition
- Confronting Direct Reports
- Informing
- Managing and Measuring Work
- Action-Oriented
- Hiring and Staffing
- Drive for Results
- Self-Knowledge

Themes in Overuse

- Too quick on the trigger
- Too wishy-washy and changeable
- Too critical
- Too selfish a self-marketer
- Overly controlling
- Prefers a class or type of people

Copyright of Human Resource Planning is the property of Human Resource Planning Society and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.