Consulting Psychology Journal: Practice and Research
2012, Vol. 64, No. 3, 157–182

1065-9293/12/$12.00 DOI: 10.1037/a0030011

EVIDENCE-BASED ANSWERS TO 15 QUESTIONS ABOUT LEVERAGING 360-DEGREE FEEDBACK

Kenneth M. Nowack and Sandra Mashihi
Envisia Learning, Santa Monica, California

Despite the popularity of 360-degree feedback, meta-analytic findings suggest that these interventions can lead to a significant change in behavior but the effect sizes are typically modest and when done poorly may lead to both disengagement and a decline in performance. The research evidence addressing practical issues for coaches to successfully implement 360-degree feedback interventions is updated since previous review studies (e.g., Craig & Hannum, 2006; Fleenor, Taylor, & Craig, 2008). This article reviews 15 specific questions that are common to most 360-degree feedback interventions (purpose and goals, methodology and psychometric properties, and process and implementation) designed to facilitate enhanced awareness and successful behavior change in individuals and teams.

Keywords: 360-feedback, multirater feedback, feedback, coaching, behavioral change, talent development, leadership development

Increasingly, 360-degree feedback systems have proliferated and are being used for diverse purposes and interventions (e.g., executive coaching, performance evaluation, talent management, and succession planning). Despite the widespread use of 360-degree feedback, coaches and consultants still seem to ignore some of the potential issues and evidence-based research highlighting the possible limitations, risks, and issues of this type of intervention for coaching and talent development (Bracken & Rose, 2011). Under the right circumstances, feedback interventions can facilitate some of the conditions required for successful behavioral change (Mashihi & Nowack, 2011), yet there are many studies showing that such processes sometimes create no measurable change whatsoever (Siefert, Yukl, & McDonald, 2003), small effects (Atwater, Waldman, Atwater, & Cartier, 2000), or may have negative effects on both engagement and productivity (Kluger & DeNisi, 1996).

The current practice and use of 360-degree feedback by coaches and consultants is often based on expert opinion, recommendations from vendors, or fads, rather than on evidence-based empirical findings or applied evaluation studies. In fact, there is a paucity of well-designed longitudinal research and evaluation studies to guide coaches in the effective design, administration, reporting, interpretation, and use of 360-degree feedback systems for initiating and sustaining new behavioral change over time (Nowack, Hartley, & Bradley, 1999). The current literature is challenging to interpret because of the use of diverse and nonstandardized competency models and definitions, different purposes and goals of the feedback process, use of 360-degree feedback across multiple job

Kenneth M. Nowack and Sandra Mashihi, Envisia Learning, Santa Monica, California.
Correspondence concerning this article should be addressed to Kenneth M. Nowack, Envisia Learning, 3435 Ocean Park Boulevard Suite 203, Santa Monica, CA 90405. E-mail: ken@envisialearning.com

levels, cross-cultural factors in comparing self-other scores, diverse response scales, and measurement issues inherent within the actual assessments being used (Caputo & Roch, 2009; English, Rose, & McLellan, 2009; Roch, Sternburgh, & Caputo, 2007).

In addition to the limited number of well-designed prospective studies showing the benefits of 360-degree feedback, there are other studies that suggest potential harm, danger, and potential limitations of its impact on both awareness and effectiveness. Despite the possible limitations of 360-degree feedback, coaches can leverage this type of intervention to maximize both awareness and behavioral change by understanding and using comprehensive feedback and individual change models which build on the theoretical work of others (Gregory, Levy, & Jeffers, 2008; Joo, 2005; London & Smither, 2002; Nowack, 2009).

Leveraging 360-Degree Feedback Based on Empirical Evidence

Trying to distill the 360-degree feedback literature into evidenced-based practice is challenging, but coaches are encouraged to review some of the earlier recommendations by Bracken, Timmreck, Fleenor, and Summers (2001), who discussed proximal and distal factors related to successfully implementing 360-degree feedback, Morgeson, Mumford, and Campion (2005), who organized 360-degree feedback research into 27 questions that focus on practical applications, Craig and Hannum (2006), who summarized relevant research findings up to that point, and Fleenor, Taylor, and Craig (2008), who wrote about "best practices" in using 360-feedback for behavior change.

Purpose and Goals	Methodology and Psychometric Properties	Process and Implementation
1. Does 360-degree feedback do more harm than good?	3. What type and how many raters should be included?	9. Can open-ended questions be emotionally damaging to clients?
2. Under what conditions and for whom does 360-degree feedback become beneficial?	4. Do ratings between rater groups agree with each other?	10. Does personality impact how people respond to 360-degree feedback?
	5. Do ratings within rater groups agree with each other?	11. How do you manage the feedback of underestimators and overestimators?
	6. Which response scale is best for 360-degree feedback?	12. What kind of training or certification is required by coaches to help clients understand and interpret 360-degree feedback reports?
	7. How many rating points should be on a 360-degree feedback scale?	13. Are there cultural differences to be considered in the use of 360-degree feedback?
	8. Should a 360-degree feedback report contain a mix of graphs, charts and responses to open-ended questions to maximize understanding?	14. Does 360-degree feedback require debriefing?
		15. How can you leverage the impact of 360-degree feedback to ensure successful behavior change?

Figure 1. Fifteen key questions about leveraging 360-degree feedback.

The intent of this article is to update prior "best practices" suggestions for coaches on 15 important questions (see Figure 1) that typically arise on the use of 360-degree feedback (Nowack, 1999). In particular, this article will update the expanding literature on questions addressed by Morgeson et al., (2005) and others, and will cite approximately 25 newer evidence-based research studies since 2008. These questions will address issues that arise in the development, implementation, interpretation, and leveraging of 360-degree feedback interventions for individual, team, and organizational effectiveness particularly in a multinational and global environment.

The first two questions addressed will focus on the *Purpose and Goals of Using 360-Degree Feedback* and whether it is always the best intervention to use for coaching, training, compensation and talent management/succession planning programs. The second set of questions will focus on *Methodology and Psychometric Properties*. The third section reviews the *Process and Implementation* of the 360-degree feedback process from selecting raters, confidentiality, and the impact of negative open-ended questions often included in most assessments. This last section will also address how to leverage the impact of 360-degree feedback as well as how to evaluate the intervention's effectiveness.

Purpose and Goals of 360-Degree Feedback

1. Does 360-Degree Feedback Do More Harm Than Good?

In general, poorly designed 360-degree feedback assessments and interventions can increase disengagement and contribute to poor individual and team performance (Ilgen & Davis, 2000; Kluger & De Nisi, 1998). In one commonly cited meta-analysis on performance feedback (607 effect sizes, 23,663 observations), Kluger and DeNisi (1996) found that although there was a significant effect across all studies for feedback interventions ($d = .41$), performance actually declined in one third of all studies analyzed for various reasons such as depth of the feedback process, how feedback was delivered, and personality of the recipient.

Several studies have also shown that individuals can experience strong discouragement and frustration when 360-degree feedback is not as positive as they expected (Atwater & Brett, 2005). Brett and Atwater (2001) found that managers who rated themselves higher than others (overestimators) reported significantly more negative reactions to the 360-degree feedback process. They noted specifically that "negative feedback (i.e., ratings that were low or that were lower than expected) was not seen as accurate or useful, and it did not result in enlightenment or awareness but rather in negative reactions such as anger and discouragement" (p. 938).

Newer neuroscience research sheds some interesting light on why negative feedback is potentially emotionally harmful. In general, stressors that induce greater social-evaluative threat elicit significantly larger cortisol and ambulatory blood pressure responses (Dickerson & Kemeny, 2004; Lehman & Conley, 2010). Recent studies confirm that emotional hurt and rejection, whether part of social interactions or poorly designed and delivered feedback interventions, can actually trigger the same neurophysiologic pathways associated with physical pain and suffering (Eisenberger, Lieberman, & Williams, 2003).

In two follow-up studies by DeWall et al., (2010), functional MRI was used to test whether a physical pain suppressant reduced behavioral and neural responses to social rejection. Their two studies confirmed that acetaminophen, relative to a placebo control, significantly reduced behavioral and neural responses associated with the pain of social rejection providing additional evidence of the substantial overlap between social and physical pain. Current findings also suggest that people report higher levels of self-reported pain and have diminished performance on a cognitively demanding task after reliving a past socially meaningful but painful event more than a past physically painful event (Chen, Williams, Fitness, & Newton, 2008).

Finally, research on individual positive psychological well-being (Schwartz, Reyonolds, Thase, Frank, Fasiczka, & Haaga, 2002), success in marriage (Gottman, 1994), and team effectiveness (Losada & Heaphy, 2004) suggest that the ratio of positive-to-negative emotions and interactions is of critical importance. For example, Losada and Heaphy (2004) unobtrusively observed actual work teams working on strategic planning tasks and coded all interpersonal interactions as positive (e.g.,

demonstrations of support and encouragement) or negative (e.g., cynicism and disapproval of others).

They identified 15 flourishing teams defined as showing uniformly high performance across three indicators: profitability, customer satisfaction, and evaluations by superiors, peers, and subordinates. Other teams were categorized as mixed ($n = 26$) or low performers ($n = 19$). Losada and Heaphy (2004) found that the optimally performing teams demonstrated an approximate 3:1 positive-to-negative ratio of interpersonal interactions, but performance decreased at 11.9:1 (i.e., teams became more dysfunctional and less productive suggesting a possible upper limit of positive-to-negative interactions). Most commonly, a ratio of 3:1 positive-to-negative interactions appears to be significantly associated with enhanced individual and team performance, individual engagement, emotional flourishing, and effectiveness (Fredrickson & Losada, 2005).

Implications. Neurobiological research hints that perceptions around status, certainty, autonomy, social relationships, and fairness (Rock, 2008) can possibly derail a 360-degree feedback process and create emotional stress in clients that can potentially interfere with insight, acceptance, and initial motivation to change behavior. The positivity-to-negativity ratio studies mentioned earlier are important to consider in light of how clients experience and interpret ratings and comments from others in a 360-degree feedback process. In particular, a preponderance of negative (vs. positive) feedback messages can interfere with both proximal (insight and motivation) and distal goals (sustained deliberate practice and overall effectiveness) in the coaching engagement.

2. Does 360-Degree Feedback Work (for Whom and Under What Conditions)?

Among researchers and coaches, there is little disagreement that under the right conditions and applying evidence-based "best practices" that 360-degree feedback can increase self-awareness and increase individual and team effectiveness (Atwater & Brett, 2006; Fleenor, Taylor, & Craig, 2008; Reilly, Smither, & Vasilopoulos, 1996). At this point, there appears to be a need for even more in-depth prospective studies that allow for a more complete evaluation of the impact of 360-feedback interventions along with potential limits of behavioral change. An earlier meta-analysis of 26 longitudinal studies by Smither, London, and Reilly (2005) suggests that 360-degree feedback does lead to significant improvements on both perceptions of improved performance and actual behavioral change and some insights about the conditions required to ensure success.

In their meta-analysis, Smither et al., (2005) examined the mean unweighted (and uncorrected) effect sizes and compared them with the mean weighted (and uncorrected) effect sizes. The mean unweighted (and uncorrected) effect sizes were .24, .12, .14, and .00 for direct report, peer, supervisor, and self-ratings, respectively. The mean weighted (and uncorrected) effect sizes were .12, .04, .10, and .03 for direct report, peer, supervisor, and self-ratings, respectively. Across rater sources (excluding self-ratings), the average effect size in the developmental purpose studies was .25 versus .08 in the administrative purpose studies.

It is important to note that 15 of the 24 studies included in the meta-analytic calculated only a single score (i.e., the average rating across all items). Smither et al. (2005) also noted whether the study involved only upward feedback (i.e., only from direct reports) versus feedback from multiples sources (i.e., direct reports, peers, supervisor) as well as being used for developmental purposes or for personnel decisions (e.g., promotion decisions, performance reviews). The authors point out that these small effect sizes might be influenced by averaging ratings across all items at Time 1 versus Time 2 (over half of the studies in their meta-analysis). They suggest a shift to evaluating change in goal progress or behavioral effectiveness might be more appropriate to measure the true impact of 360-degree interventions.

These findings suggest that expected performance improvements may be practically modest for even those most motivated and capable of changing behavior over time (Smither et al., 2005). Taken together, there is supporting evidence that feedback is a necessary and important condition for successful behavioral change and most useful for those clients with moderate to low levels of self-insight, or who express a strong motivation to improve, demonstrate poor performance on

teachable skills, and have a learning versus performance goal orientation (Leonardelli, Herman, Lynch, & Arkin, 2003).

Implications. Smither et al., (2005) presented eight important factors that play a role in determining the extent of behavioral change and performance improvement after 360-degree feedback interventions. These factors help answer questions related to for whom and under what conditions feedback can be most beneficial and impactful. The eight factors include the following: (1) the delivery and content of the feedback; (2) interpretations and emotional reactions to feedback; (3) the personality of the participant; (4) feedback orientation of the participant; (5) readiness to change, (6) beliefs about change as well as self-esteem and self-efficacy; (7) goal intentions versus implementation intentions; and (8) taking/sustaining action while managing possible relapse.

It appears that these factors would be useful for coaches to consider and for researchers to continue to focus on improving the efficacy of feedback interventions for individuals, teams, and organizations. For example, coaches who can identify their client's readiness for change level and evaluate important personality variables related to acceptance of the feedback, achievement orientation, and openness to change (e.g., conscientiousness, emotional stability) will be better able to tailor their approach to discussing the feedback results and setting development goals. Additionally, coaches should help define, clarify, and focus client development goals and translate them to implementation intentions to maximize successful behavioral change and help strategize possible coping strategies to minimize potential relapse (Mashihi & Nowack, 2011).

Methodology and Psychometric Properties

3. What Type and How Many Raters Should Be Included?

The type of raters who provide feedback to clients will depend on a number of factors, including the purpose of the 360-degree feedback process, the job level of the client, the competencies being assessed, and the relevant stakeholders who had an opportunity to provide constructive feedback. In general, individual-based 360-degree feedback processes typically include the client's manager, direct reports, team members (different job levels), peers (same job level), internal customers, or others outside the organization. Each rater type appears to provide unique and meaningful information and may become a focal point for developmental actions (Fleenor et al., 2008).

How many raters are necessary to provide meaningful and accurate 360-degree feedback? The answer, of course, is only one rater, but there is no way to know who is "all-knowing" and perfectly accurate in his or her observations. The following analogy demonstrates this. Sit down with a child to solve a puzzle and ask this question; "How many puzzle pieces do you need to assemble to have confidence that what you are making resembles the picture on the cover?" Think of this as a way to determine the number of raters needed in 360-degree feedback. The more puzzle pieces we assemble correctly, the more confident we become that we are seeing the image that is also on the cover of the box. There is no need to assemble all of the pieces to verify this. All that is needed is a "critical mass" of puzzle pieces of the assembled puzzle pieces to be confident of the true picture of the puzzle.

When clients ask raters for feedback, it is hoped that they have a clear and accurate picture of how they are behaving and being perceived by others. In fact, there is some research that suggests what this "critical mass" of feedback is to reach a level of confidence that others are accurately experiencing the client's behavior and can identify signature strengths and development opportunities. Greguras and Robie (1995) suggested that the optimum number of raters involved in most 360-degree feedback projects requires at least four supervisors, eight peers, and nine direct reports to achieve acceptable levels of reliability (.70 or higher). Of course, this statistical standard may not be practical in circumstances in which leaders have only a few direct reports or when the input of only the current manager is desired.

Implications. Recent research suggests that when two or fewer respondents provide data for a given group, this small number of rater responses may be inadequate for reliable measurement (3D Group, 2009). Inviting more, rather than fewer, raters would be helpful in ensuring accuracy and a large enough rater pool to make the 360-degree feedback findings relevant and useful. Inviting and

having too few raters in each rater category may limit the meaningfulness and accuracy of the feedback for professional and personal development.

At least one study (Nieman-Gonder, Metlay, Kaplan, & Wolfe, 2006) has explored ratings provided by selected and nonselected raters by clients using multiple accuracy measures. Results indicated that selected raters were as accurate, or more accurate, than raters who were not selected by the client. Therefore, having a critical mass of feedback is essential whether raters are nominated by the client and without input from others or selected directly by human resources or the client's manager. Ideally, the selection of both the number and type of raters should be a participative process between the client and his or her manager to optimize acceptance of the feedback results (Bracken & Rose, 2011).

4. Do Ratings Between Rater Groups Agree With Each Other?

There is an extensive literature on the relationship between rater reliability and job performance that has implications for expecting and interpreting differences between rater groups in 360-degree feedback (Le, Oh, Shaffer, & Schmidt, 2007; Murphy, 2008; Murphy, Cleveland, & Mohler, 2001). Whereas high levels of interrater reliability are necessary for adequate measurement in most performance evaluation systems, 360-degree feedback interventions are based on the assumption that raters from different levels provide unique and meaningful information (Lance et al., 2008). From this perspective, some degree of cross-source disagreement is actually desirable, and source effects are not necessarily an indicator of poor quality ratings (Hoffman & Woehr, 2009). For example, Wanguri (1995) found that multiple rater appraisals improved rating accuracy and perceptions of fairness in a meta-analysis of 113 empirical studies on performance evaluations.

Some research suggests that findings of low levels of agreement across rating sources may be largely spurious. Specifically, Libretto, Burgess, Kaiser, Archly, and James (2003) present a convincing case that estimates of interrater agreement based on intraclass and Pearson correlations are severely attenuated because of restriction of range in job performance and thus represent substantial underestimates of interpreter agreement. On the other hand, Scullen, Mount, and Goff (2000), hypothesized that observed variations in ratings might reflect actual differences in performance because an employee is likely to perform differentially in front of diverse groups of people. They suggest that rating differences are more a function of true differences in the observed performance than of variations in the observers themselves (bias). Their study found small perspective-related effects were observed in boss and subordinate ratings, but not in peer ratings.

Other researchers also support the use of multiple independent raters who have had opportunity to observe ratee performance and pooling of performance ratings across raters to improve reliability of rating scores on which organizational decisions may be based (Murphy, 2008; Ones, Viswesvaran, & Schmidt, 2008; Viswesvaran, Schmidt, & Ones, 2002, 2005). These findings suggest that different rater sources, including self-ratings, represent valid performance-related variance and most likely are not mere measurement method bias (Conway, Lombardo, & Sanders, 2001; Hoffman, Lance, Bynum, & Gentry, 2010). It is also important to note that interrater agreement and interrater reliability can but do not necessarily coexist (Liao, Hunt, & Chen, 2010). The presence of one does not ensure the other. Interrater agreement and interrater reliability are both important for the accuracy and usefulness of performance evaluation. The former indicates stability of ratings an employee receives from different raters, whereas the latter shows the consistency of ratings across different employees from different raters.

In general, self-ratings have been found to be modestly correlated with other rater perspectives (Pearson correlations .3 to .6), with a greater convergence between peer and supervisor ratings (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988; Nowack, 1992, 2002). Cumulative evidence suggests that supervisors are the most reliable source of job performance ratings (Conway & Huffcutt, 1997) and their ratings are more strongly associated with performance as measured by external criteria (e.g., promotions, salary) than are ratings from other sources (Atkins & Wood, 2002; Beehr, Ivanitskaya, Hansen, Erofeev & Gudanowski, 2001). Findings from Viswesvaran, Schmidt, and Ones (2005) also suggest that supervisory ratings of overall job performance are empirically related to other measurements of job performance and validity coefficients for promi-

nent predictors used in personnel selection or promotion are comparable for supervisory ratings and other performance indicators. In general, supervisors tend to focus more on performance-oriented behaviors compared with direct reports who tend to emphasize and filter interpersonal and relationship behaviors into subjective ratings (Nowack, 2009) and are more likely to be recalled by participants (Smither, Brett, & Atwater, 2008).

Despite some mixed evidence (e.g., Sala & Dwight, 2002), peers are able to discern future leadership potential (Nowack, 2009), leadership efficacy (Chemers, Watson, & May, 2000; Hannah, Avolio, Luthans, & Harms, 2008), accurately evaluate job performance (Law, Wong, & Song, 2004), and are particularly sensitive to negatively evaluate the personality trait of narcissism in others (Goffin & Anderson, 2007). For example, Inceoglu and Externbrink (2012) collected data for 151 international managers from a global Fortune 500 company (consumer goods sector) who participated in an internal leadership program. Results showed that assessment center (AC) ratings correlated positively with 360-degree feedback ratings for the same competency but only if rated by peers. Overlap between 360-degree feedback and overall AC ratings made by independent assessors in this managerial sample shows that peers may have a more accurate perspective of a participant's performance compared with subordinates or managers. In their analysis of leadership effectiveness with 74 executives using 360-degree feedback, Harris and Kuhnert (2007) reported that peers' evaluations added incremental information distinct from superiors in predicting overall leadership effectiveness.

These meaningful rater group differences might also be a point of possible confusion in the interpretation of 360-degree assessments for clients trying to use these results to determine specific behaviors to modify for some or all rater groups (e.g., their boss and/or direct reports). This potential ambiguity in understanding and interpreting 360-degree feedback is important in light of recent research suggesting that people who are even mildly neurotic report more distress by uncertainty within oral and written feedback than when given very direct negative feedback (Hirsh & Inzlicht, 2008).

Implications. At a practical level, clients might be challenged to understand how to interpret important differences observed by rater groups and to decide whether to focus development efforts on managing upward, downward, or laterally in light of potentially discrepant results. Coaches should be cognizant of the moderate correlations between different rater groups and help their clients to fully understand and interpret the meaning of such differences.

For example, if a client receives a report showing low ratings from both the client's manager and peers but much higher ratings from direct reports, the coach might explore the current relationship with the client's boss as well as the meaningfulness of the discrepancies in ratings by those they directly supervise and work with at a peer level. Because peers seem to focus on "followership" traits such as self-esteem, ability to accept feedback, and confidence in their ratings (Goffin & Anderson, 2007), these coworker relationships might need to be nurtured and strengthened for clients who have aspirations for seeking future leadership opportunities within the organization. Coaches should also be familiar with the findings of Brett and Atwater (2001) who reported that participants in their 360-degree feedback program had stronger emotional reactions to negative feedback from bosses and peers than they did to negative feedback from direct reports.

It appears that almost all vendors who generate 360-degree feedback reports present various ways to compare and contrast self versus other rater responses (e.g., graphs or tables showing average scores across competencies and specific behaviors). However, it is typically up to the client to discern the meaning of these differences and what actions, if any, should be considered as a result of the feedback they received. Correctly interpreting and acting on these perceptual differences by rater groups is both an opportunity and challenge inherent in 360-degree feedback interventions.

5. Do Ratings Within Rater Groups Agree With Each Other?

In an earlier meta-analytic study of 360-degree feedback performance ratings by Conway and Huffcutt (1997), the average correlation between two supervisors was only .50, for peers, it was .37, and between two subordinates it was only .30. Until a newer meta-analytic review is conducted, it appears that agreement within rater groups appears to be an important issue to discern for clients in the interpretation of their 360-degree feedback reports.

Several explanations have been offered as to why different raters of the same individual may provide discrepant ratings including: (1) raters selectively focus on different aspects of an individual's competence, personality, and/or performance; (2) raters have different opportunities to actually observe and experience the behavior of others (e.g., sampling bias); (3) raters attribute different levels of importance to the same observed behavior influencing the way they make appraisals of others; and (4) the linguistic characteristics of the actual questions in 360-degree assessments influence rater appraisals (Lebreton, Burgess, Kaiser, Atchley, & James, 2003; Kaiser & Craig, 2005).

For example, using an archival database of responses (n = 737), a study by Dai, Stiles, Hallenbeck, and De Meuse (2007) found that when a leadership competency was abstract, the agreement between self and others was lower than if the competency was concrete. Kaiser et al. (2005) argued that vendors and companies developing 360-degree feedback assessments should pay special attention to the specific questions being asked (e.g., behaviorally specific, placed in a proper context and not double barreled) to enhance interrater reliability.

Vendors who do not provide a way for clients to evaluate within-rater group agreement in feedback may increase the probability that average scores used in reports can be easily be misinterpreted—particularly if they are used by coaches to help clients focus on specific competencies and behaviors for developmental planning purposes. It is not unusual for clients to expend a great deal of energy trying to identify their potential "critics" and "supporters" in discussing the results of the 360-degree feedback assessment when there is discrepancy of ratings within rater categories (e.g., direct reports). Having a way to discern and discuss potential "outliers" in the data can help clients be more focused on their developmental goals and make informed choices about how much energy to put into "managing" their direct reports, peers, or boss.

Implications. From a practical perspective, vendors should provide one or more measures of rater agreement within each individual feedback report such as the following: (1) Including a range of scores on self-other rating summaries; (2) Showing a distribution of ratings on most and least frequent behavior summaries in a manner that ensures confidentiality; and (3) Including statistical metrics of rater agreement (e.g., based on standard deviation or standard error). All of these within-rater group agreement metrics would appear to help delineate potential outliers and clarify how to possibly interpret polarized scores at both a competency and question level.

It is easy for clients to review a 360-degree feedback report including a table ranking the most or least frequently observed behaviors (e.g., based on the calculation of average scores by rater groups) and immediately interpret specific behaviors on which to focus their development efforts. Clients often do so without an analysis of whether the average scores might truly reflect rater "clans" of supporters and critics that require additional follow-up to clarify what the average scores may mean. Having at least one approach in 360-degree feedback reports to measure and evaluate within-rater agreement, without breaking confidentiality, would appear useful for accurately interpreting results and facilitating meaningful development planning. When there are only a few raters who provide feedback to a client, the possibility of an "outlier" magnifies the importance of interpreting average scores at the item level cautiously.

6. Which Response Scale Is Best for 360-Degree Feedback?

Many studies suggest that response scales have an impact on the 360-degree feedback data, and some response scales seem to be preferable to others (Cools, Hofmans, & Theuns, 2006; Heidemeier & Moser, 2009; Viswanathan, Bergen, Dutta, & Childres, 1996). For example, Bracken and Rose (2011) suggested that commonly used frequency scales (e.g., "never" to "always") are inferior to others (e.g., satisfaction or effectiveness) because of lack of variability in the responses, but are quick to point out that the majority of research has focused on the anchors themselves and that additional research is needed to identify optimal response choices and/or anchor format.

A recent meta-analysis by Heidemeier and Moser (2009) suggested that social comparison scales (scales with relative rather than absolute anchors) were able to reduce leniency in self-ratings and should be employed much more often than in the past. Goffin and Olson (2011) also presented evidence from at least three important and quite different domains that comparative evaluative judgments of the self or others (i.e., whether a given person is higher or lower on some characteristic

than is another specific person) may be more advantageous than absolute evaluative judgments (i.e., asking the respondent to indicate the individual's level of performance, attitude, traits, or other attributes using numerical scales with verbal anchors such as low to high, unfavorable to favorable, bad to good).

Hoffman et al. (2012) explored a new method of presenting items in 360-degree feedback assessments called frame-of-reference scales (FORS) in both a laboratory and field-based study. Drawing from previous approaches to improve the quality of performance ratings, Frame of Reference Scales (FORS) add definitions and examples of effective and ineffective behaviors to a set of behavioral items associated with each dimension or competency being assessed (i.e., provides behavioral examples on the scale itself with a few critical anchors). They found that FORS were associated with a higher pattern of competency factor loadings, less overlap, and decrease of error in measurement (Hoffman et al., 2012). The overall impact of using this FORS scaling approach in both studies was found to be moderate in overall effect size but a significant improvement over the use of traditional rating scales and, as such, deserves additional research attention.

Holt and Seki (2012) encouraged vendors and coaches to explore alternative assessment approaches and scaling that will be more culturally sensitive to defining, measuring, and evaluating leadership competence in light of the use of 360-degree feedback within multinational companies. They encouraged consideration of alternatives beyond frequency-based scales that might emphasize perceptions of overdoing or underdoing behaviors (e.g., Kaiser & Overfield, 2011) and bipolar scales measuring strengths and the overuse of those strengths.

Gentry and Eckert (2012) suggested that using a dual scale of leadership expectations (ratings on the dimensions that contribute or hinder leadership in general) and perceptions (actual ratings of behavior) and exploring the alignment between these two might also improve and enhance cross-cultural measurement and development of clients receiving feedback. They argued that this method differs from traditional 360-degree measurement approaches in that perceptions per se, are neutral and that meaning is dependent on the context of "local" expectations. Therefore, the "fit" between expectations and perceptions helps coaches interpret the results with their client. They also pointed out some limitations of this approach such as overload of data to interpret for the client and how expectations of raters may shift over time (e.g., changing roles or styles of leaders required to remain competitive in the marketplace).

Implications. A review of the literature suggests that there are diverse approaches to the use of response scales in 360-degree feedback assessments (e.g., comparison vs. absolute). In general, response or rating scales are important to consider when developing customized 360-degree feedback assessments and interpreting off-the-shelf tools available from vendors.

People usually define their strengths based on traits they already possess and define their developmental opportunities more in terms of personality and abilities they lack at the moment (Dunning, Heath, & Suls, 2004). Kaiser and Overfield (2011) found that leaders were five times more likely to overdo behaviors related to their strengths, and they argue that most 360-degree feedback assessments do not "distinguish doing a lot of a behavior from doing too much of it, or distinguish underdoing from overdoing as two distinct classes of performance problems" (pp. 105–106).

The selection and use of 360-degree feedback scales should be properly matched to the purposes of their use (e.g., coaching/development emphasizing what clients can translate into deliberate practice, selection/succession planning which typically depends on comparing clients ratings to others or performance evaluation which may require a mix of the two to facilitate both appraisal of performance and future development). Coaches and researchers should continue to explore and investigate the use of alternative 360-degree scales to maximize cross-culture relevance and meaningful interpretation of results. Finally, coaches using 360-degree feedback assessments should also consider providing client and rater training which is generally accepted as an effective method to enhance the psychometric soundness of ratings (Hoffman & Baldwin, 2012; Woehr, 2008).

7. How Many Rating Points Should Be on a 360-Degree Feedback Scale?

There is no definitive agreement from researchers or vendors about the optimal number of response categories that should be used to get the most reliable data in 360-degree assessments. However, there is a general range provided in the broader survey literature that suggests the number of response categories to consider using (Cicchetti, Showalter, & Tirer, 1985; Cools et al., 2006; Dawes, 2008; Weng, 2004; Wikman & Warneryd, 1990). The most recent findings from a 2009 benchmark study (3D Group, 2009) suggest that the most popular in practice is a five-point scale (76%) followed by a seven-point rating scale (16%).

For example, Bandalos and Enders (1996) found that the reliability was highest for scales having five to seven points. Preston and Colman (2000) examined response categories ranging from two to 11 and reported that test–retest reliability was lowest for two- to four-point scales and was highest for seven- to 10-point scales (there was a noted decrease beyond 10 response categories). In a direct comparison of the reliability and validity of four-point and six-point Likert scales, Chang (2005) found that criterion-related validity was not affected by the number of scale points but reliability was higher using the four-point scale.

Research by Lozano, Garcia-Cueto, and Muniz (2008) also investigated the reliability and validity of scales ranging from two to nine response options with four different sample sizes. Their analyses suggested that having between four and seven response options was optimal. It is important to note that the number of rating points may be largely dependent on the type of scale being used (e.g., frequency, effectiveness, comparison) and/or a selection of using a curvilinear approach such as the "too little versus too much" frequency scale discussed by Kaiser and Overfield (2011).

Implications. Newer research confirms that the "sweet spot" for 360-degree feedback response scales is between four and seven for both commercially and organizationally developed assessments (3D Group, 2009). Current research suggests that short (less than three) and long (greater than seven) response scales are generally less reliable and popular in practice.

Coaches should make sure that the response scale is appropriate for the purpose of the 360-degree feedback intervention (e.g., development vs. personnel decisions) and understand the implications for interpretation of results when the full range of the response scale is not used by most raters (e.g., negative skewness). For example, some evidence suggests that standard deviations are larger across all competencies with the use of positively worded response scales (i.e., those that include more positive than negative labels as rating descriptors) in 360-degree feedback assessments (English, Rose, & McClellan, 2009). The use of positively worded response scales might be useful to increase variability in rater responses, increasing accuracy and making interpretation of feedback results more meaningful for clients.

8. Should an Individual 360-Degree Feedback Report Contain a Mix of Graphs, Charts, Tables, and Responses to Open-Ended Questions to Maximize Understanding?

Little research exists to provide definitive answers as to the best way to present 360-degree feedback results to facilitate acceptance and enhance readiness to change behavior. However, it is intuitive that clients have different learning styles, and some may prefer to favor the interpretation of either qualitative over quantitative presentations of results or vice versa. One study that does give some insight about what presentation style might maximize the acceptability, understanding, and interpretation of 360-degree feedback results comes from Atwater and Brett (2003). These researchers compared several different report presentations to participants and concluded the following:

a. Individuals appear to be significantly less positive and less motivated after receiving text feedback than after receiving numeric feedback.

b. Individuals appear to prefer numeric over qualitative scores and normative over relative comparisons of self and other ratings in 360-degree summary reports and find these useful for development purposes.

Implications. Unfortunately, practice has truly outstripped research in guiding developers of 360-feedback assessments to optimize report content and to organize the presentation of data to

enhance understanding and acceptance of results without it being confusing or overwhelming. Most 360-degree feedback reports tend to have common sections including the following: (1) Comparison of self-other rating similarities/differences using bar, line, spider, and scatterplot graphs to illustrate these gaps and trends; (2) A list of most and least frequently observed or effective behaviors that summarize strengths and development opportunities; and (3) Open-ended question narrative responses about development opportunities or strengths to leverage.

It appears that "what the crowd does" is what is now commonly accepted in how 360-degree feedback reports are organized and presented in light of a paucity of evidence-based practices to guide developers and coaches. Until more research has accumulated, the findings of Atwater and Brett (2003) would appear to support the more popular and common approaches being used today. Future research should also explore the delivery of reports as well as the type of content made available to clients given new technological opportunities to receive and view reports on personal computers, portable tablets, and smartphones either in self-directed or facilitated fashion by coaches or others (e.g., showing selected report components via video conferencing).

360-Degree Process and Implementation

9. Can Open-Ended Questions Be Emotionally Damaging to Clients?

It is common in most online based 360-degree feedback assessments to include one or more open-ended questions which are typically voluntarily and confidentially answered by raters (typically these comments are categorized by rater groups used in the assessment process or listed in a random order without reference to rater groups). In general, there has been little research evaluating the cognitive and emotional reactions of such qualitative feedback on clients in the 360-degree literature based on online data collection. Narrative comments shared by raters can possibly be evaluative, overly critical, or negative, having an adverse impact on acceptability and motivation if included in reports without editing or removal.

A study by Smither and Walker (2004) analyzed the impact of upward feedback ratings, as well as narrative comments, over a 1-year period for 176 managers. The number of raters providing comments per target manager ranged from 1 to 12 (M = 3.10, *SD* = 2.21). The number of comments per target manager ranged from 1 to 35 (M = 7.35, *SD* = 6.14). Seventy percent of the comments were coded as 3.5 or higher (1 = *unfavorable*, 5 = *favorable*). The study found that those who received a small number of unfavorable, behaviorally based comments improved significantly more than other managers, but those who received a large number (relative to positive comments) significantly declined in performance more than other managers (Smither et al., 2004). Unfortunately for coaches, the study did not quantify the exact number or ratio of negative to positive comments that might be the "tipping point" for performance declines.

Implications. When the 3:1 ratio of positive-to-negative open-ended comments begins to decrease and be of a magnitude that could create strong emotional reactions on the part of clients (Losada et al., 2009), coaches face a set of ethical questions for which clear answers are difficult to answer. The APA Ethical Principles of Psychologists and Code of Conduct (American Psychological Association, 2010) in the Avoiding Harm Standard 3.04 states "Psychologists take reasonable steps to avoid harming their clients/patients, students, supervisees, research participants, organizational clients and others with whom they work, and to minimize harm where it is foreseeable and unavoidable" and the Assessment Standard 9.02a states "Psychologists administer, adapt, score, interpret or use assessment techniques, interviews, tests or instruments in a manner and for purposes that are appropriate in light of the research on or evidence of the usefulness and proper application of the techniques."

These ethical guidelines suggest that when open-ended comments are overwhelmingly negative with little prescriptive feedforward suggestions to improve, coaches should consider reasonable options to organize and summarize the themes of the feedback and present them back to the client in a manner that will engender understanding, acceptance and the management of potential negative emotional reactions. The findings of Smither et al., (2004) highlight the necessity to minimally

follow-up with clients after they have received and reviewed their 360-degree feedback report to ensure that any negative emotions and reactions can be processed in a healthy manner.

Like quantitative results, open-ended comments create strong emotional reactions that can interfere with the acceptance of feedback and lead to diminished engagement and performance. Additional research on ways to improve gathering constructive and meaningful "feedforward" comments from raters and the impact of such ratings on client's motivation, emotional reactions, and readiness to change future behavior would appear to be warranted.

10. Does Personality Impact How People Respond to 360-Degree Feedback?

Personality appears to directly influence how clients react to 360-degree feedback, how motivated they will be to act on the observations and suggestions of others, and how likely they will be to implement and sustain new behaviors to become more effective. For example, research by Smither, London, and Richmond (2005) explored the relationship between personalities of leaders and their reactions to and use of 360-degree feedback. Leaders high in the five-factor personality facet of emotional stability were significantly more likely to be rated by a psychologist as motivated to use the feedback results for their ongoing professional development. Additionally, leaders high in extraversion were significantly more likely to have sought additional feedback six months later, whereas leaders high in conscientiousness were more likely to have actually engaged in developmental behaviors. These researchers and others have found that extraverted leaders who were also high on the personality factor of openness to experience were more likely to perceive and view negative feedback as valuable data and were most likely to seek further information about their feedback (Bell & Arthur, 2008).

Several studies support the findings that individuals with high self-esteem report more favorable attitudes toward the 360-degree feedback results than those with low self-esteem (Bono & Colbert, 2005). Feedback recipients who rated themselves highly on receptivity and the desire to make a good first impression were also perceived by feedback providers as having more positive reactions to feedback (Atwater, Brett, & Charles, 2007) as well as clients who generally rated themselves lower than others (Brett & Atwater, 2001).

Research by Bono and Colbert (2005) provided evidence that the motivation to change behavior following 360-degree feedback is related to a metapersonality concept called core self-evaluations (CSE). Specifically, they found that individuals with high levels of core self-evaluations (those with high self-esteem, generalized self-efficacy, internal locus of control and low neuroticism) will be most motivated to initiate change behavior when they receive discrepant feedback, whereas those with low levels of core self-evaluations will be most motivated when others' ratings are most similar to their own. As such, the personality of the client has a direct result in the level of readiness to change based on both the direction and magnitude of self-other agreement.

The stable personality trait of goal orientation has also been shown to influence whether an individual views feedback as a development opportunity or a challenge to his or her self-rating (Dweck, 1986). Individuals with a learning goal orientation tend to hold a view of ability as modifiable and believe they are capable of improving their level of abilities (Brett & Atwater, 2001). These researchers found that those with a learning goal orientation believed the feedback was more useful than those with a performance goal orientation.

Implications. Taken together, it seems clients are most motivated to use 360-degree feedback for development when they are conscientious, achievement oriented, extraverted, possess high self-efficacy, have an internal locus of control, a learning goal orientation, and express low anxiety. Identifying and understanding the personality of clients will help coaches to structure coaching and feedback interventions in a manner that facilitates both readiness to change and the enhancement of self-efficacy (Rhodes, Plotnikoff, & Courneya, 2009).

Coaches might also consider including a newer generation five-factor personality inventory or a structured interview in their practice to assess personality factors that might contribute or hinder to the acceptance of 360-degree feedback and overall commitment to initiate and sustain successful behavior change over time. For example, clients who report being low on conscientiousness tend to be less capable of controlling, regulating, and directing their impulses and, as such, may be less

committed to long-term behavior change and personal/professional development (Klockner & Hicks, 2008). In such cases, coaches may need to explore ways to structure behavior change efforts to maximize recognition and reinforcement after short-term goal accomplishments and schedule structured follow-up with their clients using a variety of methods (e.g., meetings, email reminders) to encourage continued efforts.

11. How Do You Manage the Feedback of Underestimators and Overestimators?

It has been estimated that 65% to 75% of the employees in any given organization report that the worst aspect of their job is their immediate boss (Hogan, 2007, p. 106). In fact, estimates of the base rate for managerial incompetence in organizations range from 30% to 75%, with the average level of poor leadership estimated at about 50% (Hogan & Kaiser, 2005). Many of these incompetent leaders tend to have inflated views of their skills and abilities, and this appears fairly common in 360-degree feedback research (Atwater & Brett, 2005). In a study by Vecchio and Anderson (2009), the tendency to overestimate one's own leader effectiveness relative to evaluations provided by others was found to be greater for males and older managers.

These self-enhancers or overestimators are often blind to accurately identifying their own strengths, less receptive to feedback from others, have negative reactions to feedback (Brett & Atwater, 2001), and are at high risk for derailment (Quast, Center, Chung, Wohkittel, & Vue, 2011). As a result, coaches might find it difficult to find the "what's in it for me" with such clients to accept the perceptions of others and commit to modifying their behavior to some degree in order to better meet the expectations and needs of those working with them.

In a recent Harvard Business Review article, Kaplan and Kaiser (2009) argue that it is just as detrimental to overuse our strengths as it is to underuse them. In their research, those expressing the right amount of strength showed a significant association with a measure of leadership success. As the authors point out, leveraging and emphasizing strength might lead to actually interfering with being flexible and adopting new behaviors.

Goffin and Anderson (2007) found in their study of 204 managers that self-rating inflation was significantly correlated with high achievement, self-esteem, and social desirability personality factors. This personality profile pattern suggests that self-enhancers might possess an exaggerated perception of their strengths resulting in potential defensiveness and resistance during 360-degreee feedback discussions with their coach or others. It should also be noted that the pattern of high social desirability and low anxiety (repressive coping) has long been shown in the health psychology literature to be significantly associated with increased cardiovascular reactivity to stress, higher blood pressure, and poor overall health outcomes (Mund & Mitte, 2011; Rutledge, 2006). This pattern, found in an earlier Goffin and Anderson (2002) study, suggests that overestimators might not only be at risk to derail in their careers but also vulnerable to negative physical health outcomes. To date, no research has directly tested this hypothesis or considered that the most vulnerable overestimator might indeed be a personality profile characterized as high in social desirability, low in negative affect (anxiety), and simultaneously high in positive affect (i.e., a "super repressor").

Another form of "cognitive distortion" that is common in 360-degree feedback processes are characterized by clients rating themselves significantly lower than the ratings of others. These underestimators are actually viewed as possessing strengths but not fully recognizing or acknowledging them relative to others giving them feedback (Nowack, 2009). Furthermore, research by Goffin and Anderson (2007) suggests that underestimators score significantly higher on negative affect than overestimators, suggesting they are likely to be more emotionally reactive, anxious, and nervous in the interpretation of their feedback results. Nowack (2009) reported that underestimators (about 25% to 30% of those taking 360-degree assessments) are typically characterized as highly perfectionist, expect high performance for themselves and others, focus on their weaknesses and look for fault, criticism, and potential deficits in their feedback from others, and reframe feedback suggesting strengths as being too complimentary.

In summary, underestimators tend to minimize the strengths seen by others and dwell on anything that isn't perfect in their summary feedback report. In practice, these clients are resistant to leverage their strengths as seen by others even when it is pointed out that they are underestimating how others are

experiencing the frequency or effectiveness of their expressed behavior. The underestimators tend to be hyper-vigilant to anything they perceive to be critical or negative in 360-degree feedback reports and emphasize what they perceive to be developmental opportunities or weaknesses.

Implications. Coaches should become familiar with the impact of client self-enhancement (both magnitude and direction) on the understanding, acceptance, and actions taken after 360-degree feedback as well as how it might predict job performance. For example, a study by Atwater, Ostroff, Yammarino, and Fleenor (1998) found that leadership effectiveness in 1,460 managers was highest when both self and other ratings were high, and when self-ratings were substantially lower than other ratings (severe underestimation). Effectiveness was the lowest for overestimators when self-rating was only moderate and subordinate ratings were low.

The discrepancy between self-ratings and other ratings can affect both emotional reactions and readiness to change behavior. Current research suggests mixed findings for the association between affect and behavioral change. For example, Atwater and Brett (2005) suggest that leaders who received low ratings and overrated themselves were actually more motivated to change than leaders who received low ratings and gave themselves low ratings. However, these overraters also had more negative reactions (e.g., were more angry) than those who did not overrate. In contrast, other research suggests that overestimators are significantly less likely to engage in developmental plans after negative feedback (Woo, Sims, Rupp, & Gibbons, 2008).

Coaches should be aware that clients who are underestimators are likely to be highly perfectionistic, self-critical, and express high negative affect making them likely to dismiss the strengths perceived by others. It should be expected that underestimating clients will not see their feedback in balance and coaches should anticipate that their clients will accentuate and focus on the negative, despite feedback from others that they are actually performing strongly or possess high competence in particular skills and abilities being rated. In practice, getting underestimating clients to leverage their strengths in developmental planning is one of the biggest challenges faced by coaches during 360-degree feedback meetings.

In contrast, clients who are overestimators are likely to be highly achievement oriented, express high self-esteem, project a socially desirable impression of their behaviors, and report little anxiety. Coaches should also note that self-enhancing assessments of poor performance ratings from others may take different forms such as attending selectively to only the positive indicators or minimizing negative indicators.

Jordan and Audia (2012) point out three common self-enhancing assessments of receiving low performance ratings from others including: (1) Downplaying the importance of the performance goals to perceive low performance in a more positive light; (2) Redefining the level of abstraction of a performance goal to make it more flexible or broad; and (3) Focusing on how things could have been worse if they had acted differently (i.e., focusing on counterfactual outcomes as comparisons).

In practical terms, sometimes coaches have to work hard to "find the crease" to allow clients to digest and accept constructive negative feedback they have received without dismissing it outright. Finally, it is important to note that people not only compare themselves with others but also with how they used to be in the past. In general, individuals evaluate their current and future selves as better than their past selves (Wilson & Ross, 2001), suggesting that coaches should focus their developmental planning efforts with clients in a future-oriented manner and help them compare one's "ideal self" with one's "real self" (Boyatzis & Akrivou, 2006).

12. What Kind of Training or Certification by Coaches Is Required to Help Clients Understand and Interpret 360-Degree Feedback Reports?

Coaches and consultants may have very diverse backgrounds and academic degrees, but familiarity with assessments in general should be required to professionally utilize 360-degree feedback systems (Nowack, 2003). In their article in *Harvard Business Review*, Sherman and Freas (2004) stated that executive coaching is "Like the Wild West of yesteryear, this frontier [executive coaching] is chaotic, largely unexplored, and fraught with risk, yet immensely promising" (pp. 82–83).

Current research on coaching differences by education and training has found that psychologists are more likely to meet face-to-face, contract for fewer sessions, and are more likely to use

360-degree assessments in their practice than nonpsychologists (Bono et al., 2009). Bono et al., (2009) found in their study of 428 coaches (256 nonpsychologists, 172 psychologists) that differences were generally small (average difference = .26) and there were as many differences between psychologists in their training and orientation to coaching (e.g., clinical, social/personality, or industrial/organizational) as between coaches and nonpsychologists. Additionally, these researchers reported that psychologist coaches were "more likely to use effective tools to diagnose the problem (e.g., multisource behavioral ratings, interview with a supervisor, interview with peers, ability/aptitude tests, and review of prior performance data" (p. 390).

Implications. Coaches should be knowledgeable and competent in the use and interpretation of assessments including 360-degree feedback. In light of current research, it would appear necessary that coaches possess adequate measurement and statistical expertise to fully explain and interpret 360-degree feedback results to their clients.

Coaches should continue to pursue continuing education and training to enhance their knowledge (e.g., use of assessments), skills (e.g., managing tricky ethical dilemmas), and abilities (handling resistance) to help their clients gain self-insight and facilitate long-term behavioral change success. For example, if an executive coach is clinically trained as a psychologist he or she might need to enhance their knowledge about business/industry, or if trained in a research-oriented industrial-organizational psychology program he or she might need to enhance core counseling skills.

13. Are There Cultural Differences to Be Considered in the Use of 360-Degree Feedback?

There is increasing use of 360-degree feedback in different cultures and countries, as multinational companies use it throughout their entire organization. Differences in 360-degree feedback rating and interpretation should be expected to some degree in other cultures. Several cultural dimensions have been thoroughly studied (Hofstede & McRae, 2004) and would appear to be meaningful to 360-degree ratings (self and others). These cultural dimensions include individualism versus collectivism, power distance, uncertainty avoidance, short-term versus long-term orientation, and gender egalitarianism.

Varela and Premeaux (2008), in their sampling of managers in Latin America, found the least discrepancy between peer and self-ratings. In their analysis, direct reports gave the highest ratings to their managers in this highly collectivistic and high power distance culture. Cultural differences between geographic regions in Asia have been found to be associated with patterns of self-ratings of managerial performance (Gentry, Yip, & Hannum, 2010). These researchers found that significant self-other discrepancies were wider in high power and individualistic cultures mainly due to the subject's self-ratings and not the ratings of others. In a comparison of U.S. managers (n = 22, 362) with an Asian sample of 3,810 managers consisting of five countries, Quast et al., (2011) found that self-other discrepancies in all countries were significantly associated with bosses' predictions of how likely a manager was to experience future career derailment. These results provide support for earlier findings that self-other rating discrepancies are associated with derailment in the United States and extend these findings to the five Asian countries included in this study (China, S. Korea, Japan, India, and Thailand).

Atwater, Wang, Smither, and Fleenor (2009) explored self and subordinate ratings of leadership in 964 managers from 21 countries, based on assertiveness, power distance, and individualism or collectivism. Self and other ratings were more positive in countries characterized as high in both assertiveness and power distance. However, Atwater, Waldman, Ostroff, Robie, and Johnson (2005) found varying multisource ratings patterns (i.e., self-other agreement) in different cultures. Their study showed that links between self-other discrepancies and managerial effectiveness varied greatly and these discrepancies were related to effectiveness in the United States but not in the European and Scandinavian countries of Germany, Denmark, Italy, and France (only others' ratings of leadership predicted managerial effectiveness in these countries).

In one of the broadest studies to date, Eckert, Ekelund, Gentry, and Dawson (2010) investigated self-observer rating discrepancies on three leadership skills on data from 31 countries. They reported

that rater discrepancy on a manager's decisiveness and composure was higher in high power distance cultures (e.g., Asian) than low power distance cultures (e.g., Americas). Self-observer rating discrepancy has also been shown to be higher (i.e., bigger or wider) for U.S. American managers than for Europeans on 360-degree ratings of managerial derailment behaviors (Gentry, Hannum, Ekelund, & de Jong, 2007). At least in the United States, higher disagreement between self and observer ratings is generally associated with lower effectiveness and job performance (Ostroff, Atwater, & Feinberg, 2004; Atwater & Brett, 2005), but some contradictory evidence has been found in other countries (Atwater et al., 2009).

Cultural relevance was compared across five countries (United States, Ireland, Israel, Philippines, and Malaysia), and this supported the overall effectiveness of the 360-degree feedback process but also revealed important differences (Shipper, Hoffman, & Rotondo, 2007). This study suggested that the 360-degree feedback process is relevant in all cultures but most effective in those low on power distance with individualistic values (e.g., United States vs. Philippines). Finally, earlier research on 360-degree feedback across 17 countries by Robie, Kaster, Nilsen, and Hazucha (2000) suggested that there were more similarities than differences across countries. For example, the ability to solve complex problems and learn quickly appears to be universally predictive of effectiveness for leaders across cultures high and low in power.

In general, interpreting the current literature of 360-degree feedback across countries and cultures is somewhat difficult given the different competencies being evaluated in the studies cited (the majority were vendor developed), the various approaches to measuring self-other differences (from simple algebraic difference scores to uses of more sophisticated polynomial regression), and the diverse performance/derailment outcome measures used (e.g., single item predictions of derailment to managerial performance ratings). Additional research is needed and will continue to guide coaches to understanding the cross-cultural implications surrounding the effective use of feedback interventions and the meaningfulness of the gaps between self and other differences.

Implications. Taken together, these newer cross-cultural 360-degree feedback studies suggest that factors such as values, norms, and beliefs have an impact on self–other rating discrepancies and their meaning. Despite some contradictory evidence, the relevance of self–other ratings appears to be important for coaches to use and interpret for both development and nondevelopment purposes and appears to impact both leadership performance and potential derailment. In general, peer ratings appear to have utility for predicting future leadership potential across cultures. Coaches who work with multinational organizations should continue to develop their own cultural competence and knowledge of relevant norms, values, and history as they interact with leaders and talent at all levels.

Additionally, alternative competency models defining cross-cultural leaders might be strongly considered for future feedback interventions given the lack of a universal taxonomy or systematic framework for evaluating the content coverage of such assessments (Holt et al., 2012). Current research suggests that many 360-degree feedback assessments with multiple competencies lack research-based frameworks, are often highly intercorrelated with each other (Hoffman & Woehr, 2009), and typically possess a small number of underlying factors (Smither et al., 2005). These are critical issues given that 360-degree feedback systems are commonly used in developmental settings and that competencies are the focus when interpreting and acting upon developmental feedback.

For example, coaches should be familiar with the research behind the competency framework used in the particular 360-degree feedback assessment they are using. They should also have some understanding of the academic debate about various leadership taxonomies such as the "Great 8 Competencies" by Bartram (2005), the four theoretically derived behavioral types of leadership based on the extension of the transactional-transformational model of leadership by Pearce et al. (2003), and the three metacategories of task, relations, and change behavior introduced by Yukl, Gordon, and Taber (2002).

Coaches should critically examine the competency-based framework behind the specific 360-degree feedback assessment they are using and align the model with the specific goals of the project and the job level of the clients they are working with. They should only use commercially based, or organizationally developed, assessments that have adequate psychometric properties (e.g., inter-

nal consistency reliability) and a factor structure to support the breakdown of specific competencies that often are targeted for development planning by participants.

14. Does a 360-Degree Feedback Report Require Debriefing?

"Best practices" in 360-degree feedback processes suggest that greater transfer of learning and goal setting occurs when a manager and/or coach helps clients understand and debrief their reports (Nowack, 2009). For example, Arthur, Bennett, Stanush, and McNelly (1998) conducted a meta-analysis of knowledge and skill decay studies and reported that one day after training, trainees exhibit little to no skill decay, but one year after training they lost more than 90% of what they learned. Some vendors and some coaches espouse the "diagnose and adios" approach to 360-degree feedback, hoping that self-directed reflection alone will result in motivated behavioral change efforts.

In one of the few prospective empirical studies conducted on the impact of executive coaching, Smither et al., (2003) reported that after receiving 360-degree feedback, managers who worked with a coach were significantly more likely to set measureable and specific goals and solicit ideas for improvement. They subsequently received significantly improved performance ratings. Thach (2002) found that in six weeks of executive coaching after 360-degree feedback, performance increased by 60%. In a much-cited study in the public sector, Olivero, Bane, and Kopelman (1997) found that employee feedback and coaching for 2 months increased productivity above the effects of a managerial training program (22.4 vs. 80.0%) for 31 clients. These coaching studies support the importance of supportive follow-up after feedback is received to facilitate developmental action planning and the practice of targeted behaviors.

Some limited support for other approaches to structured follow-up comes from a recent doctoral dissertation study evaluating the effectiveness of 360-degree feedback interventions in 257 leaders in diverse organizations (Rehbine, 2007). In this study, more than 65% of those surveyed expressed a strong interest in using some type of an online follow-up tool to measure progress and facilitate their own individual behavioral change efforts. However, newer research also suggests that the approach to coaching might be as important as the modality (self-directed using an online development and reminder system or coach directed).

A recent study by Hooijberg and Lane (2009) surveyed 232 managers from diverse organizations and investigated what makes coaching most effective for clients. One of the key questions asked was "What did your coach do that you found most effective?" From the view of the client or clients, three major categories determined feedback success: (1) interpreting results (34.8%), (2) inspiring action (27.5%), and (3) professionalism of the coach (23.3%). The majority of clients thought the best coaches were those who analyzed strengths and weaknesses, helped assimilate feedback, and made concrete developmental recommendations. This study seems to contradict much of the coaching literature and suggests that clients using 360-degree feedback expect and want their coach to take a much more active and directive role in interpreting their results and making developmental recommendations to leverage actual behavior change.

Implications. Overall, "best practices" would strongly suggest that 360-degree feedback reports be discussed with clients (or be made available with the use of highly structured online goal setting systems that are sometimes integrated with such feedback assessments). Greater use of "prescriptive" and "feedforward" suggestions on the part of the coach, manager, or consultant debriefing the report would appear to maximize readiness to change and the targeting of specific behaviors to use more, less, or differently to enhance overall performance.

Coaches should use some of the prescriptive suggestions from models of transfer of training that suggest that factors before, during, and after feedback can positively influence the extent of transfer back to the job (Grossman & Salas, 2011). Finally, based on research suggesting the importance of personality factors in feedback processes, coaches should carefully consider the mediating effects of self-efficacy, mastery goal orientation, and motivation to transfer during their debriefings to leverage learning transfer (Chiaburu, Van Dam, & Hutchins, 2010).

15. How Can You Leverage the Impact of 360-Degree Feedback to Ensure Successful Behavioral Change?

Smither, London, Flautt, Vargas, and Kucine (2003) found that leaders who received unfavorable feedback initially had negative reactions, but six months later they had created significantly more improvement goals than other leaders. They suggested "negative feedback may take a while to sink in or recipients may need some time to reflect and absorb the feedback" (p. 203) after the initial emotions have subsided. As pointed out earlier, Brett and Atwater (2001) found that individuals who received negative feedback from bosses and peers reported adverse emotional reactions. However, the impact of initial reactions seemed to lessen after several weeks and was not related to perceived feedback usefulness despite the emotionality surrounding the process.

Taken together, these findings suggest two things. The first is that it is important to manage the initial emotional reactions that clients have and to identify key personality traits that might exacerbate or temper these responses (e.g., narcissism, self-esteem, and emotional stability). Second, translating awareness into goal implementations would appear to be most critical to ensure that clients translate the 360-degree feedback experience into deliberate practice of new behaviors to accentuate what they do well, acquire new habits, or modify existing ones to become more effective (Smither, Brett, & Atwater, 2008).

New evidence suggests that perceived importance of the desired behavioral change end point is the best predictor differentiating nonintenders from those who are successful adopters of new behavior. However, self-efficacy, perceived control, and being clear about the "cons" behind behavioral change are more important in discriminating successful maintainers from unsuccessful maintainers who relapse and fall back to their older habits and routines (Rhodes, Plotnikoff, & Courneya, 2009). Finally, recent research on goal accomplishment suggests that a shift in attention and motivation level from the starting point to the end point occurred halfway through the goal so this might be one of the most important times for coaches to follow-up with their clients and discuss possible relapse prevention strategies (Bonezzi, Brendl, & De Angelis, 2011). As such, clients might require more attention, reinforcement, and follow-up in the middle of a coaching intervention than any other time based on the course of motivation over time.

Typically, the development of expertise in a complex activity requires at least 10 years and/or 10,000 hours of deliberate practice (Ericsson, 2006). Acquiring new habits also requires repeated practice so that the new behavior is automatic (unconscious competence). To investigate the process of habit formation in everyday life, 96 volunteers chose an eating, drinking, or activity behavior to carry out daily in the same context (e.g., "after breakfast") for 12 weeks. The study participants completed the self-report tracking form each day and recorded whether they carried out the behavior (Lally, Van Haarsveld, Potts, & Wardle, 2010). The number of days it took for a new behavior to become "automatic" depended on its complexity (e.g., new eating habits 65 days and exercise 91 days). Participants who missed the habit just once did not seem to experience a relapse, though those who missed it multiple times did. Of the 82 clients who saw the study through to the end, the most common pattern of habit formation was for early repetitions of the chosen behavior to produce the largest increases in its automaticity. Over time, further increases in automaticity dwindled until a plateau was reached beyond which extra repetitions made no practical difference to the automaticity achieved.

Organizations that implement a systemic approach to talent development with support from a manager and follow-up development activities tied to performance improvement will have the most effective outcomes in leadership development (Nowack, 2009). A better understanding of the role of the manager as an important internal coach and how organizational culture influences promoting and sustaining new behavior is in need of greater exploration. The manager can play a big role to reinforce and support the implementation of the development plan of their direct report.

Implications. Use of newer online goal setting and development planning/reminder systems may be promising, and recent studies seem to suggest that clients are interested in using these for translating awareness from 360-degree feedback into behavior change (Rehbine, 2007). For example, in a recent unpublished 1-year longitudinal study using an online goal setting and tracking system with a major university medical center (Nowack, 2011), significant behavioral change was

observed on an overall score averaged across all items and raters (managers, direct reports, and peers) on a post-program 360-feedback assessment, $F(2, 11) = 4.72$, $p = .03$.

Martin (2010) found a positive effect on learning transfer for peer support in a corporate field environment, with peer support and encouragement mitigating a negative work climate. Martin (2010) evaluated learning transfer on 237 managers of a manufacturing company in the Midwest United States who participated in a leadership development program. He found that leaders in a division with a more favorable climate and those reporting greater peer support showed the greatest transfer of learning but that support for peers bolstered transfer in the face of more negative work climates.

Additionally, managers who follow-up with talent who have taken 360-degree feedback assessments are more likely to set specific goals, solicit ideas for improvement, and subsequently receive improved performance ratings (Smither, London, Flautt, Vargas, & Kucine, 2003). These findings support the importance of involving the manager in the coaching intervention to leverage long-term behavior change success. This is important in light of recent findings suggesting that effect sizes for transfer of management training interventions are generally low (particularly when seen by direct reports and peers) but can be improved significantly with opportunities for structured and deliberate practice over time (Taylor, Taylor, & Russ-Eft, 2009).

Overall, these findings suggests that performance can be practically enhanced by using a 360-degree feedback process involving both peers as development partners and managers as performance coaches to hold clients accountable for creating and implementing a development plan based on 360-degree feedback results (Chiaburu & Marinova, 2005). Future research is required to replicate these findings about the facilitating roles that peers and managers might play in reinforcing new behaviors following 360-degree feedback programs.

Recent research suggests that stated goal intentions alone may not always result in successful maintenance of behavior over time (Lawton, Cooner, & McEachan, 2009). After 360-degree feedback, many clients express a strong desire and intent to become more effective and may actually try new behaviors. However, as a result of relapse these individuals might be unable to sustain them for very long. This suggests that coaches should emphasize relapse prevention techniques, facilitate self-efficacy, and become a "professional nag" to their clients to help reinforce new behavior change efforts (Rhodes, Plotnikoff, & Courneya, 2009). Coaches should also initially focus more on implementation intentions with their clients using "if-then" goal statements to maximize behavioral change commitments, planning, and maintenance over time (Gollwitzer & Sheeran, 2009).

Conclusions

In general, there has been a lack of attention in both research and practice in exploring ways to successfully help clients initiate and sustain new behavior with the use of 360-degree feedback (Bracken et al., 2011; Joo, 2005; London & Smither, 2002). Behavior change efforts are often not linear but tend to be progressive, regressive, or even static. Current findings suggest that multiple simultaneous efforts (e.g., behaviors planned to improve multiple competencies at the same time) tend to be equal or even more effective than focusing on single goals because they reinforce quick benefits (Hyman, Pavlik, Taylor, Goodrick, & Moye, 2007).

Coaches should also be familiar with the existing literature across disciplines (e.g., health psychology and management) on different models of individual behavioral change. For example, Nowack (2009) introduced the Enlighten, Encourage, and Enable behavior change model to help facilitate successful goal initiation and implementation. This model (Nowack, 2009) is based on the applied theories of individual behavioral change including the theory of planned behavior (Ajzen, 1991), self-efficacy and social–cognitive theory (Bandura, 1977), the health belief model (Becker, 1974), intentional change (Boyatzis, 2008), and the Transtheoretical/readiness to change model (Prochaska & Velcier, 1997). Each of these theories and individual change models should be useful to coaches who are attempting to extend the utility of 360-degree feedback beyond awareness and toward enhanced individual, team, and organizational effectiveness or impact.

Sustaining behavioral change for anyone is challenging in the most ideal situations. The evidence-based limitations of feedback interventions along with an earlier meta-analysis by Kluger

and DeNisi (1996) all support the idea that enhancing awareness and effectiveness of feedback depends on a complex interplay of intrapsychic, interpersonal, organizational, and external factors. Individual differences (e.g., personality) can impact the motivational level following feedback as well as the goal setting process. Finally, coaches need to make an evaluation before using 360-feedback interventions in light of the possibility that some clients are unsuitable for coaching in the first place (Goldsmith, 2009).

Attention to these various evidence-based issues and challenges can assist coaches to consider the best approach to using feedback interventions to create the desired individual, team, and organizational outcomes (Nowack, 2005). Coaches and consultants who deliver feedback or use 360-degree feedback interventions should become familiar with the questions and answers summarized in this article and the latest evidence-based research studies available to leverage the impact of this intervention on both proximal (awareness) and distal (successful behavior change) goals.

References

Ajzen, I. (1991). The theory of planned behaviour. *Organisational Behaviour and Human Decision Processes, 50,* 179–211.

American Psychological Association. (2010). American Psychological Association ethical principles of psychologists and code of conduct. Retrieved from http://www.apa.org/ethics/code/index.aspx#

Arthur, W., Jr., Bennett, W., Jr., Stanush, P. L., & McNelly, T. L. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance, 11,* 57–101.

Atkins, P. W. B., & Wood, R. E. (2002). Self- versus others' ratings as predictors of assessment center ratings: Validation evidence for 360-degree feedback programs. *Personnel Psychology, 55,* 871–904.

Atwater, L. E., & Brett, J. F. (2005). Antecedents and consequences of reactions to developmental 360-degree feedback. *Journal of Vocational Behavior, 66,* 532–548.

Atwater, L., & Brett, J. (2006). 360 degree feedback to managers: Does it result in changes in employee attitudes? *Group & Organizational Management, 31,* 578–600.

Atwater, L. A., Waldman, D., Atwater, D., & Cartier, P. (2000). An upward feedback field experiment. Supervisors' cynicism, follow-up and commitment to subordinates. *Personnel Psychology, 53,* 275–297.

Atwater, L., Brett, J., & Charles, A. (2007). Multi-source feedback: Lessons learned, and implications for practice. *Human Resource Management Journal, 46,* 285–307.

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

Atwater, L., Waldman, D., Ostroff, C., Robie, C., & 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., Wang, M., Smither, J., & Fleenor, J. (2009). Are cultural characteristics associated with the relationship between self and others' rating of leadership? *Journal of Applied Psychology*, *4,* 876–886.

Bandalos, D. L., & Enders, C. K. (1996). The effects of non-normality and number of response categories on reliability. *Applied Measurement in Education, 9,* 151–160.

Bandura, A. (1977). Self-Efficacy: Toward a unifying theory of behavior change. *Psychological Review, 84,* 191–215.

Bartram, D. (2005). The Great 8 competencies: A criterion-centric approach to validation. *Journal of Applied Psychology, 90,* 1185–1203.

Becker, M. H. (1974). The Health Belief Model and Personal Health Behaviour. *Health Education Monographs, 2,* 324–473.

Beehr, T. A., Ivanitskaya, L., Hansen, C. P., Erofeev, D., & Gudanowski, D. M. (2001). Evaluation of 360 degree feedback ratings: Relationships with each other and with performance and selection predictors. *Journal of Organizational Behavior, 22,* 775–788.

Bell, S. T., & Arthur, W. A. (2008). Feedback acceptance in developmental assessment centers: The role of feedback message, participant personality, and affective response to the feedback session. *Journal of Organizational Behavior, 29,* 681–703.

Bonezzi, A., Brendl, C. M., & De Angelis, M. (2011). Stuck in the middle: The psychophysics of goal pursuit. *Psychological Science, 22,* 607–612.

Bono, J., & Colbert, A. (2005). Understanding responses to multi-source feedback: The role of core self-evaluations. *Personnel Psychology, 58,* 171–203.

Boyatzis, R. E. (2008). Leadership development from a complexity perspective. *Consulting Psychology Journal: Practice and Research, 60,* 298–313.

Boyatzis, R. E., & Akrivou, K. (2006). The ideal self as a driver of intentional change. *Journal of Management Development, 25,* 624–642.

Bracken, D. W., & Rose, D. S. (2011). When does 360-degree feedback create behavior change? And how would we know it when it does? *Journal of Business and Psychology, 26,* 183–192.

Bracken, D. W., Timmreck, C. W., Fleenor, J. W., & Summers, L. (2001). 360 feedback from another angle. *Human Resources Management, 40,* 3–20.

Brett, J., & Atwater, L. (2001). 360-degree feedback: Accuracy, reactions and perceptions of usefulness. *Journal of Applied Psychology, 86,* 930–942.

Caputo, P. M., & Roch, S. (2009, April). *Rating formats and perceptions of performance appraisal fairness*. Poster presented at the 24th meeting of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Chang, L. (1994). A psychometric evaluation of 4-point and 6-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement, 18,* 205–215.

Chemers, M. M., Watson, C. B., & May, S. T. (2000). Dispositional affect and leadership effectiveness: A comparison of self-esteem, optimism, and efficacy. *Personality and Social Psychology Bulletin, 26,* 267–277.

Chen, Z., Williams, K., Fitness, J., & Newton, N. (2008). When hurt will not heal. Exploring the capacity to relive social and physical pain. *Psychological Perspectives, 19,* 789–795.

Chiaburu, D. S., & Marinova, S. V. (2005). What predicts skill transfer? An exploratory study of goal orientation, training self-efficacy and organizational supports. *International Journal of Training and Development, 9,* 110–123.

Chiaburu, D. S., Van Dam, K., & Hutchins, H. M. (2010). Social support in the workplace and training transfer: A longintudinal analysis. *International Journal of Selection and Assessment, 18,* 187–200. doi:10.111/j.1468-2389.2010.00500.x

Cicchetti, D. V., Showalter, D., & Tyrer, P. J. (1985). The effect of number of rating scale categories on levels of interrater reliability: A Monte Carlo investigation. *Applied Psychological Measurement, 9,* 31–36.

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

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

Cools, W., Hofmans, J., & Theuns, P. (2006). Context in category scales: Is "fully agree" equal to twice agree? *European Review of Applied Psychology, 56,* 223–229.

Craig, S. B., & Hannum, K. (2006). Research update: 360-degree performance assessment. *Consulting Psychology Practice and Research, 58,* 117–122.

Dai, G., Stiles, P., Hallenbeck, G., & De Meuse, K. P. (2007, August). Self-other agreement on leadership competency ratings: The moderating effects of rater perspectives and rating ambiguity. Paper presented at the 2007 Annual Meeting of the Academy of Management, Philadelphia, PA.

Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *International Journal of Market Research, 50,* 61–77.

DeWall, C. N., MacDonald, G., Webster, G., Masten, C. L., Baumeister, R. F., Powell, . . . Eisenberger, N. I. (2010). Acetaminophen reduces social pain: Behavioral and neural evidence. *Psychological Science, 21,* 931–937. doi:10.1177/0956797610374741

3D Group. (2009). *Current practices in 360-degree feedback: A benchmark study of North American Companies*. 3D Group Tech. Rep. No. #8326. Berkeley, CA: Data Driven Decisions, Inc.

Dickerson, S., & Kemeny, M. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin, 130,* 355–391.

Dunning, D., Heath, C., & Suls, J. (2004). Flawed self-assessment: Implications for health, education and the workplace. *Psychological Science in the Public Interest, 5,* 69–106.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41,* 1040–1048.

Eckert, R., Ekelund, B., Gentry, W., & Dawson, J. (2010). I don't see me like you see me but is that a problem? Cultural differences in rating discrepancy in 360-degree feedback instruments. *European Journal of Work and Organizational Psychology, 19,* 259–278.

Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An fMRI study of social exclusion. *Science, 302,* 290–292.

English, A., Rose, D., & McClellan, J. (2009, June). Rating scale label effects on leniency bias in 360-degree feedback. Paper presented at the 24th Annual Meeting of the Society for Industrial Organizational Psychologist. New Orleans, LA.

Ericsson, K. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson et al., (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 683–703). New York, NY: Cambridge University.

Fleenor, J., Taylor, S., & Chappelow, C. (2008). *Leveraging the impact of 360-degree feedback*. New York, NY: Wiley.

Fredrickson, B. L., & Losada, M. (2005). Positive affect and the complex dynamics of human flourishing. *American Psychologist, 60,* 678–686.

Gentry, W. A., & Eckert, R. H. (2012). Integrating ILTs and fit into the development of global leaders: A 360-degree approach. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 5,* 226–229.

Gentry, W. A., Hannum, K. M., Ekelund, B. Z., & 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., Yip, J., & Hannum, K. (2010). Self-observer rating discrepancies of managers in Asia: A study of derailment characteristics and behaviors in Southern Confucian Asia. *International Journal of Selection and Assessment, 18,* 237–250.

Goffin, R. D., & Anderson, D. W. (2002, June). Differences in self and superior ratings of performance: Personality provides clues. Paper presented at the Society for Industrial and Organizational Psychology, Toronto, Canada.

Goffin, R. D., & Anderson, D. W. (2007). The self-rater's personality and self-other disagreement in multi-source performance ratings: Is disagreement healthy? *Journal of Managerial Psychology, 22,* 271–289.

Goffin, R. D., & Olson, J. M. (2011). Is it all relative? Comparative judgments and the possible improvement of self-ratings and ratings of others. *Perspective on Psychological Science, 6,* 48–60.

Goldsmith, M. (2009). How to spot the "uncoachables". Retrieved from http://blogs.hbr.org/goldsmith/2009/03/how_to_spot_the_uncoachables.html

Gollwitzer, P. M., & Sheeran, P. (2009). Self-regulation of consumer decision making and behavior: The role of implementation intentions. *Journal of Consumer Psychology, 19,* 593–607.

Gottman, J. M. (1994). *What predicts divorce? The relationship between marital processes and marital outcomes*. Hillsdale, NJ: Erlbaum.

Gregory, J. B., Levy, P. E., & Jeffers, M. (2008). Development of the feedback process within executive coaching. *Consulting Psychology Journal: Practice and Research, 60,* 42–56.

Greguras, G. J., & Robie, C. (1995). A new look at within-rater source inter-rater reliability of 360-degree feedback ratings. *Journal of Applied Psychology, 83,* 960–968.

Grossman, R., & Salas, E. (2011). The transfer of training: What really matters. *International Journal of Training and Development, 15,* 103–120.

Hannah, S., Avolio, B. J., Luthans, F., & Harms, P. (2008). Leadership efficacy: Review and future directions. *Leadership Quarterly, 19,* 669–692.

Harris, L. S., & Kuhnert, K. W. (2007). Looking through the lens of leadership: A constructive developmental approach. *Leadership & Organization Development Journal, 29,* 47–67.

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

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

Hirsh, J. B., & Inzlicht, H. (2008). The devil you know: Neuroticism predicts neural response to uncertainty. *Psychological Science, 19,* 962–967.

Hoffman, B. J., & Baldwin, S. P. (2012). Modern managerial assessment: A comparison of assessment centers and multisource feedback. In G. Thornton, & N. Povah (Eds.), *Assessment centers and global talent management* (pp. 143–162). Burlington, VT: Gower.

Hoffman, B. J., Gorman, C. A., Blair, C. A., Meriac, J. P., Overstreet, B., & Atchley, E. K. (2012). Evidence for the effectiveness of an alternative multisource performance rating methodology. *Personnel Psychology, 65,* 531–563.

Hoffman, B. J., Lance, C. E., Bynum, B., & Gentry, W. (2010). Rater source effects are alive and well after all. *Personnel Journal, 63,* 119–151.

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

Hofstede, G., & McRae, R. R. (2004). Personality and culture revisited: Linking traits and dimensions of culture. *Cross-Cultural Research, 38,* 52–88. doi:10.1177/1069397103259443

Hogan, R. (2007). *Personality and the fate of organizations*. Hillsdale, NJ: Erlbaum.

Hogan, R., & Kaiser, R. (2005). What we know about leadership. *Review of General Psychology, 9,* 169–180.

Holt, K., & Seki, K. (2012). Global leadership: A developmental shift for everyone. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 5,* 198–217.

Hooijberg, R., & Lane, N. (2009). Using multisource feedback coaching effectively in executive education. *Academy of Management Learning and Education, 8,* 483–493.

Hyman, D. J., Pavlik, V. N., Taylor, W. C., Goodrich, G. K., & Moye, L. (2007). Simultaneous versus sequential counseling for multiple behavioral change. *Archives of Internal Medicine, 167,* 1152–1158.

Ilgen, D., & Davis, C. (2000). Bearing bad news: Reactions to negative performance feedback. *Applied Psychology: An International Review, 49,* 550–565.

Inceoglu, I., & Externbrink, K. (2012, April). Leadership development: Who knows best how well the highflyers perform? Paper presented at the 27th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Joo, B. K. (2005). Executive coaching: A conceptual framework from an integrative review of research and practice. *Human Resource Development Review, 4,* 134–144.

Jordan, A., & Audia, P. (2012). Self-enhancement and learning from performance feedback. *Academy of Management Review, 37,* 211–231.

Kaiser, R., & Overfield, D. (2011). Strengths, strengths overused, and lopsided leadership. *Consulting Psychology Journal: Practice and Research, 63,* 89–109.

Kaiser, R. B., & Craig, S. B. (2005). Building a better mouse trap: Item characteristics associated with rating discrepancies in 360-degree feedback. *Consulting Psychology Journal: Practice and Research, 57,* 235–245.

Kaplan, R. E., & Kaiser, R. B. (2009). Stop overdoing your strengths. *Harvard Business Review, 87,* 100–103.

Klockner, K., & Hicks, R. E. (2008). My next client: Understanding the Big Five and positive personality dispositions of those seeking psychological support interventions. *International Coaching Psychology Review, 3,* 148–163.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback theory. *Psychological Bulletin, 119,* 254–284.

Lally, P., Van Jaarsveld, C., Potts, H., & Wardle, J. (2009). How are habits formed: Modeling habit formation in the real world. *European Journal of Social Psychology, 1009,* 998–1009.

Lance, C. E., Hoffman, B. J., Gentry, W., & Baranik, L. E. (2008). Rater source factors represent important subcomponents of the criterion construct space, not rater bias. *Human Resources Management Review, 18,* 223–232.

Law, K. S., Wong, C., & Song, L. J. (2004). The construct and criterion validity of emotional intelligence and its potential utility for management studies. *Journal of Applied Psychology, 89,* 483–496.

Lawton, R., Conner, M., & McEachan, R. (2009). Desire or reason: Predicting health behaviors from affective and cognitive attitudes. *Health Psychology, 28,* 56–65.

Le, H., OH, I. S., Shaffer, J. A., & Schmidt, F. L. (2007). Implications of methodological advances for the practice of personnel selection: How practitioners benefit from recent developments in meta-analysis. *Academy of Management Perspectives, 21,* 6–15.

Lebreton, J. M., Burgess, R. D., Kaiser, R. B., Atchley, E. K., & James, L. R. (2003). The restriction of variance hypothesis and interrater reliability and agreement: Are ratings from multiple sources really dissimilar? *Organizational Research Methods, 6,* 80–112.

Lehman, B. J., & Conley, K. M. (2010). Momentary reports of social-evaluative threat predict ambulatory blood pressure. *Psychological Science and Personality Science, 1,* 51–56. doi:10.1177/1948550609354924

Leonardelli, G. J., Herman, A. D., Lynch, M. E., & Arkin, R. M. (2003). The shape of self-evaluation: Implicit theories of intelligence and judgments of intellectual ability. *Journal of Research in Personality, 37,* 141–168.

Liao, S. C., Hunt, E. A., & Chen, W. Comparison between inter-rater reliability and inter-rater agreement in performance assessment. *Annals of the Academy of Medicine, Singapore*. *39,* 613–618.

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

London, M., & Smither, J. W. (2002). Feedback orientation, feedback culture and the longitudinal performance management process. *Human Resource Management Review, 12,* 81–100.

Losada, M., & Heaphy, E. (2004). The role of positivity and connectivity in the performance of business teams: A nonlinear dynamics model. *American Behavioral Scientist, 47,* 740–765.

Lozano, L. M., Garcia-Cueto, E., & Muniz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *European Journal of Research Methods for the Behavioral and Social Sciences, 4,* 73–79.

Martin, H. J. (2010). Workplace climate and peer support as determinants of training transfer. *Human Resource Development Quarterly, 21,* 87–104.

Mashihi, S., & Nowack, K. (2011). *Clueless: Coaching people who just don't get it*. Santa Monica, CA: Envisia Learning, Inc.

Morgeson, F. P., Mumford, T. V., & Campion, M. 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,* 3, 196–209.

Mund, M., & Mitte, K. (2011). The costs of repression: A meta-analysis on the relation between repressive coping and somatic diseases. *Health Psychology*, Advance online publication. doi:10.1037/a0026257

Murphy, K. R. (2008). Explaining the weak relationship between job performance and ratings of job performance. *Industrial and Organizational Psychology, 1,* 148–160.

Murphy, K. R., Cleveland, J. N., & Mohler, C. (2001). *Reliability, validity and meaningfulness of multisource ratings*. In D. Bracken, C. Timmreck, and A. Church (Eds.), Handbook of multisource feedback (pp. 130–148). San Francisco, CA: Jossey-Bass.

Nieman-Gonder, J., Metlay, W., Kaplan, I., & Wolfe, K. (2006, May). The effect of rater selection on rating accuracy. Poster presented at the 21st Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

Nowack, K. (1992). Self-assessment and rater assessment as a dimension of management development. *Human Resources Development Quarterly, 3,* 141–155.

Nowack, K. (1999). 360 Degree feedback. *Intervention: 50 performance technology tools* (pp. 34–46). San Francisco, CA, Jossey-Bass, Inc.

Nowack, K. (2002). Does 360-degree feedback negatively affect company performance: Feedback varies with your point of view. *HR Magazine*, *47,* 6.

Nowack, K. (2003). Executive coaching: Fad or future? *California Psychologist, 36,* 16–17.

Nowack, K. (2005, March). Longitudinal evaluation of a 360 degree feedback program: Implications for best practices. Paper presented at the 20th Annual Conference of the Society for Industrial and Organizational Psychology, Los Angeles.

Nowack, K. (2009). Leveraging multirater feedback to facilitate successful behavioral change. *Consulting Psychology Journal: Practice and Research, 61,* 280–297.

Nowack, K. (2011). *Leveraging 360-degree feedback using an online goal setting and tracking system*. Unpublished manuscript, Envisia Learning, Inc., Los Angeles, CA.

Nowack, K., Hartley, J., & Bradley, W. (1999). Evaluating results of your 360-degree feedback intervention. *Training and Development, 53,* 48–53.

Olivero, G., Bane, D., & Kopelman, R. (1997). Executive coaching as a transfer of training tool: Effects on productivity in a public agency. *Public Personnel Management, 26,* 461–469.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (2008). No new terrain: Reliability and construct validity of job performance ratings. *Industrial and Organizational Psychology, 1,* 174–179.

Ostroff, C., Atwater, L., & Feinberg, B. (2004). Understanding self-other agreement: A look at rater and ratee characteristics, context and outcomes. *Personnel Psychology, 57,* 333–375.

Pearce, C. L., Sims, H. P., Cox, J. F., Ball, G., Schnell, E., Smith, K. A., & Trevino, L. (2003). Transactors, transformers, and beyond. *Journal of Management Development, 22,* 273–308.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica, 104,* 1–15.

Prochaska, J. O., & Velicer, W. F. (1997). The transtheoretical model of health behaviour change. *American Journal of Health Promotion, 12,* 38–48.

Quast, L. N., Center, B. A., Chung, C., Wohkittel, J. M., & Vue, B. (2011, February). Using multi-rater feedback to predict managerial career derailment: A model of self-boss rating patterns. Paper presented at the 2011 Academy of Human Resource Development International Research Conference in the Americas, Chicago, IL.

Quast, L. N., Wohkittel, J. M., Chung, C., & Center, B. A. (2011, December). Patterns of self-other rating discrepancies and predictions of managerial career derailment: Comparing Asia to the United States. Paper presented at the 10th International Conference of the Academy of HRD (Asia chap.) International Research Conference, Kuala Lumpur, Malaysia.

Rehbine, N. (2007). *The impact of 360-degree feedback on leadership development.* (Unpublished doctoral dissertation). Capella University, Minneapolis, MN.

Reilly, R. R., Smither, J. W., & Vasilopoulos, N. L. (1996). A longitudinal study of upward feedback. *Personnel Psychology, 49,* 599–612.

Rhodes, R. E., Plotnikoff, R. C., & Courneya, K. S. (2008). Predicting the physical activity intention-behavior

of adopters and maintainers using three social cognition models. *Annals of Behavioral Medicine, 36,* 244–252.

Robie, S., Kaster, K., Nilsen, D., & Hazucha, J. (2000). *The right stuff: Understanding cultural differences in leadership performance*. Unpublished manuscript, Personnel Decisions, Inc., Minneapolis, MN.

Roch, S. G., Sternburgh, A. M., & Caputo, P. M. (2007). Absolute vs. relative performance rating formats: Implications for fairness and organizational justice. *International Journal of Selection and Assessment, 15,* 302–316.

Rock, D. (2008). SCARF: A brain-based model for collaborating with and influencing others. *Neuroleadership Journal, 1,* 44–52.

Rutledge, T. (2006). Defensive personality effects on cardiovascular health: A review of the evidence. In D. Johns (Ed.), *Stress and its impact on society* (pp. 1–21). Hauppauge, NY: Nova Science Publishers.

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

Schwartz, R. M., Reynolds, C. F., III., Thase, M. E., Frank, E., Fasiczka, A. L., & Haaga, D. A. F. (2002). Optimal and normal affect balance in psychotherapy of major depression: Evaluation of the balanced states of mind model. *Behavioural and Cognitive Psychotherapy, 30,* 439–450.

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

Sherman, S., & Freas, A. (2004). The Wild West of executive coaching. *Harvard Business Review, 82,* 82–90.

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

Siefert, C., Yukl, G., & McDonald, R. (2003). Effects of multisource feedback and a feedback facilitator on the influence of behavior of managers toward subordinates. *Journal of Applied Behavior, 88,* 561–569.

Smither, J., Brett, J., & Atwater, L. (2008). What do leaders recall about multi- source feedback? *Journal of Leadership and Organization Studies, 14,* 202–218.

Smither, J., London, M., Flautt, R., Vargas, Y., & Kucine, I. (2003). Can working with an executive coach improve multisource feedback ratings over time? A quasi-experimental field study. *Personnel Psychology, 56,* 23–44.

Smither, J., London, M., & Reilly, R. (2005). Does performance improve following multisource feedback? A theoretical model, meta-analysis, and review of empirical findings. *Personnel Psychology, 58,* 33–66.

Smither, J., & Walker, A. G. (2004). Are the characteristics of narrative comments related to improvement in 360-degree feedback ratings over time? *Journal of Applied Psychology, 89,* 575–581.

Smither, J., Walker, A., & Yap, M. (2004). An examination of web based versus paper and pencil upward feedback ratings. *Educational and Psychological Measurement, 64,* 40–61.

Taylor, P., Taylor, H., & Russ-Eft, D. (2009). Transfer of management training from alternate perspectives. *Journal of Applied Psychology, 94,* 104–121.

Thach, E. (2002). The impact of executive coaching and 360-feedback on leadership effectiveness. *Leadership and Organization Development Journal, 23,* 205–214.

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

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.

Viswanathan, M., Bergen, M., Dutta, S., & Childres, T. (1996). Does a single response category in a scale completely capture a response? *Psychology and Marketing, 13,* 457–479.

Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2002). The moderating influence of job performance dimensions on convergence of supervisory and peer ratings of job performance: Unconfounding construct-level convergence and rating difficulty. *Journal of Applied Psychology, 87,* 345–354.

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

Wanguri, D. M. (1995). A review, an integration, and a critique of cross-disciplinary research on performance appraisals, evaluations, and feedback: 1980–1990. *The Journal of Business Communication, 32,* 267–293.

Weng, L. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement, 64,* 956–972.

Wikman, A., & Warneryd, B. (1990). Measurement errors in survey questions: Explaining response variability. *Social Indicators Research, 2,* 199–212.

Wilson, A. E., & Ross, M. (2001). From chump to champ: People's appraisals of their earlier and present selves. *Journal of Personality and Social Psychology, 80,* 572–584.

Woehr, D. J. (2008). On the relationship between job performance ratings and ratings of job performance: What do we really know? *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1,* 161–166.

Woo, S., Sims, C., Rupp, D., & Gibbons, A. (2008). Development engagement within and following developmental assessment centers: Considering feedback favorability and self-assessor agreement. *Personnel Psychology, 61,* 727–759.

Yukl, G., Gordon, A., & Taber, T. (2002). A hierarchical taxonomy of leadership behavior: Integrating a half century of behavior research. *Journal of Leadership and Organizational Studies, 9,* 15–32.

Received June 19, 2012
Latest revision received August 13, 2012
Accepted August 14, 2012 ■