

Feedback that fuels **success**<sup>TM</sup>  
**Whitepaper**



## When Does 360-Degree Feedback Create Behavior Change? And How Would We Know It When It Does?

David W. Bracken  
Dale S. Rose



510.463.0333  
[www.3dgroup.net](http://www.3dgroup.net)

# When Does 360-Degree Feedback Create Behavior Change? And How Would We Know It When It Does?

David W. Bracken · Dale S. Rose

© Springer Science+Business Media, LLC 2011

**Abstract** 360-degree feedback has great promise as a method for creating both behavior change and organization change, yet research demonstrating results to this effect has been mixed. The mixed results are, at least in part, because of the high degree of variation in design features across 360 processes. We identify four characteristics of a 360 process that are required to successfully create organization change, (1) relevant content, (2) credible data, (3) accountability, and (4) census participation, and cite the important research issues in each of those areas relative to design decisions. In addition, when behavior change is created, the data must be sufficiently reliable to detect it, and we highlight current and needed research in the measurement domain, using response scale research as a prime example.

**Keywords** 360-Degree feedback · Behavior change · Organization change · Multisource feedback · Human resource systems

## Introduction

The prevalence and popularity of 360-degree feedback has had the effect of causing research and viewpoints to be published at an ever-increasing pace. This, in turn, has created a need for summaries in the forms of books

(Edwards and Ewen 1996; Bracken et al. 2001a; Lepsinger and Lucia 1998, 2009; Tornow and London 1998; Waldman and Atwater 1998; Fleenor et al. 2008), meta analyses on behavior change (Kluger and DeNisi 1996; Smither et al. 2005), meta analyses on rater behavior (Conway and Huffcutt 1997; Heidemeier and Moser 2009; Harris and Schaubroeck 1988), research reviews (Craig and Hannum 2006; Morgeson et al. 2005), many articles/book chapters (e.g., Bracken 1994, 1996; Bracken et al. 2001b; Church et al. 2002; Nowack 2009), and benchmark studies (3D Group 2002, 2004, 2009).

The purpose of this article is to (1) review the design elements within 360-degree feedback processes that facilitate behavior change, (2) identify the conditions needed to detect those changes when they do occur, and (3) bring to the forefront those elements and conditions related to using 360 processes to create behavior change that are most important and in need of future research. Our intent is to ground our conclusions in established research where applicable and to suggest future research where there are glaring omissions in the knowledge base. In some cases, controlled research is missing but industrywide best practices are virtually ubiquitous.

## 360-Degree Feedback and Behavior Change

360-degree feedback *can* create behavior change under the right circumstances (cf. Goldsmith and Underhill 2001; Goldsmith and Morgan 2004; Smither et al. 2005). Yet there are many studies that seem to indicate that 360-degree feedback processes sometimes create no measurable change (Siefert et al. 2003), and, at times, may actually have negative effects (Pfau and Kay 2002). We propose that it is meaningless to make any kind of blanket

---

D. W. Bracken (✉)  
OrgVitality, LLC, PO Box 88491, Atlanta, GA 30356, USA  
e-mail: davidbracken@orgvitality.com

D. S. Rose  
3D Group, Emeryville, CA, USA

statement about the effectiveness of 360-degree feedback in creating behavior change since 360-degree feedback processes vary so widely (3D Group 2002, 2004, 2009). At the conclusion of their meta analysis, Smither et al. (2005) state, “it is time for researchers and practitioners to ask, ‘Under what conditions and for whom is multisource feedback likely to be beneficial?’ (rather than asking ‘Does multisource feedback work?’)” (p. 60). We agree.

360-degree feedback is an extremely complex process that requires dozens of nuanced decisions in its design and implementation. The relative effectiveness in creating behavior change depends on those decisions, and a taxonomy of those decisions is required to systematically guide the design and implementation process in an informed manner. Since the early 1990s, a number of comprehensive taxonomies have been published that identify the various dimensions of 360-degree feedback processes to consider during design (e.g., Bracken et al. 2001a, 2001b; Bracken and Timmreck 2001b; Morgeson et al. 2005; Nowack 2009). The list of possible design elements is large, and some design decisions and the resulting practices are so powerful as to cause a total 360-degree feedback process to succeed or fail in a multiplicative fashion.

### Creating Sustainable Behavior Change

Our focus in this article is *not* on creating behavior change with a sample of one person, but rather how to best design a 360-degree feedback process that can be most effective across a large number of participants within a workplace system or subsystem. A great deal of effort has been expended in identifying individual-level variables that can enhance or hinder behavior change, including personality (Bono and Colbert 2005), self–other discrepancies (Brett and Atwater 2001; Atwater and Brett 2005; Heidemeier and Moser 2009; Atwater et al. 2009), openness to feedback (Smither et al. 2005), cultural differences (Gillespie 2005; Shipper et al. 2007), etc. We assume that many individual-level characteristics are resistant (or impossible) to change, and, are therefore factors that might be of interest to a coach in helping someone work through their individual results, but of less use when designing practices for larger populations.

When practitioners consider each 360-degree feedback design decision for its individual effect on the success and sustainability of behavior change, which in turn is compounded by all the possible interactions, the task becomes quite daunting. We hope to bring some type of focus to this list of possible factors by prioritizing the few most critical design factors, the relevant existing research, and, most significantly, some related research questions that might help to move the practice forward in considering how to affect behavior change.

In discussing ways to triage the potential list of critical design factors in need of research, we came up with the following scenario to help frame the discussion:

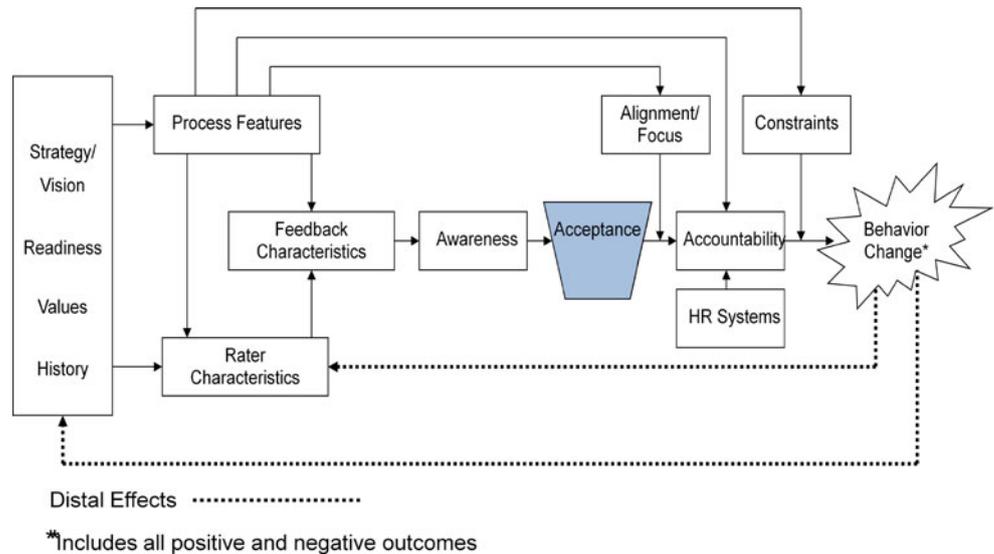
Imagine the organization you are working with (either as an internal or external consultant) has adopted your recommendation to implement a 360-degree feedback process with the objective of creating aligned sustainable behavior change across senior leadership (director and above). The vice president of human resources (your primary client and sponsor) tells you that you have the authority to design most aspects of the feedback process but he or she wants you to list the four design features that you feel are critical to achieving his or her goal of creating sustainable behavior change across the organization. He or she states that each of these four is fully supported in planning, implementation, and budgeting but he or she suggests you choose wisely because you are held accountable for the success of the project in terms of its effect on behavior change.

### A Systems View of Behavior Change

With our combined 40+ years of experience implementing 360-degree feedback programs and conducting best practices research in this area, we have put ourselves in the position of the consultant needing to select design features for that are most likely to increase the probability of (a) aligned sustainable behavior change across the most leaders (Bracken et al. 2001b) and (b) sustaining the 360 process for a period of time that allows for repeated administration for the foreseeable future. We have built our recommendations on a systems model first articulated by Bracken and Timmreck (2001a). This model captures the critical systemwide factors that drive the two objectives in our fictitious case (see Fig. 1).

This systems view of a 360-degree feedback process is based on the proposition that aligned sustainable behavior change is the ultimate definition of validity (Bracken et al. 2001b), a definition of validity that is unique to 360 processes that acknowledges that reliable/valid measurement is necessary but not sufficient. The keystone event in this model is Acceptance, i.e., the point at which the ratee is willing to accept that the feedback is accurate and worthy of consideration in guiding future behavior. This is the point at which the ratee actually decides to change. The elements of the model to the left of Acceptance define the elements that must be designed and implemented in such a way so as to maximize the probability of Acceptance, reinforcing that Awareness alone is not sufficient (Nowack 2009; Church et al. 2002).

**Fig. 1** A systems view of 360 feedback (adapted from Bracken and Timmreck 2001a)



The elements of the model to the right of Acceptance indicate that the journey between acceptance and sustained behavior change is not necessarily a short direct path. It includes considerations such as how to motivate the participant, opportunities to involve others (boss, coach, and raters) in the post feedback phase, possible barriers (constraints) to behavior change, and integration with other HR systems.

The model makes the assumption (but not the requirement) that the process is repeated and that the validity of the process is affected by prior administrations and the experiences of the participants (raters and ratees). Each time a rater, for example, has an experience in a 360 process, it becomes part of the context for their next 360 participation (for good or bad), whether it is the same process or some different 360 process. These experiential effects are labeled distal factors because they have a delayed impact, in contrast to proximal factors that have immediate results (Bracken et al. 2001b). In that way, the system and its participants evolve over time as they learn how to provide better feedback, how to use it more productively, and, maybe most importantly, whether anyone (participant, management, HR, or the organization at large) really cares.

**Critical Design Factors**

Considering the possible antecedents to creating sustainable behavior change and a sustainable 360 process, we have identified four critical design factors: (1) relevant content, (2) credible data, (3) accountability, and (4) census (organizationwide) participation. For each of these, we share our reasoning, relevant research, and some important research questions. Table 1 summarize the main design features (decision points), some associated research questions, and a

sampling of existing research. Clearly, further research is needed in all of these areas, and, in some cases, research is almost nonexistent.

**Relevant Content**

An American football commentator (though it can also apply to the “other” football) recently remarked, “There is nothing worse than having a big, fast player running at full speed in the wrong direction.” This is the alignment part of our definition of success, i.e., ensuring that leaders are in synch with the values and leadership competencies that are uniquely defined by their organization. This is found in the left edge of the model (see Fig. 1), and relates directly to the corresponding need to design and use instruments that are tailored to the specific business and people strategies of each organization. Although it is possible to find good approximations with standardized tools, the best fit comes from a custom survey (Church et al. 2002).

We also understand that the use of standard versus custom surveys has been a passionately debated topic, but not one that is rich in research to support the positions of either camp. One basic problem is that there is little consensus as to what “success” looks like. Two primary choices for criteria for success would be (a) supports reliable measurement and (b) creates sustainable behavior change.

Given the overlap in methods between 360 processes and other assessment methods, it is worth considering lessons learned from research in practice areas where data are generated by someone other than the target person. Such practice areas include assessment centers, employee surveys, and performance appraisals. Borrowing from research in these related fields was especially useful when

**Table 1** Sample design decisions related to four essentials in effective 360-degree feedback processes

Design factor (choice)	Research questions	Sample-related research/articles
1. Relevant content		
Instrument Content	<ul style="list-style-type: none"> <li>• How do standardized versus custom instruments compare in areas such as:               <ul style="list-style-type: none"> <li>· Response rates?</li> <li>· Reliability?</li> <li>· Acceptance of results by ratees?</li> <li>· Behavior change?</li> <li>· Sustainability?</li> </ul> </li> </ul>	Heidemeier and Moser (2009)
Languages	<ul style="list-style-type: none"> <li>• How is reliability affected by using only English in other cultures/nations?</li> <li>• How well do leadership models hold up across cultures?</li> </ul>	Gillespie (2005) and Heidemeier and Moser (2009)
Write in comments	<ul style="list-style-type: none"> <li>• What types/formats of comments are most useful in creating positive behavior change?</li> </ul>	Smither and Walker (2004), Rose et al. (2004), and Smither et al. (2004)
2. Credible data		
Item design	<ul style="list-style-type: none"> <li>• What type (behavioral, trait, competency, and dimension) of item is best?</li> </ul>	Heidemeier and Moser (2009)
Response scale	<ul style="list-style-type: none"> <li>• What are the best response scales to reduce leniency, improve reliability, demonstrate validity, and raise acceptance levels?</li> </ul>	Heidemeier and Moser (2009), Bracken and Paul (1993), Kaiser and Kaplan (2006), and English et al. (2009)
Rater selection	<ul style="list-style-type: none"> <li>• Which rater sources provide the best feedback?</li> </ul>	Mount et al. (1998)
Number of raters	<ul style="list-style-type: none"> <li>• What is the optimal number of raters for each rater type?</li> </ul>	Greguras and Robie (1998)
Rater training	<ul style="list-style-type: none"> <li>• What methods of rater training are most effective (cost/benefit)?</li> </ul>	Antonioni and Woehr (2001)
Frequency	<ul style="list-style-type: none"> <li>• What is the optimal time between reassessing leaders?</li> </ul>	Smither et al. (2005)
3. Accountability		
Role of a coach	<ul style="list-style-type: none"> <li>• Does a coach help in creating behavior change?</li> </ul>	Smither et al. (2003), Smither et al. (2005), and Thach (2002)
Role of boss	<ul style="list-style-type: none"> <li>• What are the effects (good or bad) when raters and ratees know the boss will have access to feedback data?</li> </ul>	London and Smither (2002), Yukl and Lepsinger (1995), and Atwater and Waldman (1998)
Integration with HR systems	<ul style="list-style-type: none"> <li>• What are the perceptions of raters, ratees and the organization if 360 data feeds into HR processes?</li> </ul>	DeNisi and Kluger (2000), Waldman et al. (1998), and Heidemeier and Moser (2009)
Reviewing results with raters	<ul style="list-style-type: none"> <li>• Is behavior change more likely if ratees engage raters in reviewing and clarifying feedback?</li> <li>• Does the process of discussing results improve rater performance in subsequent administrations?</li> </ul>	Goldsmith and Morgan (2004) Unknown
4. Census (organizationwide) participation		
Required participation	<ul style="list-style-type: none"> <li>• Is there a benefit of perceived and real fairness if all managers must participate?</li> </ul>	Unknown
Support systems	<ul style="list-style-type: none"> <li>• What are the most effective ways to support raters, ratees, and users in higher volume situations (e.g., training, coaching, instrument formats, and technologies)?</li> </ul>	Smither et al. (2004); Bracken et al. (2001b)

the field of 360-degree feedback was relatively new with little published research of its own.

We bring this point up here in the context of standard versus custom models (and instruments), but note limitations to solutions used in other processes, such as assessment centers. Just consider differences in the data generators, i.e. the raters. In assessment centers, raters are

carefully selected, trained, motivated to perform well, practiced over repeated administrations, operate in an isolated environment, work to reach consensus, and have never met the ratee before the assessment. We usually say exactly the opposite of most raters in a 360-degree feedback process. One way we believe we can increase motivation and engagement is to use custom instruments that

have meaning and relevance. The process also creates an opportunity to reinforce organizational messages (e.g., values and desired leadership behaviors) that are less important in an assessment center because they touch a very small unique population of raters.

### Credible Data

Credibility comes from both reliable data (as scientists may define it) and what augments the perception of “good data” in the eyes of all stakeholders. There are potentially multiple users of the data depending on the purpose, including feedback recipients, feedback providers, coaches, managers (bosses), other managers (e.g., hiring managers in other parts of the business), and human resources (on behalf of the organization). The reliability of the data (real and perceived) is doubly important because it touches on some of the other issues we address here in our wish list of features, as well as the topic of measurement embedded in the discussion of detecting change elsewhere in this paper.

Real and perceived data credibility come from a number of sources, including: (1) having a sufficient number of raters selected (and respond), (2) selecting raters who have sufficient opportunity to observe the ratee, (3) having the feedback recipient choose the raters with manager approval, (4) an instrument that is professionally constructed with clear behavioral items, (5) an instrument that does not confuse or trick the rater by using randomization and reverse wording, (6) using a rating scale that is relevant, clear, and reduces rating errors (e.g., leniency and halo), and (7) having rater training. If our hypothetical VP of human resources buys our request for credible data, we will be requesting each of these design features. Each on its own is a rich area of research, and some also contribute to reliable measurement, as discussed later.

### Accountability

Moving to the right side of Acceptance on Fig. 1, accountability is a key consideration in getting from Awareness/Acceptance to sustainable behavior change. This concern was identified early in the evolution of 360-degree feedback, most notably by London et al. (1998) where they labeled Accountability as the “Achilles’ heel of multisource feedback.”

Accountability is a difficult construct to observe and measure, which may in turn make it a challenging but fruitful topic for more research. For example, can we attribute some aspect of externally driven accountability to findings of increased likelihood of behavior change when results are shared with raters (Walker and Smither 1999;

Goldsmith and Underhill 2001; Goldsmith and Morgan 2004), ratees have coaches (Smither et al. 2005), and/or involved their boss in using results (Hazucha et al. 1993)? The fact is that many practitioners see accountability as a factor in influencing process design, though there are many other forces at work in these practices. We know of no research that tries to separate feelings of accountability created by certain design features, and then relate them to outcomes, such as behavior change, action planning activity, and so on.

### Impact of Follow-Up

Let us take this opportunity to look deeper into one such design factor, i.e., meeting with direct reports (and possibly other raters). Marshall Goldsmith (Goldsmith and Underhill 2001; Goldsmith and Morgan 2004) has been publishing data on this effect and incorporating it into his consulting for many years. The data are compelling, spanning multiple corporations, different development programs, and global environments. In the Goldsmith and Morgan (2004) study of more than 86,000 raters and more than 11,000 leaders, responses to the degree of leader “follow-up” reported by the raters are highly predictive of perceived change in leader effectiveness. In fact, not following up with raters appears to be a good predictor of *decreased* effectiveness for a high percentage of cases. Note that Goldsmith typically does not measure actual behavior change, but instead reports of whether a leader has changed, and in what direction. If it were just Marshall reporting this, we might withhold judgment, but the effect has intuitive appeal and is confirmed by our experience, and our own observations. The magnitude and consistency of the effect is also compelling. Goldsmith and Morgan (2004) also point out that the “follow-up” they refer to is with coworkers, not with coaches, though Smither et al. (2005) include having a coach as a moderate predictor of behavior change.

“Follow-up” is a complex process that creates many dynamics. One dynamic is accountability, and that may well operate when the interaction is with a coach. But following up with raters has many other potential components and benefits, such as: (1) demonstrates the value of feedback and a commitment to change from the ratee, (2) provides raters an opportunity to further clarify the meaning (intention) of the feedback (see Alignment/Focus in Fig. 1), (3) engages raters in the action-planning process, including getting their support (a form of rater accountability), and (4) makes raters better observers of behavior moving forward (calibrates their use of the instrument).

Discussing results with others is one of a few powerful design factors that has the potential to override, at least to some extent (yes, this is another research question), individual

resistance to change. For example, if a 360 process was designed such that a ratee would receive a significant year-end bonus if they change their behavior on one or two stated goals derived from 360-degree feedback results, might this design feature shift even a reluctant individual's motivation? Whereas only 33% of companies require participants to use 360 results to create a development plan with their manager (3D Group 2009), some firms, including Center for Creative Leadership (Dalton 1997), often actively *discourage* this approach. Thus, it seems that instead of asking “Does 360-degree feedback create change?” the better question may be “Do we actually want to use 360-degree feedback to create change?” and if so, “Are organizations willing to do what it takes to design systems with change as the goal?”

### Role of the Boss

We suspect that when practitioners see Fig. 1 and the Accountability box, their mental model jumps to the role of the boss in making sure that the ratee follows through on their action planning commitments, often in the form of a development plan. We (and others) do indeed see the boss as being a critical player for many reasons (Atwater and Waldman 1998; Yukl and Lepsinger 1995), and are dismayed when “experts” condone or explicitly recommend that the boss be excluded from the equation in the use of the feedback. There is a catch here, though, because we also believe that the boss is a representative of the organization, and therefore the involvement of the boss is the same as having the organization co-own the feedback.

If we, the consultants, then receive approval for Accountability from the VP of HR as one of our four design factors, we are recommending features such as: (1) provide each ratee with a minimum of one coaching (or facilitated feedback) session, (2) require that the boss is involved in the action-planning process that is in turn integrated into the development plan in the performance management system, (3) require that the ratee review their results and action plan with their direct reports and other raters (the coach and boss can provide guidance for how this should be done, which requires training for bosses and coaches), and (4) have consistent policies and practices for use of 360 results in HR processes, such as staffing, succession planning, high-potential selection/development, training/development selection, performance management, and compensation.

## Census (Organizationwide) Participation

### Benefits of Census

Having all leaders/managers in an organization (company, division, region, etc.) participate in the 360 process has multiple potential benefits:

- Maximizes the message, i.e., exposes all employees to expectations for leaders
- Communicates an expectation of accountability for leadership behavior
- Creates efficiencies and potential cost savings for technology and training
- Creates a climate of consistency and fairness for all stakeholders
- Provides the organization with data on all leaders, which is, in turn:
  - Useful for decisions (e.g., succession planning, staffing, and high-potential identification)
  - Serves as an early warning mechanism for dysfunctional managers
  - Builds an internal norm database and data for research (reliability and validity)
- Improves rater ability and reliability
  - Better observers of behavior
  - Increased honesty
  - Better users of the instrument through experience and calibration (learning how their input is interpreted and used)

### Creating Organization Change

Census participation is a central requirement of 360 processes that have the objective of creating organization change (Church et al. 2002). In fact, an organizationwide 360 process that is integrated into its culture can be a powerful tool for communicating and instituting change, rapidly touching all members of the organization when new markets, strategies, values, and structures are introduced into the system. We note that Church et al. (2002) also endorse integration with HR systems, as shown in Fig. 1, but with the caveat that the purpose be “development only” (i.e., not used for decision making). We see this as being somewhat misleading because (a) most HR systems require decisions (succession planning, high-potential identification, and staffing), and (b) even “development only” uses require decisions, such as access to training programs, development resources, job assignments, etc.<sup>1</sup>

As seen in Table 1, this is an area where we were unable to find relevant research. This may be because this decision is somewhat philosophical and related to process sustainability than individual behavior change per se. Yet we do think there is a major question as to whether raters do get

<sup>1</sup> The authors have had frequent interaction with Allan Church on this topic over the years, and note that his views and practices have evolved to be more supportive of the use of 360-degree feedback for decision-making purposes.

better at being observers and reporters of behavior when they have repeated exposure to the instrument as well as to a leader, and are motivated by seeing that their feedback is being used constructively.

### The Power of Single Design Elements

We suspect that there are many “sine qua non” process elements besides follow-up with raters that seriously affect the perceived and/or real behavior change in participants when they are absent. These, too, are areas for more research. To an unknown degree, these factors and others are believed to enhance or suppress behavior change and, ultimately, the utility of the 360 process.

If a design factor such as follow-up with raters is as powerful as suggested by Goldsmith’s research, then generalizations regarding the effectiveness of 360-feedback may be moot. Our observations, as noted earlier, are that features that drive change (follow-up with raters, as an example) are rarely required, sometimes encouraged, and sometimes actively discouraged. If a majority of studies in the Smither et al. (2005) meta analysis, again as an example, did not require follow-up, then it should not be surprising that observed behavior change is modest. When follow-up in the form of a coach was present, results were moderately positive. The bottom line is that, before generalizing any studies, we need a very rigorous set of parameters to apply to each 360 process so we know we are examining a process that was actually designed to effect change. Even more important, we need to be clear with organizations about what the minimum requirements are for a 360 process when they say they want to use 360-degree feedback to drive systemwide change. It is hardly reasonable to demand that 360 processes be used to create change, and then not design them with that in mind.

### Detecting Change

Beyond examining the characteristics of a successful 360 process, there is a question of equal magnitude, i.e., when we are successful in creating behavior change, will we know it? That is, can we somehow minimize the probability of Type II error in not detecting behavior change when it does occur?

360-degree feedback processes are susceptible to rating errors because of the potential consequences associated with the results. For example, employees completing surveys may believe (true or not) that their ratings have a significant impact on the ratee’s employment status (Longenecker et al. 1987), though this generalization has inconsistent support (Smith and Fortunato 2008). If/when this effect does occur, employee motivation may shift from

providing accurate ratings to providing ratings that help retain (or expel) a particular ratee. It is also not uncommon for companies to prioritize short surveys over valid surveys, sometimes even ignoring clear data suggesting that extremely short surveys do not predict outcomes as well as longer ones (e.g. Healy and Rose 2003). Associated measurement problems for establishing reliability and validity make it challenging to detect behavior change effects.

Table 1 includes the most important design decisions not only to create behavior change but also *relating to reliable measurement* that allow us to detect change. Each of these is potential areas for future research, including some that have been studied to some extent but need of verification.

### Response Scales

Let us look at one important design variable that has widely varying practices as an example: response scales. The fact is that some response scales are better than others at generating distributions with adequate variability in the responses. Given the importance of variance in determining reliability and validity, the type of response scale is critical to the success of detecting behavior change, a point ignored by many practitioners in our experience. Many studies (Bracken and Paul 1993; Kaiser and Kaplan 2006; Caputo and Roch 2009; English et al. 2009; Heidemeier and Moser 2009) have demonstrated that the response scale can have a major effect on the 360-degree feedback data, and some response scales are indeed preferable. One potentially useful observation from this line of research is that commonly used frequency scales (e.g., never to always) are inferior to others (Bracken and Paul 1993; Kaiser and Kaplan 2006; Heidemeier and Moser 2009). Most of this research has focused on the anchors themselves. Add to that the questions of the optimal number of response choices, odd/even number of choices, and use of midpoint, and the permutations become quite numerous. Clearly, more research is needed to identify the optimal anchor format and number of anchors, perhaps similar to research by Bass et al. (1974), to promote variance and accuracy, which in turn should make detecting effects much easier.

### Interactions

We can easily compound the challenge of isolating the effects of our design choices when we acknowledge that they interact with each other. We worked with one client, for example, who wished to address highly skewed (inflated) ratings by doing two things that have been shown to potentially alleviate this problem, i.e., modify the rating

scale anchors (English et al. 2009) and conduct rater training (Antonioni and Woehr 2001). Their efforts were very successful in changing the response pattern to something closer to a normal distribution but we still do not know whether their success was a main effect for the anchors, the training, or an interaction.

Some design factors affect both of the issues we are discussing in this article, i.e., both creating and detecting behavior change. Consider the research of Greguras and Robie (1998) who document how the number of raters used in each rater category (direct report, peer, and manager) affects the reliability of the feedback, with direct reports being the least reliable and therefore requiring more participation to get a stable result. From a measurement perspective, we know that many 360 administrations permit very small rater samples that are only subsets of a rater group (for example, only half of the direct reports or a small number (three to four) of a large population of peers). In fact, recent evidence suggests that as many as 26% of companies report results when two or fewer respondents provide data for a given group (3D Group 2009). These small samples may be totally inadequate for reliable measurement to occur. Use of microsamples is often justified in the name of cost (hard and soft) savings. At the same time, we expect that the likelihood of behavior change depends, in part, on the acceptance of the feedback as being accurate (see Fig. 1), which we expect to be influenced by both the number of raters who respond and who those raters are (i.e., how familiar they are with the focal person).

## Conclusions

Meta analyses such as the one carried out by Smither et al. (2005) are powerful contributions to our understanding of the complexity of using 360 processes to create behavior change. Yet meta analysis as a methodology has many limitations when examining 360-degree feedback, not the least of which is the need to combine data from multiple 360 processes of unknown quality and with a wide variability of design.

Our taxonomy of existing and future research hopefully can inform researchers and practitioners of the current state of knowledge in many areas of 360 practice, and then guide (and hopefully encourage) future research to help us further enlighten those of us who have to make design and implementation decisions.

One of our fears is that an article such as this will raise (or increase) doubts and cynicism about the efficacy of 360 systems. Of course, our intent is just the opposite. We hope that this article can make the reader aware of some research that might have not been familiar before. Maybe we can

cause some “experts” (including ourselves) to reflect on some areas where we might benefit from a little more uncertainty before we make conclusive statements about what is the “right” answer to a design question. Perhaps we can encourage organizations that use 360-degree feedback for creating behavior change to design their 360 systems with that goal in mind.

## References

- Antonioni, D., & Woehr, D. J. (2001). Improving the quality of multisource rater performance. In D. W. Bracken, C. W. Timmreck, & A. H. Church (Eds.), *The handbook of multisource feedback*. San Francisco: Jossey-Bass.
- Atwater, L. E., & Brett, J. F. (2005). Antecedents and consequences of reactions to developmental 360° feedback. *Journal of Vocational Behavior*, *66*, 532–548.
- Atwater, L. E., & Waldman, D. (1998). Accountability in 360-degree feedback. *HR Magazine*, *43*, 96–104.
- Atwater, L. W., Wang, M., Smither, J. W., & Fleenor, J. W. (2009). Are cultural characteristics associated with the relationship between self and others' ratings of leadership? *Journal of Applied Psychology*, *94*, 876–886.
- Bass, B. M., Cascio, W. F., & O'Connor, E. J. (1974). Magnitude estimations of expressions of frequency and amount. *Journal of Applied Psychology*, *59*(3), 313–320.
- Bono, J., & Colbert, A. (2005). Understanding responses to multisource feedback: the role of core self-evaluations. *Personnel Psychology*, *58*, 171–203.
- Bracken, D. W. (1994). Straight talk about multirater feedback. *Training & Development*, *48*, 44–51.
- Bracken, D. W. (1996). Multisource (360-degree) feedback: surveys for individual and organizational development. In A. I. Kraut (Ed.), *Organizational surveys*. San Francisco: Jossey-Bass.
- Bracken, D. W., & Paul, K. B. (1993, May). *The effects of scale type and demographics on upward feedback*. Paper presented at the Society for Industrial and Organizational Society Annual Conference, San Francisco, CA.
- Bracken, D. W., & Timmreck, C. W. (2001a). Success and Sustainability: a systems view of multisource feedback. In D. W. Bracken, C. W. Timmreck, & A. H. Church (Eds.), *The handbook of multisource feedback*. San Francisco: Jossey-Bass.
- Bracken, D. W., & Timmreck, C. W. (2001b). Guidelines for multisource feedback when used for decision making purposes. In D. W. Bracken, C. W. Timmreck, & A. H. Church (Eds.), *The handbook of multisource feedback*. San Francisco: Jossey-Bass.
- Bracken, D. W., Timmreck, C. W., & Church, A. H. (2001a). *The handbook of multisource feedback*. San Francisco: Jossey-Bass.
- Bracken, D. W., Timmreck, C. W., Fleenor, J. W., & Summers, L. (2001b). 360 degree feedback from another angle. *Human Resource Management*, *40*(1), 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., & Roch, S. (2009, April). *Rating formats and perceptions of performance appraisal fairness*. Paper presented at the Society for Industrial and Organizational Psychology Annual Conference, New Orleans, LA.
- Church, A. H., Walker, A. G., & Brockner, J. (2002). Multisource feedback for organization development and change. In J. Waclawski & A. H. Church (Eds.), *Organization development:*

- a data-driven approach to organizational change. San Francisco: Jossey-Bass.
- 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.
- Craig, S. B., & Hannum, K. (2006). Research update: 360-degree performance assessment. *Consulting Psychology Journal: Research and Practice, 58*, 117–122.
- Dalton, M. A. (1997). When the purpose of using multi-rater feedback is behavior change. In D. Bracken, et al. (Eds.), *Should 360 degree feedback be used only for development purposes?*. Greensboro, NC: Center for Creative Leadership.
- DeNisi, A. S., & Kluger, A. N. (2000). Feedback effectiveness: can 360-degree appraisals be improved? *Academy of Management Executive, 14*, 129–139.
- Edwards, M., & Ewen, A. (1996). *360° feedback: the powerful new tool for employee assessment and performance improvement*. New York: AMACOM.
- English, A. E., Rose, D. S., & McClellan, J. (2009, April). *Rating scale label effects on leniency bias in 360-degree feedback*. Paper presented at the Society for Industrial and Organizational Psychology Annual Conference, New Orleans, LA.
- Fleenor, J. W., Taylor, S., & Chappelow, C. (2008). *Leveraging the impact of 360-degree feedback*. San Francisco: Pfeiffer.
- Gillespie, T. L. (2005). Internationalizing 360-degree feedback: are subordinate ratings comparable? *Journal of Business and Psychology, 19*(3), 361–382.
- Goldsmith, M., & Morgan, H. (2004). Leadership is a contact sport: the “follow-up” factor in management development. *Strategy + Business, 36*, 71–79.
- Goldsmith, M., & Underhill, B. O. (2001). Multisource feedback for executive development. In D. W. Bracken, C. W. Timmreck, & A. H. Church (Eds.), *The handbook of multisource feedback*. San Francisco: Jossey-Bass.
- Greguras, G. J., & Robie, C. (1998). A new look at within-source interrater reliability of 360-degree feedback ratings. *Journal of Applied Psychology, 83*, 960–968.
- 3D Group (2002). *Benchmark study of North American 360-degree feedback practices*. 3D Group Technical Report #8214. Berkeley, CA: Data Driven Decisions, Inc.
- 3D Group (2004). *Current practices in 360-degree feedback*. 3D Group Technical Report #8251. Berkeley, CA: Data Driven Decisions, Inc.
- 3D Group (2009). *Current practices in 360-degree feedback: A benchmark study of North American companies*. 3D Group Technical Report #8326. Berkeley, CA: Data Driven Decisions, Inc.
- Harris, M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology, 41*, 43–62.
- Hazucha, J. F., Hezlett, S. A., & Schneider, R. J. (1993). The impact of 360-degree feedback on management skills development. *Human Resource Management, 32*(2–3), 325–351.
- Healy, M. C., & Rose, D. S. (2003). *Validation of a 360-degree feedback instrument against sales: content matters*. Paper presented at the 18th annual convention of the Society for Industrial and Organizational Psychology, Orlando, FL.
- 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.
- Kaiser, R. B., & Kaplan, R. E. (2006, May). *Are all scales created equal?* Paper presented at the Society for Industrial and Organizational Psychology Annual Conference, Dallas, TX.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254–284.
- Lepsinger, R., & Lucia, A. D. (1998). *The art and science of 360 degree feedback*. San Francisco: Jossey-Bass.
- Lepsinger, R., & Lucia, A. D. (2009). *The art and science of 360 degree feedback* (2nd ed.). San Francisco: Jossey-Bass.
- London, M., & Smither, J. W. (2002). Feedback orientation, feedback culture and the longitudinal performance management process. *Human Resource Management Review, 12*, 81–100.
- London, M., Smither, J. W., & Adsit, D. L. (1998). Accountability: the Achilles’ heel of multisource feedback. *Group and Organization Management, 22*, 162–184.
- Longenecker, C. O., Sims, H. P., & Gioia, D. A. (1987). Behind the mask: the politics of employee appraisal. *The Academy of Management Executive, 1*(3), 183–193.
- 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: Research and Practice, 57*, 196–209.
- Mount, M. K., Judge, T. A., Scullen, S. E., Sytsma, M. R., & Hezlett, S. A. (1998). Trait, rater, and level effects in 360-degree performance ratings. *Personnel Psychology, 51*, 557–569.
- Nowack, K. M. (2009). Leveraging multirater feedback to facilitate successful behavioral change. *Consulting Psychology Journal: Practice and Research, 61*(4), 280–297.
- Pfau, B., & Kay, I. (2002). Does 360-degree feedback negatively affect company performance? Studies show that 360-degree feedback may do more harm than good. What’s the problem? *HRMagazine, 47*(6), 54–60.
- Rose, D. S., Farrell, T., & Robinson, G. N. (2004). *Are narrative comments in 360-degree feedback useful or useless?* Technical Report #8253. Berkeley, CA: Data Driven Decisions, Inc.
- Shipper, F., Hoffman, R. C., & Rotondo, D. M. (2007). Does the 360-feedback process create actionable knowledge equally across cultures? *Academy of Management Learning & Education, 6*(1), 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 Psychology, 88*(3), 561–569.
- Smith, A. F. R., & Fortunato, V. J. (2008). Factors influencing employee intentions to provide honest upward feedback ratings. *Journal of Business and Psychology, 22*, 191–207.
- Smither, J. W., & Walker, A. G. (2004). Are the characteristics of narrative comments related to improvement in multirater feedback ratings over time? *Journal of Applied Psychology, 89*, 575–581.
- Smither, J. W., London, M., Flautt, R., Vargas, Y., & Kucine, I. (2003). Can working with an executive coach improve multi-source feedback ratings over time? A quasi-experimental field study. *Personnel Psychology, 56*, 23–44.
- Smither, J. W., Walker, A. G., & Yap, M. K. T. (2004). An examination of the equivalence of web-based versus paper-and-pencil upward feedback ratings: Rater- and ratee-level analyses. *Educational and Psychological Measurement, 64*(1), 40–61.
- Smither, J. W., London, M., & Reilly, R. R. (2005). Does performance improve following multisource feedback? A theoretical model, meta-analysis and review of empirical findings. *Personnel Psychology, 58*, 33–66.
- Thach, E. (2002). The impact of executive coaching and 360 feedback on leadership effectiveness. *Leadership and Organization Development Journal, 23*, 205–214.

- 
- Tornow, W., & London, M. (1998). *Maximizing the value of 360-degree feedback: a process for successful individual and organization development*. San Francisco: Jossey-Bass.
- Waldman, D. A., & Atwater, L. E. (1998). *The power of 360° feedback*. Houston, TX: Gulf Publishing Company.
- Waldman, D., Atwater, L., & Antonioni, D. (1998). Has 360 feedback gone amok? *Academy of Management Executive*, 12, 86–94.
- Walker, A., & Smither, J. W. (1999). A five-year study of upward feedback: what managers do with their results matters. *Personnel Psychology*, 52(2), 393–423.
- Yukl, G., & Lepsinger, R. (1995). How to get the most out of 360-degree feedback. *Training*, 32, 45–50.