

360 FEEDBACK FROM ANOTHER ANGLE

David W. Bracken, Carol W. Timmreck, John W. Fleenor, and
Lynn Summers

How do you know if a 360 process has been implemented successfully? What are the factors that influence its success? How can those factors be controlled by those responsible for its implementation? Despite its popularity, there has been little effort to build a comprehensive model that addresses these fundamental questions about 360 feedback. In a quest for such a model, we identify a host of key factors organized according to whether they exert their influence proximally or distally. We discuss how each factor contributes to successful implementation. After identifying how design features of a 360 process affect these key factors, we recommend how to enhance the probability of implementing 360 feedback successfully and sustaining the process over time.
© 2001 John Wiley & Sons, Inc.

Introduction

Despite the proliferation of 360 (multisource) feedback processes (where individuals receive anonymous, questionnaire-based feedback from peers, subordinates, supervisors, customers, and other co-workers), the jury is still out on the answers to two questions: (1) How well does it work, and (2) Is it a good idea? And the answer to the second question is very much determined by the answer to the first. These questions become more salient as organizations contemplate using the results for decision-making, such as performance appraisals, pay determinations, succession plans, job placement, or even downsizing (Bracken, Dalton, Jako, McCauley & Pollman, 1997), although we maintain that even “development-only” systems also ultimately depend on “good” information to have any hope of success.

During the past few years, the literature on 360 feedback has expanded rapidly, including books,¹ book chapters,² and special issues of *Human Resource Management* (Tornow, 1993a) and *Group and Organization Management* (Church & Bracken, 1997). The general sense in this proliferation is much like the blind men groping the elephant, each arriving at a different conclusion depending on the “angle.” We do not see any of these offerings providing a comprehensive model of the underlying factors that lead to successful 360 feedback processes.

The reasons behind effective (and ineffective) 360 feedback systems are elusive. In interviews of 27 administrators of multisource processes who had previously indicated that they were using or planning to use the feedback for appraisal purposes, we (Timmreck & Bracken, 1997) found that in more than half of the cases, multisource feedback was work-

The reasons behind effective (and ineffective) 360 feedback systems are elusive.

ing *successfully* in conjunction with performance management. We tended to focus on the “half full” interpretation that multisource feedback for decision making *can* work under the right conditions.

Maxine Dalton, quoted in Lepsinger & Lucia (1997b), points instead to the “half empty” interpretation that some systems do indeed “fail” (although she overstates the failure rate). In reviewing our findings, it was very difficult to determine the reasons for failure (i.e., discontinuation); they were certainly many, diverse, and sometimes unclear, ranging from changes in management, to legal advice, to rebellion in the ranks, just to name a few. We also had a nagging feeling that some of the dead/dying processes probably deserved their fate due to insufficient care in their implementation.

We suggest that the field needs a systematic review of 360 feedback and development of a model that includes the factors contributing to its success, or lack thereof. As Table I indicates, “success” is defined differently depending on the constituent group one is trying to satisfy. Raters, for example, would consider a 360 process successful if ratees appeared to accept the feedback and change their behaviors, if their working relationships with ratees improved, if they were not retaliated against for providing honest feedback, and if the overall feedback climate improved. Similarly, each of the other principal constituencies (ratees, bosses, and the organization itself) has its own definition of the factors that contribute to successful implementation of the process.

While acknowledging these different and sometimes conflicting definitions of “success”, we propose that ultimately it is determined by the organization, and can be defined as a process that:

Creates focused, sustained behavior change and/or skill development in a sufficient number of people so as to result in increased organization effectiveness.

Our definition purposely excludes behavior change targeted at one individual at a time. While it is possible that one person’s 360 feedback could affect organizational effectiveness if that individual were strategically placed, we

have chosen to address what we believe to be the more significant (and challenging) opportunity via system-wide implementations of 360 feedback. We also propose that this definition of success will not be achieved if the organization does not at least consider the success definitions of all the constituent groups.

The complexity of 360 feedback lies in the many ways it can be implemented and the ultimate effect implementation has on the accuracy, usefulness, and acceptance of the feedback (Bracken, 1994; Brutus, Fleenor & London, 1998b; Lepsinger & Lucia, 1997a). We propose that each decision made in the implementation of 360 feedback has implications for the success of the process, and that it is our responsibility to assist users in making informed decisions as to the possible ramifications of their decisions.

With this in mind, we present a new model of successful 360 feedback that can be used by practitioners to design effective (i.e., valid) multisource processes in their organizations. Our bias is to create a model that will withstand the scrutiny of users of 360 feedback for decision making, but also will improve purely developmental processes. Without bringing to bear any new research of our own, we present the entire process as a system with interacting and dynamic elements. By providing a model that is often prescriptive, accompanied by the rationale for recommendations, we hope to equip the practitioner with the tools to make informed decisions during 360 design and implementation.

A New Model of Validity for 360 Feedback

Traditionally, notions of validity have come from relatively controlled, static measurement settings, the most classic of which is pre-employment testing. Standards for testing go to great lengths to standardize the test parameters, focusing on the test environment, a validated instrument, and consistent use of results in decision making. Once a person becomes an employee, the organization might do further assessments, such as an assessment center or some derivative, where it still strives for standardization but introduces a potential source of unreliability by relying on observers/raters to generate the scores.

The complexity of 360 feedback lies in the many ways it can be implemented and the ultimate effect implementation has on the accuracy, usefulness, and acceptance of the feedback.

TABLE I The Success of a 360 Feedback Process as Defined by Its Different Constituents.

Constituent Group	Definitions of Success	Factors Contributing to Success
Rater	<ul style="list-style-type: none"> • Perception that feedback was accepted and ratee behavior improved • Improved working relationship • No negative repercussions • Improved feedback climate 	<ul style="list-style-type: none"> • Anonymity • Ratee accountability • Multiple administrations • Policies and practices communicated • Hotline to report problems
Ratee	<ul style="list-style-type: none"> • Received feedback that was fair, constructive, valid, actionable, credible, reliable • Improved communications • Improved working relationship • Improved feedback climate • Improved performance 	<ul style="list-style-type: none"> • Consistency • Confidentiality • Rater training • Rater accountability (i.e., honesty) • Consistent rater selection • Sufficient number of raters with knowledge (i.e., census) • Valid, reliable instrument • Organization commitment (e.g., resources) • Access to raters for follow-up • Development and action planning
Boss	<ul style="list-style-type: none"> • Access to timely, quality feedback appropriate for use in performance management and/or development • Feedback aligned with organization goals and/or values (i.e., relevant, important) • Improvement in ratee's behavior consistent with the process' objectives • Improvement in ratee's workgroup • Improved feedback climate • Not overly intrusive (time, cost) 	<ul style="list-style-type: none"> • Boss receives data that are clear, actionable, relevant • Boss receives training on proper use of data • Ratee accountability for follow-through • Organization commitment (e.g., resources) • Process is a business priority • Process is administered regularly, consistently • All bosses held accountable • Valid, reliable instrument • Efficient data collection, reporting
Organization	<ul style="list-style-type: none"> • Sustained focused behavioral change in large number of individuals that leads to organizational effectiveness • Viability of the process (sustainable) • Improved feedback climate • Legal defensibility • Supports learning environment 	<ul style="list-style-type: none"> • Top management commitment • System-wide implementation • Resources to develop valid process • Sufficient resources to sustain the process • Valid, reliable instrument • Alignment • Efficient, cost effective data collection, processing and reporting

Some of the traditional definitions of validity imply that they are referring to a measurement “event” using some kind of instrument:

“... a measuring instrument is valid if it does what it is intended to do.” (Nunnally, 1978)

“... the degree to which inferences from scores are justified or supported by evidence.” (SIOP, 1987)

Clearly, these definitions fall short of addressing the many factors that affect the “validity” of a process that might include the use of an “instrument” but ultimately depends upon other sources (e.g., raters) for reliable data. In addition, we maintain that it is insufficient for an instrument to be judged “valid” in a vacuum. An apparently “valid” instrument can be rendered “invalid” if the information it provides is misused, ultimately not doing “what it is intended to do.”

The measurement process that comes closest to 360 feedback in its challenges and the resulting validity issues is the performance appraisal.

The measurement process that comes closest to 360 feedback in its challenges and the resulting validity issues is the performance appraisal. The appraisal traditionally relies heavily on the skills of one individual, the supervisor, to generate reliable, valid information, which is often subsequently used for decision making (e.g., pay, promotions, development). Some guidelines for valid performance appraisal are offered by Bernardin & Beatty (1984), and include contextual variables such as rater training and ensuring adequate opportunity to observe that apply equally well to 360 feedback.

Three-sixty feedback is a very complex process that draws both the good and bad features of testing, performance appraisal, large-scale data collection (e.g., employee surveys), assessment centers, and employee development. Like performance appraisals and assessment centers, 360 feedback depends on the capability to generate reliable data from unreliable sources (i.e., people). Whereas rater training is required for assessment center raters and strongly recommended for supervisors doing appraisals (Bernardin & Beatty, 1984; Malos, 1998), the potential magnitude of the task of training raters for a 360 feedback process often prevents it from happening (Timmreck & Bracken, 1995). Not only are these raters often inexperienced, they also are typically not held accountable for the quality of their ratings, unlike supervisors in appraisal settings (London, Smither & Adsit, 1997).

Another characteristic of many 360 feedback processes that make them similar to appraisals but different from other measurement systems is their ongoing, repeated administration. An ongoing, repeated 360 process is a dynamic procedure where participants (raters, ratees, managers) have experiences that will shape their future interaction with the process. These experiences and their consequences can either enhance or decrease the effectiveness of the process.

There are some definitions of validity for 360 feedback that seem to acknowledge this systems perspective:

“... a measure of the usefulness of a procedure, technique, or device.” (Jones & Bearley, 1996)

“... validity is integrity.” (Van Velsor, Leslie & Fleenor, 1997)

These two definitions appear to recognize that the validity of a process cannot be known until the process is implemented, and that a value judgment must be made as to whether validity depends on its “usefulness” or its “integrity”.

We propose our own definition of validity for 360 feedback processes, which we feel is specific to this particular area of practice:

Sustained, observed improvement in behaviors valued by the organization.

In this definition, we stop short of requiring outcomes beyond behavior change. There is certainly demand for “bottom line” evidence for 360 feedback, and gradually some is forthcoming (Gross & Levy, 1998). While this type of research is needed, we see instances where behavior change is sufficient as an end product. An example would be where the behavioral model on which the 360 process is built is derived from the organization’s values statements. We would argue that, in those situations, an individual who changes his or her behavior to be consistent with these values as a result of receiving 360 feedback has contributed evidence of the validity of the process. Note that this definition can include behaviors which are not specifically assessed in the instrument but are developed in the total process.

Two Types of Validity Factors

When we think of factors that might cause a 360 feedback process to generate “bad” (invalid) outcomes (i.e., no behavior change, misguided behaviors), one category of such factors would, of course, include things that happen around the questionnaire and its administration. While such factors are undoubtedly important, we propose that successful 360 feedback implementations are not one-time events. Our definition of “success” requires *sustained* behavior change (London & Tornow, 1998; Wimer & Nowack, 1998). When viewed from this longer-term perspective, there is a whole set of other factors that

can invalidate or improve the process. These factors may not surface until the second or later administrations.

Table II lists factors that determine the validity of 360 feedback, sorted into two primary categories, proximal and distal, with a third category representing a mix of the two. Proximal factors occur at Time 1, the initial administration. These factors exert an immediate effect on the validity of the feedback for participants in the present administration. Distal factors reflect the fact that things happen (or don't happen) at Time 1 whose effects are not realized until Time 2, a subsequent administration of 360 feedback to the same population within a reasonable period of time (i.e., two years or less). As shown in Table II,

some factors are so powerful as to have effects at both Time 1 and Time 2 as proximal *and* distal factors (i.e., dual factors), but in different ways that will be described below.

Whether proximal, distal or both, each factor has implications for the design of the 360 process. (For further discussions of design features, an excellent treatment is provided by Dalessio, 1998.³) Table II illustrates the major design features potentially affected by each validity factor and some associated recommendations.

Proximal Factors

Proximal factors affect the quality of the feedback at the time of its current administration,

We propose that successful 360 feedback implementations are not one-time events.

TABLE II 360 Feedback Validity Factors and Associated Design Features.

Proximal Factors

Validity Factors	Design Features	Recommendations
Alignment	Instrument Design Report Format Feedback to Raters Integration with HR Systems	Custom design content Use internal norms Require meeting with raters Common content with appraisal
Accuracy	Processing Resources Quality Control Instrument Design	Ability to do high volume, secure reporting Processes to ensure zero errors Precode with important information
Clarity	Instrument Design Rater Training Pilot Administration	Clear instructions/readability Training sessions to give instructions Test understanding of participants
Cooperation	Instrument Design Rater Selection Rater Training Administration Process	Keep length reasonable (40-60 items) Limit demands on rater (number of forms) Communicate need for rater cooperation Do on company time
Timeliness	Administration Process Rater Training Integration with HR Systems Rater Training	Do as frequently as is reasonable/needed Train raters against recency error Schedule to coincide with system needs Deliver results as soon as possible
Reliability	Item Writing Instrument Design Rating Scale Rater Selection Rater Training Rater Selection	Clear, behavioral, actionable Conduct statistical analyses Use clearly defined anchors Select raters with opportunity to observe Train on proper use of scale Report rater groups separately
Insight	Instrument Design Report Formats Report Content Feedback to Raters	Use item ratings (not categories) Provide as much information as possible Report verbatim write-in comments Require meeting with raters

TABLE II Continued.*Distal Factors*

Validity Factors	Design Features	Recommendations
Ratee Accountability	Ratee Training Integration with HR Systems Feedback to Raters	Communicate expectations for ratees Set consequences for noncompliance Require meeting with raters
Commitment	Administration Process Participation of Management Developmental Resources Integration with HR Systems	Administer on company time Visible participation of top management Provide access to internal/external training Use results in decision making
Acceptance	Participation Rater Selection Administration Process Integration with HR Systems Instrument Design Ratee Training Developmental Resources	Require ratee participation Ratee selects raters, concurred by mgr. Administer consistently across unit Treat process as a business priority Content clearly tied to strategies, goals Train on how to use results Provide support (workshops, coaches, etc.)

Dual Factors

Validity Factors	Design Features	Recommendations
Consistency	Participation Rater Selection Administration Process Feedback to Raters Integration with HR Systems	Apply consistently across the organization; <i>When not possible</i> , test for possible unfairness
Anonymity	Administration Process Rater Selection Rater Training Report Features Feedback to Raters	Use outside vendor All subordinates; 4-6 in other groups Communicate how anonymity is provided Never report groups <3 (except supervisor) Don't try to identify raters
Census	Rater Selection Administration Process Data Manipulations	All subordinates; 4-6 in other groups Use methods to optimize response rates Do not use Olympic scoring
Communication	Rater Training Ratee Training Integration with HR Systems Pilot Test Policies/Procedures	Train all raters (not just instructions) Train all ratees Communicate how results are used Check for understanding Acknowledge need for sufficient time
Constraints	Integration with HR Systems Developmental Resources Ratee Training Manager Training	Tie development to appraisal Ensure fair access to development Train ratees how to access development Train managers on how to use feedback
Rater Accountability	Rater Training Feedback to Raters Rater Feedback Systems	Communicate role expectation as rater Require meeting with raters Online systems to give real time feedback

i.e., they have an immediate effect. In some ways, every proximal factor can have a distal (long term) effect due to issues such as loss of confidence in the process, which in turn affects the willingness of participants (both raters and ratees) to commit to future administrations. In fact, a proximal factor can be sufficiently botched to have the distal effect of terminating the process. Here are a few factors that are primarily proximal:

Alignment. This is probably the most direct and traditional definition of validity; that is, the extent to which the content of the feedback (e.g., competencies, skills, behaviors) is relevant to the success of the individual and, in turn, to the organization. If we are measuring the “wrong” things, then the process is by definition invalid.

Alignment is most powerful and defensible when tied to the strategies, values, and goals of the organization, then translated into competencies and/or desired behaviors for the entire employee population. 360 feedback instruments derived from some less obvious models of behavior, or standardized off-the-shelf instruments, require that validity be demonstrated for that specific organization and its employees (Dalessio, 1998; Van Velsor et al., 1997).

One of the advantages of well-designed 360 feedback processes is that alignment can be reinforced three times. The first opportunity is in the instrument design, as described above. A second opportunity for alignment occurs if the ratee meets with the raters (including the “boss”) to share results and development plans, in the process ensuring that the interpretation is correct and that the development plans are consistent with the expectations of the raters and the organization (Lepsinger & Lucia, 1997a; Walker & Smither, 1999). A third opportunity for alignment occurs if the feedback is integrated into a human resource system, such as performance appraisal, and the results are reviewed and action taken in the context of organization goals and values (Jako, 1997). Missing or mismanaging any one of these opportunities to ensure alignment certainly reduces, if not negates, the validity of the process.

A fairly common sentiment in 360 feedback circles is that organizations are trying to

create a feedback climate where questionnaires will become unnecessary (Dalton, 1998). If one truly believes that 360 feedback reinforces alignment by signaling the organization’s direction and associated behaviors, then 360 feedback processes will always have a place. If, for example, the organization is trying to sustain “organizational learning”, which involves rapid response to changes in the external environment, 360 feedback can be invaluable in reinforcing the new direction and ensuring that employees are giving feedback to each other on the “right” behaviors.

Accuracy. Here we are referring to the actual process of collecting and processing data and reporting the feedback—with no errors. This is an obvious but not so simple *sine qua non* of 360 feedback, exacerbated by the practice of providing feedback from relatively small groups of employees to a large number of ratees (Bracken, 1994). Lepsinger & Lucia (1997a) provide sample causes and solutions for accuracy problems. Suffice it to say that this is where “Murphy’s Law” is in full force.

Clarity. Even if raters are motivated and able to “perform” their role as feedback providers, it is incumbent on the administrators to ensure that raters are given clear information as to how to fulfill that role (i.e., correctly complete the survey and return it in a timely fashion to the right place). Other typical errors that raters unknowingly commit include miscoding (or misunderstanding) who they are rating, misusing the response scale, and providing inappropriate write-in comments. It must be ensured that participants simply understand what they are supposed to do—a task at once difficult to do and to know when it hasn’t been successful. The most obvious example is clear instructions, but good communications and training also can go a long way towards a successful implementation (Leslie & Fleenor, 1998).

Cooperation. The quality of feedback is very much dependent upon the motivation of the raters to participate, both in an absolute sense as reflected in response rates, and in a relative sense as reflected in the quality of their responses. The design features targeted here are those that determine the magnitude of the task, such as survey length and the number of surveys a given rater must complete. Various

One of the advantages of well-designed 360 feedback processes is that alignment can be reinforced three times.

symptoms of poor quality are unreturned surveys, incomplete surveys, and potential (perhaps unknown) effects of rater fatigue on the feedback. It is our experience that instrument designs using importance ratings and “gap analysis” (observed vs. expected performance) are of questionable value and serve largely only to add more burden on the raters with an associated impact on the quality and quantity of their responses.

Timeliness. A proximal factor that is often compromised in practice—timeliness in providing feedback—is important both for accuracy and acceptance by the ratee. For example, King & Downey (1998) report that delays in providing feedback can be a factor in increasing rating error, particularly when evaluating specific behaviors (vs. overall performance). A vast majority of 360 feedback processes are administered on an annual basis (Timmreck & Bracken, 1995), and raters can be expected to exhibit recency error. One method we have seen that attempts to avoid this particular type of error has been in team-based organizations where project teams form and disband on an ongoing basis. In some cases, 360 feedback is collected at the completion of each project, accumulated over the year, and then fed back to each ratee as cumulative results for all projects worked on during that period. While this addresses the timeliness issue identified by King & Downey (1998) as leading to rater errors, it does not solve the problem of delayed feedback to the ratee and the implications for acceptance of the feedback. Giving the results to the ratee immediately can work as long as rater anonymity (real and perceived) is not threatened due to small group sizes and the ratee is supported in the full and proper use of the feedback.

Reliability. The survey instrument serves as the primary—albeit artificial—communication vehicle between raters and ratees. In addition, some action should result from the collected feedback, whether it is development planning, training, or administrative decisions, such as compensation, staffing and succession planning, which can possibly add legal considerations. Some acceptable level of reliability for 360 ratings, therefore, must be demonstrated.

We are discovering that implementing 360 feedback causes us to question what reliabil-

ity means in this context, how it should be measured, and what is “acceptable” reliability (Bracken, 1996; Dalessio, 1998). For example, test-retest measures can be called into question due to real or perceived changes in ratee behavior in the intervening time (Tornow, 1993b) and/or changes in the raters themselves (e.g., attitudes, knowledge, opportunity to observe).

Another indicator of reliability is provided by measures of internal consistency, such as coefficient alpha for dimension scores (Van Velsor et al., 1997). However, recent findings also challenge our traditional thinking. Fecteau (1998) presents data suggesting that factor structures may differ between rater groups using the same instrument, which implies that different dimensions would be needed for each rater group (e.g., subordinates, peers, bosses, etc.). To some observers (e.g., Dalessio, 1998), this is not a surprising finding. More recent and encouraging finding of equivalence across rater sources (Maurer, Raju, & Collins, 1998; Frame, Ayman, Raju, & Goff, 1999; Woehr, Sheehan, & Bennett, 1999), even taking ratee gender into consideration (Frame et al., 1999), contradict the Fecteau (1998) conclusions and support the robustness of some instruments.

We challenge Fecteau (1998) and others examining measurement equivalence to first assure us that the instrument itself is not at fault. For example, we might expect vague items to be more susceptible to differences in meaning (and associated constructs) than very specific behavioral items. For example, we propose that items that begin with the following verbs (among others) are doomed to failure: “manages,” “oversees,” “values,” “knows,” “supports,” “ensures,” “supervises,” “understands,” and “believes.” The same would be true for double and triple barreled items which are proliferating in the attempt to shorten questionnaires. Unfortunately, Fecteau was unable to share the actual instrument that produced different factor structures.

Another possible cause of this phenomenon could be the way raters interpret and use the rating scale (Maurer et al., 1998). In the case of the Fecteau (1998) study, a five-point frequency scale (“always” to “never”) was used. Putting aside for the moment potential seri-

Giving the results to the ratee immediately can work as long as rater anonymity (real and perceived) is not threatened due to small group sizes and the ratee is supported in the full and proper use of the feedback.

ous deficiencies in frequency scales (Bracken & Paul, 1993), we wonder if frequency interacts with opportunity to observe (e.g., peers having less opportunity than subordinates) to create this artifact. Systematic research on rating scales is needed due to the lack of an accepted standard and the resulting proliferation of scale formats (Timmreck & Bracken, 1995).

Another common indicator of reliability is inter-rater agreement. In the case of 360 feedback, we can consider agreement between and within rater groups with different perspectives. Meta-analyses by Harris & Schaubroeck (1989) and Conway & Huffcutt (1997) both indicate moderate levels of inter-rater agreement within groups, although subordinates are often found to have the lowest reliabilities. Additionally, Fleenor, McCauley, & Brutus (1996) reported acceptable intraclass correlations for peers (.43 to .69) and subordinates (.47 to .70). Greguras & Robie (1998) go one step further and prescribe larger rater groups (9 subordinates, 8 peers, and 4 managers) than those normally found in practice to produce acceptable (.70) reliabilities with a five-point scale. We believe that such numbers should not be dismissed as unreasonable, and are consistent with our observations about census administration discussed below. At the same time, the Greguras & Robie (1998) results are reminiscent of the findings of Mount, Judge, Scullen, Sytsma & Hezlett (1997) that supervisor ratings are of higher reliability than peers or subordinates to the extent that the separate reporting of only supervisor scores is justified. In other words, they propose that only supervisor ratings should be broken out separately on the feedback reports—all other rating groups should be combined into an overall group.

Finally, agreement between rating sources has typically been shown to be low. This is not unexpected because it represents the very reason we do 360 feedback (i.e., different perspectives on performance) (Tornow, 1993b). On the other hand, Brutus, Fleenor, & London (1998a) calculated agreement between the supervisor, one randomly selected peer, a randomly selected subordinate, and the self rating using $r_{wg(j)}$ (James, Demaree & Wolf, 1984) and found values ranging from .95 to .97 across six industries! These results pro-

vide support for the findings of Mount et al. (1997) that there appears to be as much agreement across rater groups as within rater groups with 360 ratings.

The implication of such findings for the implementation of 360 feedback is unclear. While breaking out the results by rater group may or may not make statistical sense, these groups do represent an important reality in the organization, and ratees usually expect to see these data. As noted in the discussion of Alignment and Insight (below), it is often critical that participants fully understand the significance of the ratings obtained from these groups, facilitated both by report breakouts and discussions with raters during the action-planning phase.

Insight. In order for the ratee to take actions which are aligned with both the organization's desires and the feedback of the raters, he/she should be provided with the maximum amount of reliable information. The implications are primarily for the instrument design and report format but could include additional opportunities for insight created by meeting with raters, for example.

Some instruments solicit ratings at a category level, maybe providing a definition or behavioral examples to supposedly assist the rater. Some online systems allow the rater to link to a list of behaviors by clicking on the category label. Each of these approaches places an inordinate amount of faith in the raters, assuming that they do indeed read and use this information. More importantly, the ratee has no basis for determining the specific actions which have resulted in their scores. Systems that ask for (or require) written comments from raters cannot be expected to solve this problem and in effect would actually increase the burden for the rater.

Some report formats provide only mean scores with no indication of response distribution, presumably in an effort to protect rater anonymity and fears of retribution. While it is clearly important to maintain real and perceived anonymity of raters (as discussed below), we are doing a disservice to the ratee and hence the validity of the system if the ratee is forced to make decisions based on data presented in an incomplete fashion. Report features which help understanding include displays of actual rating distributions (frequen-

In order for the ratee to take actions which are aligned with both the organization's desires and the feedback of the raters, he/she should be provided with the maximum amount of reliable information.

cies), indicators of range (min/max, standard deviations), internal/external norms, and write-in comments. Again, each feature has its pros and cons and can place significant importance on the need for ratee training.

Distal Factors

Distal factors are products of Time 1 or intervening events whose effects are not realized until the second administration (Time 2). More often than not, the effects of distal factors are not anticipated and some surprises (pleasant or unpleasant) occur at Time 2. This indicates the usefulness of conducting evaluations immediately after Time 1 using focus groups or questionnaires to anticipate and maybe head off the threats to validity caused by problematic distal factors.

Ratee Accountability. Successful 360 feedback requires accountability for each of the three constituent groups: the raters (to provide honest feedback), the ratees (to use the feedback), and the organization (to support the process), whose “representative” is the “boss” (London et al., 1997). The distal factor we are addressing here is *ratee accountability*. While raters may have some preconceived notion of the probability that the ratee will use the results productively (and give or not give the benefit of the doubt accordingly the first time through), it is not until the ratee has the first opportunity to use the results that raters have the chance to see how the feedback is being used *by the ratee*. We have observed that rater participation and honesty will be differentially affected at Time 2 based on what they observed to happen at Time 1. To the extent that raters become less motivated to return surveys or give honest feedback, the reliability of feedback at Time 2 will be negatively affected. Conversely, raters who see that their feedback is being used productively should be expected to either start or continue to fulfill their accountability for giving accurate, honest feedback.

Ratee accountability is at the very heart of the debate over 360 feedback for development-only versus use for decision making (e.g., appraisal, pay, staffing, succession planning) (Bracken et al., 1997; Fleenor & Brutus, 2001). Table III shows many of the typical fea-

tures of development-only versus decision-making 360 processes. While there are many exceptions to these observations, they seem to be true more often than not (Bracken, 1996). One of these differences addresses the expectations for ratees in using the results. In development-only uses, ratees are usually given the final decision on what is done with the feedback. They usually receive the only copy of the feedback report and are not required (or even encouraged) to share results with raters or with their supervisor. Many of the development-only proponents would go so far as to say that this guarantee of confidentiality is a requirement for objective consideration and use of the feedback by the ratee (Dalton, 1996; 1997).

Unfortunately, experience suggests that, under these conditions, little behavior change is likely to occur for the majority of participants (Hazucha, Hezlett, & Schneider, 1993; Kluger & DeNisi, 1996). 360 feedback systems, when used for decision making, provide the results to the supervisor and often require the ratee to share results with the raters (e.g., subordinates) as well. This has a dual benefit. First, it increases the probability that the ratee will seriously consider the results and generate a development plan, which in turn can be used to secure the supervisor's support. Secondly, it “engages” the raters by showing them that their feedback is being used, and that their responsibility for helping the ratee doesn't end when they turn in a questionnaire. In these meetings, the ratee asks for clarification of the feedback and seeks support for the proposed development plan. This notion seems to be confirmed by Walker & Smither (1999) who report that, over a five year period of annual administration, managers who consistently held feedback sessions to discuss their upward feedback with direct reports improved more than other managers.

An extremely important issue, therefore, is the clear determination of the purpose of the process before any design begins (Bracken, 1996; Lepsinger & Lucia, 1997a; Wimer & Nowack, 1998). Practitioners should consider design features based on the purpose of their 360 feedback process, and realize that it is difficult to serve both purposes with one system. The possibility of parallel processes being run

Ratee accountability is at the very heart of the debate over 360 feedback for development-only versus use for decision making

TABLE III 360 Feedback Design Features by Purpose.

<i>Design feature</i>	<i>Purpose</i>	
	Decision Making	Development-Only
“Owner” of the feedback	Organization	Participant
Questionnaire content	Core competencies	Job-specific competencies
Length of questionnaire	Shorter (30-50 items)	Longer (80+ items)
Response scales	Encourage between-person comparisons	Encourage within-person rankings of skills
Ratee participation	Required	Voluntary
Administration schedule	As required (e.g., annual)	Ad hoc (on request)
Report format	Between-person (normative) comparisons	Within-person (skill ranking) comparisons
Copy of results	Participant and supervisor	Participant only
Share results with raters?	Expected or required	Optional or discouraged
Action plan	Required	Recommended
Typical uses	<ul style="list-style-type: none"> • Performance management • Succession planning • Staffing • Pay (e.g., bonuses) 	<ul style="list-style-type: none"> • Employee development • Career planning • Training
Adapted from Bracken (1996)		

for different purposes is an attractive alternative (Bracken, 1996; Tornow & London, 1998).

Commitment. Commitment is defined here as accountability demonstrated by the sponsors and administrators of the 360 process. Evidence of commitment comes in many forms, including public support of the process, a long-term commitment to achieving success, participation in the process, sharing of feedback results by senior managers, providing sufficient staff and resources for quality implementation, providing training for raters and ratees, and offering developmental resources for participants.

Acceptance. We believe that ratees must accept the feedback before acting on it in the spirit intended (i.e., consistent with individual and organizational goals.) Ratees (first timers in particular) predictably go through phases of shock, anger, and rejection of the results before being able to accept the feedback. This

certainly speaks to the need to acknowledge ratee reactions during the feedback process, and to have methods for helping individuals who cannot easily work through these stages, a far too common occurrence.

Dual Factors (Proximal and Distal)

Some factors operate both proximally and distally. These are complex factors that affect validity in a number of ways and over time.

Consistency. Acceptance and conformance in a 360 feedback setting, as in many other situations, are greatly influenced by perceptions of fairness. In many cases, the most accessible and verifiable evidence of fairness is found in consistency or lack thereof. Examples include: Is everyone required to participate? How are raters selected? Do I have to show my results to my supervisor? Will my results affect my per-

Acceptance and conformance in a 360 feedback setting, as in many other situations, are greatly influenced by perceptions of fairness.

formance evaluation? Do I have to share my results with my raters?

Inconsistent application engenders feelings of unfairness. For example, if some ratees are required to be assessed by all of their subordinates while others are free to ask some but not all of their subordinates to rate them, the system is likely to be perceived as unfair, and may indeed be unfair! Systems that are perceived to be unfair will not be supported by raters and ratees alike and will die a timely death.

Consistency operates proximally to the extent that participants (raters and ratees) understand, believe, and accept communications and policies about the process going into Time 1. It operates distally first as they see if those initial communications were accurate and then as they see how the feedback is used. For example, if they discover that everyone who was supposed to participate did not, or discover that most key managers seem not to be doing anything with their feedback, the validity of the process when administered at Time 2 will be diminished due to less reliable feedback from raters and less commitment and acceptance on the part of ratees.

Some inconsistencies that appear to be benign on the surface may also have the effect of creating unreliability (and threatening validity) when placed under scrutiny, thereby having a proximal effect. One example was reported by a company where part of the organization had the supervisors hand out their questionnaires while another part had the questionnaires mailed to the raters. Not surprisingly, but unfairly as it turns out, the supervisors who handed out their surveys got higher ratings! Other anecdotal observations suggest that raters give different scores when responding on a computer than when using paper and pencil versions (personal communication, M. Edwards, 1997). In both cases it is reasonable to expect claims of unfairness for ratees who are potentially adversely impacted by inconsistent practices and the resulting decisions, with the obvious legal implications.

Anonymity. Most 360 feedback systems provide for the anonymity of the raters. In fact, some would contend that, if raters are not anonymous, then it is not what has come to be known as "360 feedback" but something else, such as peer review, T-groups, or some-

thing comparable where participants give feedback face to face. Some would also argue that this is preferable to anonymous feedback, and that ideal feedback systems should not require anonymity nor questionnaires for the feedback providers (Dalton, 1998).

Most practitioners believe that anonymous feedback will be more honest than open feedback, especially when subordinates are asked to rate their supervisors. If anonymity is ensured through communications and instructions, more reliable (honest) feedback at Time 1 might be expected if a climate of trust exists—that is, if the raters believe that anonymity is ensured. But, if through the experiences of raters at Time 1, they come to believe that anonymity is not assured, then less honesty can be expected at Time 2 with a corresponding decrease in reliability and validity (Antonioni, 1994; London & Wohlers, 1991).

We purposely differentiate anonymity from confidentiality (Van Velsor, 1998), and encourage others to do the same. Confidentiality requires that the results (i.e., data, reports, comments) are shared only with those who have access according to predetermined and communicated policies. Some would contend that strict confidentiality where only the ratee sees the results is a requirement for behavior change (Dalton, 1997). We hold to the opposite belief that a ratee who is not required to share results and action plans with some significant party (i.e., supervisor, direct reports, peers) is unlikely to demonstrate behavior change (Hazucha, Hezlett & Schneider, 1993).

Census. The most reliable feedback will be collected from the population of coworkers (and customers) who have the best opportunity to observe the ratee, not a just a sample of observers. (As noted, the potential impact of requiring raters to complete an excessive number of evaluations needs to be considered and handled by policy and procedures; there are solutions to this problem.) The issues we describe below often leave us with relatively small numbers in rater groups, especially when reported separately by perspective (e.g., peers, subordinates). Serious problems can occur (e.g., low reliability) when feedback is based on the resulting sample rather than on a census (Greguras & Robie, 1998), with corresponding implications for acceptance of the feedback by the ratee.

Systems that are perceived to be unfair will not be supported by raters and ratees alike and will die a timely death.

Some of the design features that lead to a lack of rater census include:

1. *Policy.* The organization sets an upper bound limit as to how many raters can be selected by the ratee. This practice clearly treats the ratee as the least important “customer” of the process, giving greater consideration to the raters (by limiting the number of surveys to complete) and the organization (by limiting the cost, both soft and hard).
2. *Response rates.* Even if policies are implemented that encourage or require census rater selection (e.g., all direct reports), anything less than a 100% response rate reduces the respondent pool, resulting in greater measurement error and potential loss of confidence in the results on the part of the ratee. The implications for system design here are focused on processes that prevent or demotivate participation (e.g., survey length, improper instructions, perceived lack of anonymity, perceived risks, lack of opportunity, etc.).
3. *Data reduction.* Even when policies and practices encourage census selection and 100% response rates, some reporting procedures turn a potential census into a sample through the arbitrary elimination of responses. Often called “Olympic scoring” (Edwards & Ewen, 1996; Lepsinger & Lucia, 1997a; 1997b), these processes result in the removal of high and low scores under the premise that these responses are somehow less accurate. As discussed by Bracken (1996), this premise seems fatally flawed. Here we add the argument that reducing the number of reported responses further increases measurement error and reduces ratee acceptance.

The proximal impact of not using a census approach is at least threefold. Most obviously, it affects the reliability of the feedback due to reporting from small numbers with potentially astronomical measurement error

(Greguras & Robie, 1998). Most likely, ratees also will perceive the feedback as less reliable as well. Secondly, there may be an unknown amount of influence on a rater’s feedback when he/she feels “exposed” as part of a small cadre of selected raters, who may wonder why they were selected and if they need to provide ratings that reflect the opinions of other employees who were not selected. And, as noted already, the third issue is the impact of Olympic scoring on the reliability of reported results. An additional organizational issue that is raised by not using a census of subordinates is the resulting feeling of exclusion by those not asked for feedback. The excluded subordinates will feel less engaged (i.e., valued) in the total process, perhaps negatively affecting both their willingness to assist the ratee in developmental activities (with implications for Alignment) and the quality of their ratings if they are selected for a subsequent process. We would contend, for example, that raters become better observers after they have experienced the feedback process one or more times, through their increased awareness of the desired behaviors.

We always worry about the acceptance of the feedback (and commitment to behavior change) when the feedback is perceived as being unreliable by the ratee. In addition to ratee acceptance, a sampling approach may inhibit the ratee from meeting with raters (especially subordinates) to discuss results, a very important step in maintaining the viability of an ongoing feedback process (Walker & Smither, 1999).

Communication. Communication is a vehicle for stating the “contract” with the process participants—raters and ratees alike. As with all contracts, there are expectations for both sides of the agreement, including expectations for how the raters and ratees will perform in support of the process. In the descriptions of Consistency and Anonymity above, it should be clear that the communications at Time 1 are key to establishing trust and understanding in the minds of the raters, who must be able to provide honest, accurate feedback. They in turn expect the organization to fulfill its side of the contract in the way it protects the raters and supports the development of the ratees.

An additional organizational issue that is raised by not using a census of subordinates is the resulting feeling of exclusion by those not asked for feedback.

Williams & Lueke (1998) confirmed that Perceived System Knowledge (PSK) (Williams & Levy, 1992) is positively correlated with participant reactions to the 360 feedback system, self-efficacy to change, and higher intentions to improve, as well as being negatively correlated with situational constraints (see below). PSK showed no relation to amount of experience with the 360 feedback process, indicating the importance of good communication and understanding from the start.

Constraints. Because our definition of validity requires that observed behaviors change, we must acknowledge that ability and motivation are not enough; the individual must also have the “opportunity” to perform (i.e., exhibit new behaviors). Lack of opportunity to perform, or constraints, can come in the forms of lack of time to practice new skills, lack of resources, or lack of information (Peters & O’Connor, 1985; Wimer & Nowack, 1998). Constraints can also come in the form of lack of support, or even fear of negative consequences, from a boss who doesn’t value the new behaviors. Constraints might also be perceptual, such as the belief that the new behaviors will not be accepted in the existing culture. With a 360 feedback process, Williams & Lueke (1998) confirm that the degree of perceived constraints is negatively correlated with satisfaction and intention to change.

Rater Accountability. It has been long established that people *can* be reliable, accurate observers and reporters of the behavior of others. The challenge continues to be how to enable and motivate them to do so, particularly in the open, uncontrolled environment of the workplace. Murphy & Cleveland (1995) go as far as to maintain that differences between actual and rated performance should not be considered to be measurement error but “forces in the organization that discourage accurate rating” (p. 29). Unfortunately, that turns 360 feedback into something more like a climate survey.

On the ability side of the equation, certain types of rater training have been shown to improve accuracy (Hauenstein, Facticeau, & Schmidt, 1999; Woehr & Huffcutt, 1994). Typically, in the 360 feedback arena, the concern is higher on the motivational side. How

do we motivate raters so that their ratings reflect their observations? The “carrot” approach to motivation is to show the rater the value of being honest through communication and follow through by the ratee. A “stick” approach is to discard clearly invalid rating patterns (e.g., ratings all of the same value). Online administrations may give us the opportunity to give raters immediate feedback on their rating distributions.

It is unlikely that organizations would (or should) sacrifice anonymity in order to establish direct accountability (i.e., requiring raters to justify their ratings). London et al. (1997) propose that having raters participate in the review of the ratee’s results in the spirit of ongoing development creates a moderate level of accountability, an indication to the raters that their role is not fulfilled by simply completing a questionnaire. This view of 360 feedback as an ongoing process, not an event, is central to the views we are also presenting.

We are concerned about blanket statements that ratings produced under conditions of accountability and/or use for decision making are less accurate than those collected under “safe” conditions (e.g., for developmental purposes). While there is evidence that scores are indeed higher in the former (accountable) situations (Antonioni, 1994), there is no basis to conclude that either situation produces *more accurate* ratings. In their review of the literature on this topic, Haeggberg & Chen (1999) are unwilling to make any directional hypotheses about the effect of accountability since the research to date has been totally inconclusive and contradictory. Their research suggests the need to differentiate between observational accuracy and rating accuracy, which is analogous to the ability vs. motivational components of rater performance noted above.

Conclusion

We have presented a comprehensive model of factors that we believe determine the “validity” (as defined by us) of a 360 feedback process. We contend that both proximal and distal factors can either bolster or adversely affect the validity of a 360 feedback process. Each validity factor has implications for the

With a 360 feedback process, Williams & Lueke (1998) confirm that the degree of perceived constraints is negatively correlated with satisfaction and intention to change.

design features of the system, shown in Table II. We acknowledge that it is probably impossible to optimize all the validity factors simultaneously. On the other hand, we suggest that all validity factors need to be considered simultaneously, and that the implications of design feature decisions be carefully weighed in this context. A 360 system that is poorly designed could (and should) not be sustained for any period of time. We feel that a process must be sustained to be successful (again by our definition). We strongly suggest that practitioners carefully consider the factors in Table II when designing and implementing 360 feedback systems, and furthermore publish their findings as they become available. An attempt to further translate these factors into design decisions has been published as “Guidelines for multisource feedback when used for decision making” (Bracken & Timmreck, 2001).

A central theme in our systems approach is that it is impossible to simultaneously optimize all elements. As practitioners make decisions and the resulting tradeoffs, a key consideration is the question of “Who is the customer?” Our sense is that too many organizations are designing their 360 processes with an implicit acknowledgment that the rater is the primary customer as evidenced by very short instruments and/or very small rater group sizes, for example. Another take on such practices is that cost containment is paramount, placing the organization ahead of the ratee as a customer. While each constituent group must be considered, we would propose that our definition of success requires that the ratees’ needs should come before the raters and organization.

We realize that our model of the effectiveness of 360 feedback is deficient in some important ways, not the least of which is that it does not recognize the individual differences in the ratees, either demographically or experientially (Brutus, Fleenor & McCauley, 1999). For example, several studies (Atwater, Roush & Fischthal, 1995; Smither, London, Vasilopoulos, Reilly, Millsap, & Salvemini, 1995; Walker & Smither, 1999) have reported that only participants who were initially rated poorly showed significant improvement in their performance over time. Our model, like many in our field, is designed to apply as widely as possible and to increase the overall probability of behavior change. To the extent that we can “change a lot of people just a little”, we believe that overall organizational effectiveness will be improved.

Multi-source feedback looks deceptively easy. It is like watching a proficient athlete and saying “I could do that” without any knowledge of the training and preparation required for excellence. In much the same way, an effective 360 feedback system that creates sustainable, focused behavior change requires both good tools and good implementation. Systems that fall short in one or more of the factors cited in this article will probably not “hurt” anyone, and may actually survive in a benign way. The problem is that they will not create behavior change, at least in the manner envisioned by the organization. At best they are a waste of money. At worst they divert time and resources towards activities that will not bear fruit and may cause employee morale problems.

Multi-source feedback looks deceptively easy. It is like watching a proficient athlete and saying “I could do that” without any knowledge of the training and preparation required for excellence.

DAVID W. BRACKEN is Director of Research Consulting for Mercer Delta Consulting, LLP in New York City. He is best known for his work in the area of multisource (360 degree) feedback with emphasis on its use for decision making. He is senior editor and contributor to *The Handbook of Multisource Feedback* (with co-editors Carol W. Timmreck and Allan H. Church). He holds an A.B. degree from Dartmouth College and M.S. and Ph.D. degrees in industrial and organizational psychology from Georgia Tech.

CAROL W. TIMMRECK is an organization development consultant in Shell Oil Company, where she has been responsible for the development and annual implementation of multisource feedback processes for nearly all of Shell's USA managers. She co-founded The Multisource Feedback Forum, a consortium of large companies with active MSF processes. She has addressed numerous national audiences on the topic of multisource feedback, and is a co-editor and contributor to *The Handbook of Multisource Feedback*. She received her Ph.D. degree in industrial and organizational psychology from the University of Houston.

JOHN W. FLEENOR is Director of Knowledge Management at the Center for Creative Leadership. As leader of feedback research at the Center, he was responsible for managing research on the development and use of 360-degree feedback instruments. He is a co-author of three books and numerous articles and book chapters on multisource feedback. He is a member of the editorial board of *Human Resource Management Journal* and is listed in *Who's Who in the Management Sciences*. His Ph.D. is in industrial and organizational psychology from North Carolina State University, where he is an adjunct member of the psychology faculty.

LYNN SUMMERS is co-founder and vice president of research at Performaworks, Inc., located in Raleigh, NC. Performaworks is an outsource provider of Web-based performance management services (www.performaworks.com). He has served as staff psychologist at Baltimore Gas & Electric, VP of training and management development at Hardee's Food Systems, head of the HR program at Peace College, and book review editor of *Personnel Psychology*. His doctorate in industrial and organizational psychology is from the University of South Florida.

REFERENCES

- Antonioni, D. (1994). The effects of feedback accountability on upward appraisal ratings. *Personnel Psychology*, 47, 349–356.
- Atwater, L., Roush, P., & Fischthal, A. (1995). The influence of upward feedback on self and follower ratings of leadership. *Personnel Psychology*, 48, 35–59.
- Bernardin, H.J. & Beatty, R.W. (1984). *Performance appraisal: assessing human behavior at work*. Boston: Kent.
- Bracken, D.W. (1994). Straight talk about multirater feedback. *Training and 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: Tools for assessment and change* (pp. 117–143). San Francisco: Jossey-Bass.
- Bracken, D.W., Dalton, M. A., Jako, R.A., McCauley, C.D., & Pollman, V.A. (1997). *Should 360-degree feedback be used only for developmental purposes?* Greensboro, NC: Center for Creative Leadership.
- Bracken, D.W., & Paul, K.B. (1993, May). The effects of scale type and demographics on upward feedback. Paper presented at the 8th Annual Conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.
- Bracken, D.W., & Timmreck, C.W. (2001). Guidelines for multisource feedback when used for decision making. In D. Bracken, C. Timmreck, & A. Church (Eds.), *The handbook of multisource feedback* (pp. 495–510). San Francisco: Jossey-Bass.
- Bracken, D., Timmreck, C., & Church, A. (Eds.). (2001). *The handbook of multisource feedback*. San Francisco: Jossey-Bass.
- Brutus, S., Fleenor, J.W., & London, M. (1998a). Does 360-degree feedback work in different industries? A between-industry comparison of the reliability and validity of multi-source performance ratings. *Journal of Management Development*, 17, 177–190.
- Brutus, S., Fleenor, J.W., & London, M. (1998b). Elements of effective 360-degree feedback. In W. Tornow & M. London (Eds.), *Maximizing the*

- value of 360-degree feedback: A process for successful individual and organizational development (pp. 11–27). San Francisco: Jossey-Bass.
- Brutus, S., Fleenor, J., & McCauley, C. (1999). Demographic and personality predictors of congruence in multi-source ratings. *Journal of Management Development*, 18, 417–435.
- Church, A.H., & Bracken, D.W. (Eds.). (1997). Special issue: 360-degree feedback systems. *Group & Organization Management*, 22(2).
- Conway, J., & Huffcutt, A. (1997). Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, supervisor, peer and self-ratings. *Human Performance*, 10, 331–360.
- Dalessio, A.T. (1998). Using multi-source feedback for employee development and personnel decisions: Practice and research. In J.W. Smither (Ed.), *Performance appraisal: State-of-the-art methods for performance management* (pp. 278–330). San Francisco: Jossey-Bass.
- Dalton, M.A. (1996). Multirater feedback and conditions for change. *Consulting Psychology Journal*, 48, 12–16.
- Dalton, M.A. (1997). When the purpose of using multi-rater feedback is behavior change. In D. Bracken et al. (Ed.), *Should 360-degree feedback be used only for developmental purposes?* Greensboro, NC: Center for Creative Leadership.
- Dalton, M.A. (1998). Best practices: Five rationales for using 360-degree feedback in organizations. In W. Tornow & M. London (Eds.), *Maximizing the value of 360-degree feedback: A process for successful individual and organizational development* (pp. 59–77). San Francisco: Jossey-Bass.
- Edwards, M., & Ewen, J. (1996). 360° feedback: The powerful new model for employee assessment & performance improvement. New York: AMACOM.
- Facteau, J.D. (1998, April). Are performance appraisal ratings obtained from different rating sources comparable? Paper presented at the 13th Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.
- Fleenor, J., & Brutus, S. (2001). Multisource feedback for personnel decisions. In D. Bracken, C. Timmreck, & A. Church (Eds.), *The handbook of multisource feedback* (pp. 335–351). San Francisco: Jossey-Bass.
- Fleenor, J., McCauley, C., & Brutus, S. (1996). Self-other rating agreement and leader effectiveness. *Leadership Quarterly*, 7, 487–506.
- Frame, M.C., Ayman, R., Raju, N.S., & Goff III, M. (1999, April). Executive level multi-rater performance ratings: Measurement equivalence across source and gender. Paper presented at the 14th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.
- 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.
- Gross, F.A., & Levy, P.E. (1998, April). Do 360-degree feedback appraisals predict managerial effectiveness? Paper presented at the 13th Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.
- Haeggberg, D., & Chen, P.Y. (1999, May). Can accountability assure the accuracy of upward appraisals? Paper presented at the 14th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.
- Harris, M., & Schaubroeck, J. (1989). A meta-analysis of self-boss, self-peer, and peer-boss ratings. *Personnel Psychology*, 41, 43–62.
- Hauenstein, N.M.A., Facteau, J.D., & Schmidt, J.H. (1999, May). Rater variability training: An alternative to rater error training and frame-of-reference training. Paper presented at the 14th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.
- Hazucha, J.F., Hezlett, S.A., & Schneider, R.J. (1993). The impact of 360-degree feedback on management skills development. *Human Resource Management*, 32, 325–351.
- Jako, R.A. (1997). Fitting multi-rater feedback into organizational strategy. In D. Bracken et al. (Ed.), *Should 360-degree feedback be used only for developmental purposes?* Greensboro, NC: Center for Creative Leadership.
- James, L.R., Demaree, R.G., & Wolf, G. (1984). r_{wg} : An assessment of within-group interrater agreement. *Journal of Applied Psychology*, 78, 306–309.
- Jones, J.E., & Bearley, W.L. (1996). 360° feedback: Strategies, tactics, and techniques for developing leaders. Amherst, MA: HRD Press.
- King, C.B., & Downey, R.G. (1998, April). Cognitive processes in performance appraisal: Prior impressions and delayed evaluations. Paper presented at the 13th Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.
- Kluger, A., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254–284.
- Lepsinger, R., & Lucia, A.D. (1997a). The art and science of 360° feedback. San Francisco: Pfeiffer.
- Lepsinger, R., & Lucia, A.D. (1997b, September). 360 feedback and performance appraisal. *Training and Development*, 62–70.
- Leslie, J.B., & Fleenor, J.W. (1998). Feedback to managers: A review and comparison of multi-rater feedback instruments for management development (3rd ed.). Greensboro, NC: Center for Creative Leadership.
- London, M., Smither, J.W., & Adsit, D. J. (1997). Accountability: The Achilles' heel of multisource feedback. *Group & Organization Management*, 22, 162–184.
- London, M., & Tornow, W.W. (1998). 360-degree feedback—More than a tool! In W. Tornow &

- M. London (Eds.), Maximizing the value of 360-degree feedback: A process for successful individual and organizational development (pp. 1–8). San Francisco: Jossey-Bass.
- London M., & Wohlers, A.J. (1991). Agreement between subordinate and self-ratings in upward feedback. *Personnel Psychology*, 44, 375–390.
- Malos, S.B. (1998). Current legal issues in performance appraisal. In J. W. Smither (Ed.), *Performance appraisal: State of the art in practice* (pp. 49–94). San Francisco: Jossey-Bass.
- Maurer, T.J., Raju, N.S., & Collins, W.C. (1998). Peer and subordinate performance appraisal measurement equivalence. *Journal of Applied Psychology*, 83, 693–702.
- Mount, M., Judge, T., Scullen, S., Sytsma, M., & Hezlett, S. (1997, April). Trait, rater and level effects in 360-degree performance ratings. Paper presented at the 12th Annual Conference of the Society for Industrial and Organizational Psychology, St. Louis, MO.
- Murphy, K., & Cleveland, J. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. Thousand Oaks, CA: Sage.
- Nunnally, J. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Peters, L.J., & O'Connor, E.J. (1985). Situational constraints and work outcomes: The influences of a frequently overlooked construct. *Academy of Management Review*, 5, 391–398.
- SIOF. (1987). *Principles for the validation and use of personnel selection procedures*. College Park, MD: Society for Industrial and Organizational Psychology.
- Smither, J.W., London, M., Vasilopoulos, N.L., Reilly, R.R., Millsap, R.E., & Salvemini, N. (1995). An examination of the effects of an upward feedback program over time. *Personnel Psychology*, 48, 1–34.
- Timmreck, C.W., & Bracken, D.W. (1995, May). Upward feedback in the trenches: Challenges and realities. In W.W. Tornow (Chair), *Upward feedback: The ups and downs of it*. Symposium conducted at the 10th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.
- Timmreck, C.W., & Bracken, D.W. (1997). Multisource feedback: A study of its use in decision making. *Employment Relations Today*, 24, 21–27.
- Tornow, W.W. (Ed.). (1993a). Special issue on 360-degree feedback. *Human Resource Management*, 32 (2&3).
- Tornow, W.W. (1993b). Perceptions of reality: Is multi-perspective measurement a means or an end? *Human Resource Management*, 32, 221–229.
- Tornow, W.W., & London, M. (1998). Challenges and implications for maximizing 360-degree feedback. In W. Tornow & M. London (Eds.), *Maximizing the value of 360-degree feedback: A process for successful individual and organizational development* (pp. 249–258). San Francisco: Jossey-Bass.
- Van Velsor, E. (1998). Designing 360-degree feedback to enhance involvement, self-determination and commitment. In W. Tornow & M. London (Eds.), *Maximizing the value of 360-degree feedback: A process for successful individual and organizational development* (pp. 149–195). San Francisco: Jossey-Bass.
- Van Velsor, E., Leslie, J.B., & Fleenor, J.W. (1997). *Choosing 360: A guide to evaluating multi-rater feedback instruments for management development*. Greensboro, NC: Center for Creative Leadership.
- Waldman, D., & Atwater, L. (1998). *The power of 360-degree feedback: How to leverage performance evaluations for top productivity*. Houston: Gulf.
- Walker, A.G., & Smither, J.W. (1999). A five-year study of upward feedback: What managers do with their results matters. *Personnel Psychology*, 52, 393–425.
- Ward, P. (1997). *360-degree feedback*. London: Institute of Personnel Development.
- Williams, J.R., & Levy, P.E. (1992). The effects of perceived system knowledge on the agreement between self ratings and supervisor ratings. *Personnel Psychology*, 45, 835–847.
- Williams, J.R., & Lueke, S.B. (1998, April). 360-degree feedback system effectiveness: Examination of participant reactions. Paper presented at the 13th Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.
- Wimer, S., & Nowack, K.M. (1998). 13 common mistakes using 360-degree feedback. *Training and Development*, 52(5), 69–80.
- Woehr, D.J., & Huffcutt, A.I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, 67, 189–205.
- Woehr, D.J., Sheehan, M.K., & Bennett, W. (1999, April). Understanding disagreement across rating sources: An assessment of the measurement equivalence of raters in 360 feedback systems. Paper presented at the 14th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.

ENDNOTES

1. Bracken, Timmreck & Church, 2001; Edwards & Ewen, 1996; Jones & Bearley, 1996; Lepsinger & Lucia, 1997a; Leslie & Fleenor, 1998; Tornow & London, 1998; Waldman & Atwater, 1998; Ward, 1997.
2. Bracken, 1996; Dalessio, 1998.
3. In addition, we refer you to Bracken, 1994; Bracken & Timmreck, 1999; Lepsinger & Lucia, 1997b; Van Velsor, 1998; Waldman & Atwater, 1998; and Ward, 1997.