MANAGING THE 360 PROCESS: A PRIMER

W. Christopher Musselwhite, MSIE, EdD

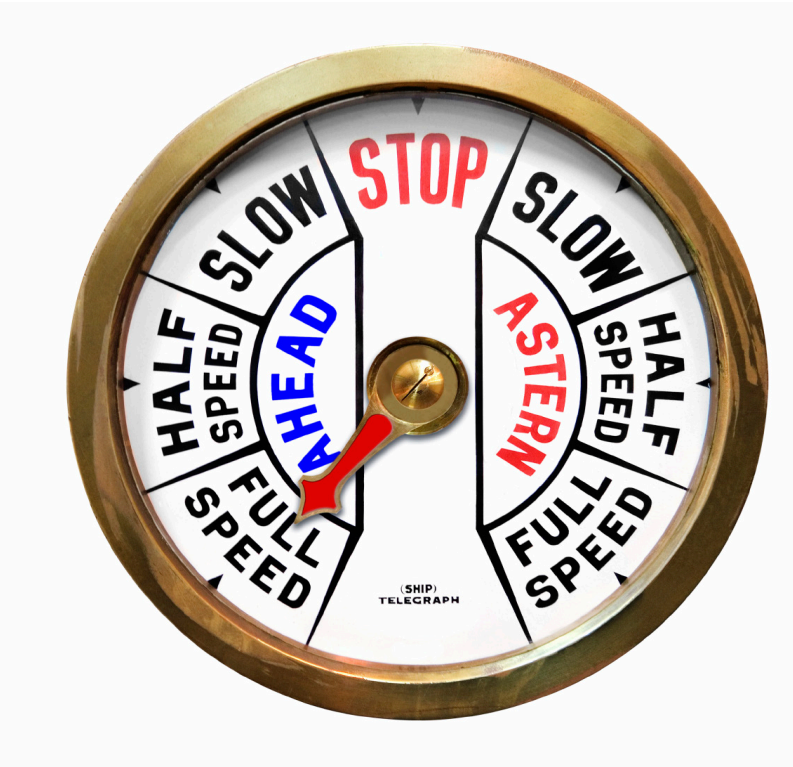
Discovery Learning Inc.

All About 360s: Done well, 360s (multi-rater leadership assessments) can provide excellent data for development. Done badly, they can cause confusion and distrust. The result is either an organization with enhanced capacity for effective operation or one in which ineffective behavior is reinforced or increased. As a result, leadership and management competency assessment using 360's is a challenging process in any organization. But taking a considered approach and following a few simple guidelines can help any organization use this proven development tool effectively.

Research shows it's cheaper to develop and retain talent than it is to recruit and hire. Companies seeking to keep existing talent via the use of more development are using 360's, along with other development tools, more frequently. Because research consistently ranks growth and development as top contributors to job satisfaction and motivation, using development tactics to retain top talent is a wise strategy.

By revealing personal blind spots that might negatively affect an individual's managerial or leadership effectiveness, 360s can provide a very focused and efficient road map for personal development. When individuals can quickly change behaviors that negatively affect those who work closest to them, they can adjust to quickly change perceptions of their effectiveness.

In addition to helping individuals recognize development areas, 360s can do the same for organizations. Composite data can reveal trends in areas ripe for development across the organization. This is valuable to organizations that seek to learn the strengths and weaknesses in their managerial culture. Plus, it can provide insight into why particular companywide initiatives stall. Smart organizations realize cultural change starts with individual change, and then they complement it with an organization development strategy. With 360s you can do both.



For 360s to Work

If you're thinking about using a 360 in your organization, take your time. To get the most value from your investment in a 360 and to ensure you select the one that's best for your organization, there are several factors you should consider.

To make a 360 payoff, you must think of the 360 survey as a component of a larger development system. Having a larger process in place helps people understand their development needs, provided a road map for the development process and allows for the crucial follow-up and/or closure to measure their progress.

A successful assessment process must include clear purpose and intention, selection of a quality assessment tool, the process for survey administration, an easy to understand feedback report, materials for guiding post-survey participant development, and support materials for consultants managing the process. Discovery Learning, Inc. offers the following guidelines to assist in the selection and use of multi-rater leadership assessment tools for a successful leadership assessment process.

Getting started

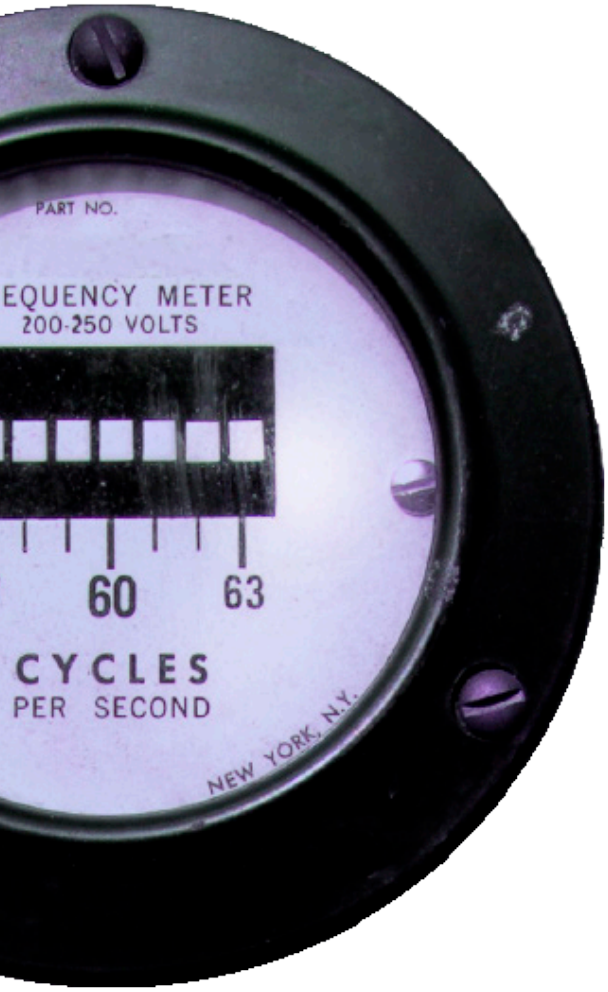
1 **Guideline: Purpose.**
The purpose of a 360 assessment process should be clear to everyone involved. 360 assessments are utilized for development, placement or promotion. The most common use is for development. Whatever the purpose, clear communication is critical. Never mix purposes.

2 **Guideline: Confidentiality.**
Trust is dependent on the confidentiality of data collected in a 360 process. If an environment of trust does not exist, data collected may be inaccurate. People only give honest feedback if they feel safe and trust the process. Unless

otherwise stated, the participant should be the only recipient of results. If a target participant's boss is to view the data, this must be clearly communicated from the onset of the process.

3 **Guideline: Issues of anonymity.**
Typically, a boss' feedback is presented to a participant separately from other data. For other rater groups, such as peers and direct reports, feedback is usually presented as consolidated data. Generally, at least three raters' data must be collected in order to receive feedback in any rater category. If fewer than three respond in any category, that data is presented only in the composite category of all raters. If open-ended questions are part of the assessment process, it should be stated that the responses will be reported verbatim.

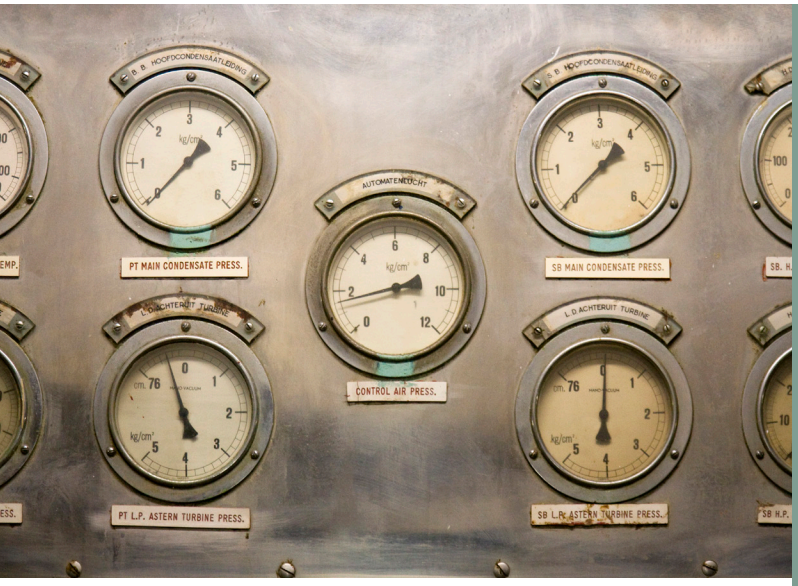
4 **Guideline: Participant support.**
An often overlooked aspect of a 360 assessment is upfront participant support. Participants should be clear about the purpose of the assessment process and obtain the boss' support. The boss should meet with the participant prior to the 360 assessment, state the purpose of the assessment, clarify how data will be utilized, and clarify his/her role in the process. If a coach or support person is to be used, this role should be clarified and assigned. The boss' interaction with the participant following the assessment process should also be defined. For example, if the boss expects to meet with the participant and either review or receive a report on the feedback data, this should be stated upfront.



Selecting the Survey

5 Guideline: Survey instructions.
There are two categories of survey instructions. The first includes instructions to the participant, consisting of information for selecting, contacting, and communicating with raters and on returning data to the scoring center. The second category includes specific instructions for answering the survey items. These instructions are important to both the participant and the selected raters since both must complete the survey. Check these instructions for clarity. Administer and review a sample survey before sending to your participants.

6 Guideline: Design of survey items.
Survey items may be designed to measure behaviors, skills or perspectives. In general over the past few years there has been a shift from measuring traits to measuring behaviors. Survey items should be well constructed. Several factors should be considered in assessing item quality and reliability. (Reliability refers to the consistency and dependability of items.) These factors include:



- *Unidimensionality – Each item should measure only one behavior or skill. Watch out for "ands" which could combine two different behaviors. A person may exhibit one without the other and this leaves the rater uncertain of how to respond.*
- *Item clarity – Wording should be clear and easy to understand. Language should not exceed a 10th grade level.*
- *Observability – The skills or behaviors being measured should be observable. Avoid items that require assumptions about attitudes and intentions.*
- *Qualifiers – Good items should avoid qualifiers, such as "very" and "extremely".*

Qualifiers may conflict with the response scale and make responses to items more difficult. The response scale is designed for measuring the "extent" or "degree" of a behavior or skill.

- *Possibility of development – The behavior or skill measured should be amenable to change. Through goal setting and action planning, participants should be able to improve the behavior or skill with effort over time. Certain personality traits may not meet this criteria. The ability to change a measured behavior or skill is a form of construct validity that is not often measured. This is because of the difficulty with measuring behavior change.*

- *Test-retest reliability – Over a short period of time, (six to 12 weeks) responses to items on the survey should produce similar results, assuming that no intervention occurs between the test-retest period. If test-retest responses are not similar it may mean that survey items are ambiguous and can have multiple interpretations. Of course, the goal in real-world applications is to see measureable* improvement over time.

7

Guideline: Rater response scales.
These scales are the rating choices offered to raters. Response scales fall into two categories; frequency (how often) and mastery (how well). Frequency scales tend to be the norm and are easier for raters. One problem with frequency scales is that performing a behavior more frequently does not necessarily correlate with performing it more effectively. When frequency scales are utilized it is desirable to include the opportunity for raters to also indicate if they would like to see more or less of the behavior.

Examine the number of points (numbers) on a scale. Five point scales are most frequently used, and scales that include more than seven points may provide too many options, requiring too much time for completion. Less than five points will not provide enough options for raters to choose from. Remember that most managers being assessed tend to be reasonably effective and competent, thus falling on the positive end on the scale. As a result when the data is normed for all managers (converted to percentile scores), ratings that are slightly or moderately positive can fall into the lower percentile scores simply because so many people tend to score high on the survey. The consequence is that raw scores that are in the positive range (for example between 3 and 4 on a 5 point scale) will fall into the lower range of the percentile scores (for example possibly 25th to 40th percentile). The consequence is that the feedback may appear lower than the rater intended. The fact is that raters use the positive numbers on the rating scale much more than the negative numbers. Look for scales that are well distributed. The best scales should include more positive points than negative, using one or two negative points to three to four positive points.

Look at mean scores and standard distributions. The mean score should be near the middle of the scale or slightly to the positive side. The distribution of scores along the scale is usually measured by standard deviation (the range on either side of the mean (average score) that contains 67% of the responses). If the standard deviation is too small, then scores will cluster near the mean score and a very small change in raw score could result in a much larger change in the equivalent percentile score. This can confuse and mislead participants.

It is critical that assessment scales include a response option for skills or behaviors that cannot be, or have not been, observed. An example of such a rating would be "not

observed". Otherwise, respondents often select the most neutral rating, thus distorting feedback.

8

Guideline: Behavior/Skill Categories.
Feedback categories should be well constructed. Feedback categories are groupings of items that define a larger construct. For example, managing conflict may be a construct that includes individual items such as: Seeks win-win outcomes, Responds to criticism without becoming defensive, Does not avoid conflict, etc. The quality of these categories can be determined by internal consistency and construct validity.

- *Internal consistency measures whether items making up a single scale or construct are measuring the same thing and is based on average correlations among items and the number of items in a scale. The measure of fit among the items is referred to as an alpha score. An alpha score of 0.7 (range can be from 0 to 1.0) is considered acceptable. If the alpha score is too low, items are measuring different constructs; if the alpha score is too high, there is excessive overlap among the items and some of the items are likely redundant.*

- *Construct validity tests the assumption that the instrument measures what it purports to measure. Construct validity can be measured through item-scale correlations, hypothesis testing, and comparisons to other instruments.*

9

Guideline: Survey Length.
Leadership development is a broad topic and most leadership surveys strive to be comprehensive. But, surveys that are too long create survey fatigue in both the participant being surveys and in the larger organization. This is especially true if multiple people are being surveyed simultaneously.

For example, you are planning a leadership development program and decide to include the top twenty five leaders in the organization. Since these people are each other's bosses, peers or direct reports, everyone is having to rate multiple people. If each person rated requires thirty minutes you could be looking at four to five hours the executives are being asked to commit to the process. Raters may start to move through the surveys quickly giving less and less thought to their responses. 360s that have fewer items often deliver better data, even though they may not be as comprehensive. Forty to seventy items should be capable of covering the necessary leadership behaviors.

Administering the Survey

10

Guideline: Selection of raters.
Raters (boss, peers, direct reports, customers, etc,) should be selected by the participant. Typically, these are people who work closely with the participant and are in a position to observe his/her behavior. When possible, a minimum of four raters in the peer and direct report category should be identified. Thus if one rater does not complete the survey, the participant will still receive three ratings needed to produce the separate feedback report for the rater group.

11

Guideline: Management and administration.
On-line administration of 360 assessments has greatly simplified the data collection process. Most surveys require the vendor to manage the process. Some surveys are designed to enable the organization to administer they surveys if they so choose. The recommendation is a survey that gives you both options.

The Feedback Report

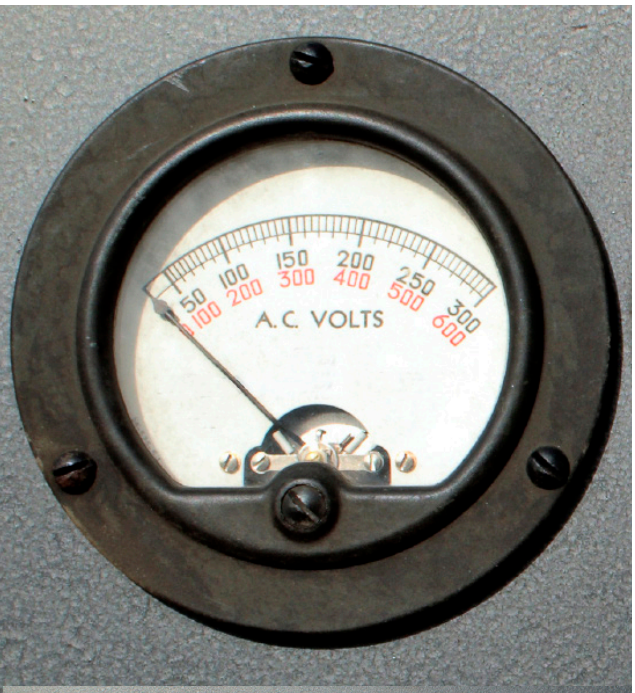
12

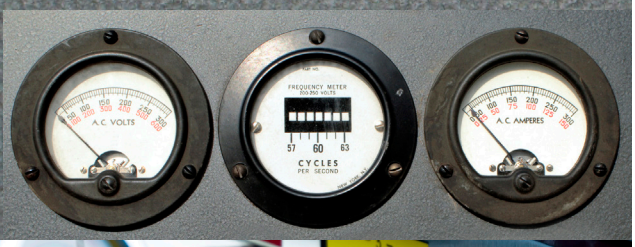
Guideline: Report formats.
Feedback report format refers to the method used to display the feedback. Look for:

- *Breakouts by respondent groups - self, boss, peers, direct reports, customer, etc.*
- *Visual presentations that are easy to read and understand without having to analyze graphs and charts.*
- *Gap analysis - highlighted differences in perspectives between self assessment and other's assessment.*
- *High/low items - easy to access highest and lowest scored items in the survey.*
- *Do more/do less - an indication of whether raters wanted to see the participant do more or less of a behavior. Measurement of behavioral frequency alone does not always indicate whether the behavior is effective.*

13

Guideline: Comparison to norms.
Survey data can be presented as raw scores or normed scores. Raw scores represent the actual rating given by the raters using the rater scale (such as 1 to 5). Normed data compares the participant scores to those of other managers. A combination of raw and normed is desirable. As mentioned in guideline 7, percentile data alone can be misleading and confusing. If normed data is used, it is important to know the database profile against which raw scores are normed. This data profile might include; number of managers in the database, organizational level, educational level, gender ratios, years of experience, and the date of the most recent database update.





Post-Survey Participant Development

14

Guideline: Developmental materials.
The survey should be accompanied by developmental materials which will not only help the participant understand the survey and feedback, but also provide ideas for development based upon feedback results.

15

Guideline: Follow-up assessment.
Of great value, but typically not included, is the opportunity for the participant to conduct a follow-up survey of effectiveness with a few items from the original survey. Re-administering the entire survey can lead to resistance, excessive time and survey fatigue. Ideally, this follow-up survey should be based upon participant goals and should occur within 8 to 16 weeks. A focused follow up assessment can be a great coaching tool.

Facilitator Support and Development

16

Guideline: Survey documentation.
A facilitator or trainer should have access to good support material. A facilitator guide should include the theoretical model behind the survey, the research basis for the survey, reliability and validity data, suggestions for presenting and interpreting survey feedback, and database information.

17

Guideline: Training/facilitation/presentation materials.
In addition to a facilitator guide, support material could include PowerPoint presentation of the survey and theoretical model; feedback sample reports; interpretation exercises and industry and professional norms and samples.

Resources

These guidelines where adapted from the following sources and the experience of Christopher Musselwhite, MSIE, EdD Discovery Learning, Inc.

Feedback that Works (2000). Sloan R. Weitzel, Center for Creative Leadership, Greensboro, N. C.

Feedback to Managers Volume I: A Guide to Evaluating Multi-rater Feedback Instruments (1991). Ellen Van Velsor and Jean Brittain Leslie, Center for Creative Leadership, Greensboro, N. C.

Feedback to Managers (1998). Jean Brittain Leslie & John W. Fleenor, Center for Creative Leadership, Greensboro, NC.

How to Design an Effective System for Developing Managers and Executives (1996). Maxine A. Dalton, Center for Creative Leadership, Greensboro, N. C.

Leadership Development (1987). John W. Gardner, Independent Sector, Washington, D. C.

Maximizing the Value of 360-Degree Feedback (1998). Walter. W. Tornow & Manuel London, Jossey-Bass Publishers, San Francisco.

ISBN 1-931194-31-9 © 2004, Discovery Learning, Inc. Published by Discovery Learning Press
431 Spring Garden Street, Greensboro, NC 27401 USA 336.272.9530 www.discoverylearning.com