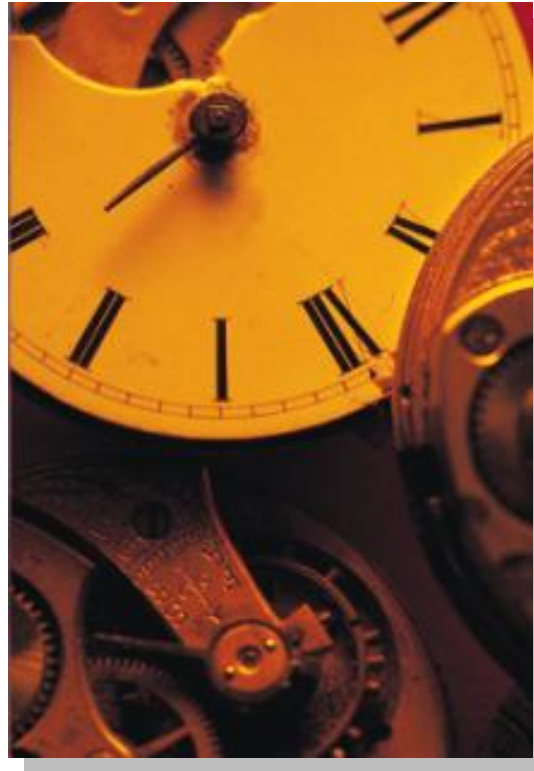


Measuring Leaders

Issues to Consider in Evaluating or Developing a Leadership Assessment Instrument

By Cornelia Sawyer and Robert Kabacoff



Choosing which assessment instrument to use in organizational and individual leadership development is an important but often bewildering decision. There are numerous instruments available, and these vary widely in terms of price, quality and usefulness. In addition, many organizations consider developing their own questionnaires internally. Some seek to measure competencies that have already been created for them. Some desire a customized instrument that reflects their unique culture and language. Some believe it would be more cost effective to develop their own instrument rather than buy one “off the shelf.”

Following are a number of important issues that an organization needs to consider, whether evaluating an instrument that is available commercially, or designing and developing its own customized instrument:

- **Deciding What to Measure**
- **Determining How to Measure It**
- **Proving That Your Measurement is Accurate**
- **Deciding How to Present the Results**
- **Collecting, Managing and Using the Data**
- **Choosing Whether to “Make” or “Buy”**
- **Considerations When Taking U.S. Based Assessment Instruments Global**

A professionally developed, well-researched assessment instrument should take into consideration and meet all of the criteria described in this article.

DECIDING WHAT TO MEASURE

There are three things you need to think about when deciding what to measure: what your underlying model of leadership is, the dimensions of leadership you want to measure, and the actual items you will include in the assessment instrument.

Underlying Model of Leadership: The underlying model needs to make sense and have practical applications to real leadership/management situations.

- Is the instrument based on an underlying model of leadership that is derived from empirical data and/or well-grounded theory? (E.g. underlying, “timeless” human behavior vs. current fads/trends)
- Is it broad enough to describe all of the important components of the leadership role?
- Is it able to accurately assess individual differences in the way people approach the leadership role?

Dimensions (Areas of Interest): The areas of interest that are to be measured need to be carefully chosen and defined.

- Can they be readily measured, or will they be very difficult to measure? (E.g. “trustworthiness”)
- Are they specific and uni-dimensional? If they are too broad or multi-dimensional (e.g. “develops others”), it can be difficult to interpret the meaning of an individual’s score, because a high score can be caused by many different factors.
- Are they areas that can be observed by others? (E.g. behaviors vs. motivations)
- Are they amenable to development or difficult to change? (E.g. behaviors vs. personality traits)

Scale Items: Writing good items is tricky and requires expertise. In evaluating an instrument, it is important to consider the following:

- Are the items carefully written to minimize response bias and social desirability?
- Are there enough items to adequately sample the area of interest?
- Do the items truly measure the area of interest?
- Is the cognitive level required to read and understand the items appropriate for the questionnaire’s intended audience?

DETERMINING HOW TO MEASURE IT

Once you have decided what you want to measure, you have to decide how you want to measure it. This means you have to think about a measurement methodology.

Choice of a Measurement Methodology (Questionnaire Format): The choice of a questionnaire format has broad implications. It affects the users’ reactions and responses to the instrument, the reliability and validity of the instrument, and its usefulness and

appropriateness for its intended purpose. Therefore, it is important to understand and evaluate the advantages and disadvantages of any questionnaire format that is being considered.

- Is the format appropriate for the instrument's intended use? (E.g. measuring knowledge of a content area vs. measuring perceptions of leadership behavior)
- Will the format provide an accurate, objective, and differential measurement of the dimensions? (E.g. scales that will be reasonably independent rather than highly inter-correlated)
- Will the format minimize the effects of response set distortion? (E.g. making it difficult for a respondent to give ratings that are all high or all low, or to manipulate the results in other ways)

PROVING THAT YOUR MEASUREMENT IS ACCURATE

Establishing an instrument's reliability and validity is a critical part of questionnaire development. Reliability and validity information should be readily available for any commercially developed instrument that is being considered, and this information should be carefully evaluated.

Reliability: Reliability refers to the stability or consistency of measurement over a variety of conditions. It is critical for an instrument to demonstrate adequate reliability. If it does not, or if reliability studies have not been conducted, there is no way to tell whether differences in scores represent true differences or merely instability in the scales. Without reliability, an instrument cannot demonstrate validity.

- Have reliability studies been conducted and, if so, what kind? (Internal consistency considers the homogeneity of the scales; test-retest reliability considers the consistency of responses over short periods of time; inter-rater reliability considers the consistency across raters when observer ratings are involved).
- Does the instrument demonstrate adequate reliability?

Validity: Demonstrating validity is an ongoing, cumulative process. It requires many studies to establish that an instrument is appropriate and valid for its intended use. Some of the questions that need to be answered are:

- Does the instrument actually measure what it claims to be measuring? (Construct validity)
- Are dimension scores actually related to effective performance? (Concurrent validity)
- Can dimension scores predict future performance? (Predictive validity)

DECIDING HOW TO PRESENT THE RESULTS

A good deal of attention also needs to be paid to the way questionnaire results are presented to its users. Unless the feedback reports are clear, understandable and meaningful to the recipients, the usefulness of the questionnaire results may be undermined or even negated. Careful attention should be paid to the issue of norming scale scores.



This can be an important feature because it gives a benchmark to help individuals/organizations make sense of their feedback. Being able to differentiate the effect of function, level, industry and other important factors on individuals' scores can provide important insight to the organization and can boost developmental efforts. (Two important issues for organizations to consider if they plan to develop their own customized instrument are how to create a large enough pool of cases to provide appropriate normative reference groups, and whether using a benchmark that consists solely of people within their own organization will be helpful enough to the feedback recipients.)

- Are the scale scores represented in a manner that is clear, understandable, and useful to the recipient?
- Are scale scores normed?
- Who comprises the normative population, and is it a relevant group for the intended feedback recipients?

Individual and Organizational Development: The fundamental purpose of a leadership assessment is to provide insight and information that will enhance leadership effectiveness. Thus, one of the most important elements to consider is how helpful the feedback results will be in prompting and sustaining ongoing individual and organizational development.

- Is interpretation of the scale scores provided?
- Is the feedback useful for development?
- Are developmental options or suggestions provided?
- What kind of information can be gained from the assessment that could be used to promote organizational development?

COLLECTING, MANAGING AND USING THE DATA

Developing a leadership model and an assessment instrument are only the beginning. Administering the instrument and working with the data collected present a new set of challenges.

Administration and Feedback Systems: Administering, collecting and scoring the questionnaires, preparing and delivering feedback, and tracking developmental efforts and progress can be complex, costly and time-consuming – especially if the process involves observer ratings for 360-degree feedback. Nevertheless, these are all critical components which must be planned, developed and implemented in order to ensure the success of a leadership development initiative.

- What processes and systems will be needed for administering the instrument to participants?
- What systems will be needed for scoring the completed instruments and generating feedback reports?

- Are these processes and systems readily available, or would they need to be created and maintained in-house?
- What internal resources would be required to deliver feedback to recipients, provide coaching for their developmental efforts, track their progress, and align these activities with organizational objectives?

Research: Any time an assessment instrument is used to gather data in an organization, a valuable opportunity arises. Using a well-designed instrument in a carefully planned initiative can generate data that may be enormously helpful to the organization. Research conducted on that data may help the organization to learn more about itself and to use that knowledge in its efforts to enhance its effectiveness.

- Does the instrument contain a section that gathers information on outcome variables (i.e. leadership effectiveness) as well as individual leadership practices?
- Is the statistical expertise available (either through the vendor or within the organization) to conduct high quality research studies on the data collected?
- What types of research reports will be generated, and how can these be used to best advantage by the organization?
- What types of safeguards exist to protect the confidentiality of data and the confidentiality of the individuals who initially provided the data?

CHOOSING WHETHER TO “MAKE” OR “BUY”

There are advantages for an organization in developing its own customized leadership assessment instrument. The dimensions chosen and language used may reflect the organization’s culture more exactly. The instrument may be designed to measure specific competencies that have already been developed and set in place in the organization. And, it may be easier to align the feedback results and ensuing developmental efforts with existing programs or initiatives in the organization.

However, there are disadvantages to be considered as well. It requires a great deal of work to properly develop not only the instrument, but also the attendant scoring, feedback and development systems.

In addition, organizations which develop their own customized instruments are likely to find it difficult to establish the reliability and validity of these instruments. Sometimes they simply do not do so because:

In the end, an organization may find that it has spent a great deal of effort to develop an instrument that measures dimensions which, rather than being startlingly new or different, are already contained in one or more commercially available instruments.

- They consider the instrument to be OK because it appears to have face validity (although this is not a true form of validity).
- It is expensive and time consuming, and they may not have the time or resources to invest in these research activities.
- It requires a sophisticated level of statistical/psychometric knowledge that may not be available within the organization.

It is critical to recognize that, if an instrument’s reliability and validity are not established, it is impossible to know if the instrument actually measures what it claims to measure – and

therefore impossible to know whether the results are accurate, meaningful and useful. And, if the instrument is used for evaluation purposes (hiring, promotions, performance appraisal), but the results have not been shown empirically to be related to job performance (through specific validity studies), the organization could be challenged on its use and could face serious legal liabilities.

CONSIDERATIONS WHEN TAKING U.S. BASED ASSESSMENT INSTRUMENTS GLOBAL

Cultural assumptions that are widely accepted in the United States about the nature of leadership, the potential for development of individuals, and the interpersonal dynamics of relationships between bosses and subordinates may not be shared by other cultures. If a U.S.-based instrument is being brought into other cultures, a wide variety of issues must be addressed.

One of the better discussions of cultural issues is included in “Maximizing the Value of 360-degree Feedback: A Process for Successful Individual and Organizational Development”, edited by Walter Tornow and Manuel London (1998, Jossey-Bass), *Chapter 8: Understanding Cultural Influences in the 360-degree Feedback Process*, by Jean Leslie, Maxine Dalton, and Nur Gryskiewicz. We have summarized their advice below.

In preparing for implementation, find out:

- Do individuals in the target country believe adults can develop/learn/improve through this process?
- Do individuals believe the workplace has a role in providing developmental feedback/opportunities for individuals?
- Are psychometric instruments an accepted convention in the target country?
- Do individuals in the workplace accept the idea that asking for information from their boss, peers, direct reports and customers is appropriate and useful?
- Do individuals in the target country accept and believe the feedback will be kept confidential and raters will remain anonymous?
- Is the model of management/leadership represented by the instrument sensible and useful for the target organizational level and culture?
- Is each question in the instrument culturally appropriate?
- What mechanisms are currently in place for post-feedback reaction to the 360-degree process?

When buying an instrument from a vendor, review the technical document to see:

- Are appropriate normative comparisons available?
- Are test adaptation procedures appropriate and sufficient for the intended use?
- Is there evidence of ongoing international research with the instrument to address the unanswered questions?

Enter the country and the feedback session as a respectful colleague and partner to the feedback recipient. Learn from each other the usefulness and limits of the process for that person in his/her organization and country.

Provide training to raters and feedback givers on:

- How the instrument is designed to be used.
- What the cultural influences on the use of response scales are.
- The value of using all the points on the scale.

Acquire knowledge in cross-cultural training on:

- How compatible the theoretical frameworks of development are in the target culture.
- How the feedback will be received by the participants.
- How to introduce 360-degree instruments in cultures that may not be ready for the experience.
- How to coach managers in different cultures to be effective and successful based on the feedback they have received.
- How to create on-the-job developmental experiences for managers to prepare them for assignments outside their home country.

In selecting a 360-degree instrument for use with a multi-cultural audience, look for:

- Evidence to support a claim of comparability and cultural fit.
- Administration instructions that are in target languages to minimize variation across populations.
- Evidence that the choice of testing techniques such as answer formats for questions (i.e. multiple choice, forced-choice, essay), conventions and procedures in giving instructions (i.e. layout, use of graphics), and presentation (i.e. paper and pencil, PC-based software, scannable answer sheets) are familiar to all intended populations.

CONCLUSION

Using an assessment instrument as the foundation for a leadership development initiative can be very beneficial for an organization. It provides a level of knowledge and insight that cannot be gained in any other way, and the data collected can be extremely helpful in both individual and organizational development. However, it is absolutely critical to use an instrument that will provide accurate and meaningful data.

While companies are likely to find that it is very expensive and time-consuming to properly develop and validate a customized leadership assessment instrument, it is critical that they either commit the resources to do this, or carefully evaluate and use a well-researched and validated instrument that is available commercially, in order to ensure the success of their organization's leadership development initiatives.

About the Authors:

Cornelia Sawyer, is the former Director – Product Quality, Management Research Group®, Portland, Maine USA

Robert Kabacoff, Ph.D., is V.P of Research and Development, Management Research Group®, Portland, Maine USA

Additional Management Research Group® Practice Guides and Research Reports can be found on our website at www.mrg.com.



Management Research Group®
Ardeen Mews, 10/11 Marine Terrace
Dun Laoghaire, Co. Dublin
Ireland
Telephone: +353 1 280 4430
Fax: +353 1 280 4434
e-mail: info@mrg.com

Management Research Group®
14 York Street, Suite 301
Portland, Maine 04101 USA
Telephone: + (207) 775-2173
Fax: + (207) 775-6796
e-mail: info@mrg.com