

DEVELOPING ASSESSMENT INSTRUMENTS

Management Research Group®'s Unique Approach to Questionnaire Design

By Cornelia Sawyer and Robert Kabacoff



For over three decades, Management Research Group® (MRG) has been a global leader in the development of assessment instruments used for individual and organizational development. During that time we have faced and overcome most of the common challenges inherent in creating scientifically robust instruments that also collect and report back information in a way that can easily be understood and used by end-clients.

The goal in developing any assessment instrument is to carefully measure a specific area of interest in a way that will produce accurate, meaningful and useful data. The design process consists of three steps, each of which is equally important in determining its utility:

- Deciding what to measure
- Deciding how to measure it
- Deciding how to present the results (feedback)

In this article we will be discussing common methodologies for questionnaire design and why MRG® has chosen a unique design strategy. We use this design approach for all of our instruments, but to help readers gain a better understanding of what MRG has done and how it works, we will be using our *Leadership Effectiveness Analysis*™ (LEA) questionnaires as an example.

Deciding What to Measure

How do you determine the elements of the entity you propose to measure? For example, if you wish to develop a leadership assessment, you need to break down the concept of “leadership” into its components, and then design an assessment instrument that measures these components.

The LEA questionnaires represent MRG's efforts to do exactly that. The LEA's underlying model of leadership is based on knowledge gained from years of experience in the field and observations of thousands of individuals in organizations. These were real world situations - real leaders in real businesses. The test developers observed the kinds of issues these leaders faced, and the behavioral strategies they used to deal with their issues and to operate, both effectively and ineffectively, in their specific leadership roles.

These observations led MRG's test developers to create a "neutral" model of leadership - one that contained no inherently "good" or "bad" leadership behaviors, but that showed universal behaviors that could be observed and applied, and used to describe any type of leadership role in any type of setting. This was a model that measured elements of leadership based not on current trends and fads, but on underlying human behavior. It was a model that could differentially describe the wide variety of ways that people deal with the leadership challenges they face on daily basis. Years of empirical testing has allowed us to refine the 22 leadership practices which comprise the model, as well as the questions used to assess it.

Deciding How to Measure the Areas of Interest

Once you have decided upon the dimensions you wish to measure, the question becomes, "How do you measure them?" This is accomplished by writing a pool of items that sample the attributes you believe to be associated with each dimension. The items for each dimension will be summed together to form a scale; it is the scale which is used to represent a person's standing on a particular dimension. The way the items are crafted and assembled will depend to a large extent on the type of questionnaire format to be used. Thus, the next step is to choose a questionnaire format.

Questionnaire formats abound, and each has its own proponents and set of advantages and disadvantages. Test developers face at least two major problems in designing questionnaires that measure attitudes and behaviors:

- Raters tend to rate people in a global, all-or-none fashion.
- Test takers attempt to present themselves in a favorable light.

These factors can reduce a questionnaire's ability to provide a valid measurement of the dimensions it is supposed to be measuring.

In an attempt to offset these tendencies, MRG® has created a questionnaire design approach that makes it very difficult for respondents to unconsciously affect or consciously manipulate results. Our unique questionnaire design (semi-ipsative) blends elements from two distinctly different formats (ipsative/forced-choice and anchored rating scales) in order to maximize the strengths and minimize the weaknesses of each.

A major consideration in questionnaire design and development is the issue of response sets and the impact these have on a questionnaire's ability to accurately and validly measure both the underlying dimensions and differences within and among respondents. Response sets are patterned ways people respond to questionnaires, in-

cluding acquiescence (saying “yes” to everything), nay-saying (saying “no” to everything), and using a limited range of the response scale (giving ratings that are all high, all moderate, or all low). In addition, individuals may respond based on social desirability (choosing the “right” answer or the one that makes the respondent look better), and other conscious efforts to choose answers that will manipulate the results in a certain way.

In anchored rating scales (see Figure 1), each item is a separate question that is considered and rated independently of all the other items. Dimensions are not compared to each other, and the rating given one item has no explicit effect on other items. Anchored rating scales are particularly susceptible to response set distortion (although this can be controlled to some extent by careful writing of items).

Figure 1: Sample Anchored Rating Scale Format

Instructions: Circle the number that represents your agreement with the following statements.

	Least like me	Not much like me	Neutral	Somewhat like me	Most like me
1. In supervising people, I am tactful.	1	2	3	4	5
2. In supervising people, I am demanding.	1	2	3	4	5
3. In supervising people, I am easy to please.	1	2	3	4	5

In a forced-choice question format (used for ipsative measurement; see Figure 2), each question consists of several items, each of which represents a different dimension. Respondents cannot consider and rate each item of a question separately, but are “forced” to compare and rank-order them. The rating given to one item in a question affects the scores of the other items paired with it. If the respondent is given equally positive or socially desirable items to choose from, a forced-choice format can be quite effective in limiting response sets - a major asset in questionnaire design.

Figure 2: Sample Ipsative Format

Instructions: Rank order the following statements by indicating which is most like you, next closest to you, and least like you:

In supervising people, I am	Rank Order:
a. tactful	most
b. demanding	least
c. easy to please	next

For MRG’s purposes, however, ipsative measurement has a major drawback. While it shows the relative order or preference within the individual, it does not show the magnitude of the preference; thus, it does not allow comparisons between individuals. Anchored rating scales do provide a measure of magnitude - an “absolute” amount that allows comparison among individuals. MRG considered this to be an essential feature for its role-specific assessment instruments, whose purpose is to differentially describe how people approach their work roles. Thus, MRG incorporated a rating scale into the forced-choice question format. It is this feature that makes the MRG design *semi*-ipsative - a hybrid of the ipsative and anchored rating scale formats (see Figure 3).

Figure 3: MRG's Semi-Ipsative Format

Instructions: Choose the item that is MOST like you and rate it a 5 or a 4; then choose the item that is NEXT closest to you and rate it a 3 or a 2; leave the last item blank. (Blank item is assumed to be LEAST like you and is scored as 0.)

In supervising people, I am	MOST	NEXT
a. tactful	5 4	3 2
b. demanding	5 4	3 2
c. easy to please	5 4	3 2

MRG has found that a semi-ipsative format, with items balanced for social desirability, has a number of benefits:

- It limits response set distortion by respondents (a benefit of the forced-choice format).
- It is more complex and thus may be less transparent, which makes it more difficult for respondents to consciously manipulate results (a benefit of the forced-choice format).
- It explicitly shows the relative ordering or preference within an individual for the dimensions being measured (a benefit of the forced-choice format).
- It provides a measure of magnitude that allows comparison among individuals (a benefit of the anchored rating scale format).
- Its dimensions are quite independent, with low scale inter-correlations (a benefit of the hybrid semi-ipsative format). This increases the questionnaire's ability to validly measure its underlying dimensions. (Anchored rating scales may have higher scale inter-correlations, again due to response sets. Scale inter-correlations in ipsative measurement are not meaningful, because they are artificially produced by the questionnaire design.)
- Scores tend to discriminate well among individuals (a benefit of the hybrid semi-ipsative format). In contrast to many anchored rating scales, the semi-ipsative approach typically yields distributions of scores that are widely dispersed over the range of possible values. This makes the questionnaire more accurate and effective in measuring and displaying real differences within an individual and also among individuals. (Anchored rating scales often demonstrate significant skewing. For example, on a 7 point rating scale, most people may choose 5, 6 or 7, while relatively few people choose 1, 2, 3 or 4. This limits the degree to which individuals can be differentiated.)
- Forming the forced-choice questions in triads may also help to decrease rater fatigue and resistance because it makes the questionnaire seem shorter than it actually is. (For example, the LEA Self Questionnaire seems to contain only 84 questions, although it actually consists of 256 questions).

As with any questionnaire design, there are some drawbacks to MRG's semi-ipsative format:

- It is complex; this makes it more difficult to construct and, once constructed, very difficult for the test developer to change.
- It requires a higher cognitive level for the test taker in order to read and understand the instructions.
- There are some useful statistical techniques that work well with anchored rating scales but do not apply to this form of measurement because of its forced-choice format.

On balance, however, we feel that the benefits of this design outweigh these drawbacks and make a strong case for its use.

Deciding How to Present the Results (Feedback)

Clearly, a good questionnaire needs to have a solid theoretical base and to be well constructed. However, a good deal of attention also needs to be paid to the way the 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.

To calculate results, items are summed to produce a raw score for each scale. However, without a context it is difficult to interpret the meaning of these raw scores. To aid in interpretation, we - like most test developers - transform the raw scale scores into normed scores. This allows individuals to gauge how their scores compare to a large sample of similar individuals.

When norming results, the important issue is who will comprise the normative sample. In order to make useful comparisons, the people in the normative sample should be as relevant to the individual as possible. For example, in assessing leadership practices, it would not be particularly helpful to compare the CEO of a multinational corporation to a normative population of college students.

The databases for the LEA questionnaires have unusual breadth for commercial psychological tests. This is due to the centralization of test processing and the generous contribution of data by questionnaire respondents. To date, over 400,000 individuals have completed the LEA Self and Observer questionnaires, and LEA instruments have been used within more than 5,000 organizations worldwide. A database of this magnitude allows MRG to create both specific and comprehensive normative groups. This provides recipients with the assurance that the norms used in preparing their reports reflect a reference group that makes sense for their specific situation.

In deciding how to present the normed results in the feedback reports, MRG has chosen to convert raw scores into percentile ranks rather than standard scores, because these are familiar to and therefore more easily understood and interpreted by the

people who are receiving feedback from these instruments. Thus, they know from their scores how many individuals in the large normative sample scored above and below their own scores.

Finally, we should note that in MRG reports, feedback is provided at the scale, but not at the item level. MRG has chosen this method of feedback for two important reasons. First, individual items tend to be poor single measures of the dimensions we are interested in. It is the summation of numerous items measuring the same thing that is useful. Second, item level reporting vastly increases the amount of information to be considered, which can often confuse rather than help the recipient. Thus, we feel that presenting scale scores rather than item scores makes the feedback more reliable, accessible, understandable and, in the end, more actionable.

Conclusion

Ultimately, the test of any questionnaire is the practical utility it possesses. The hybrid questionnaire design developed by MRG incorporates a number of strengths from the standpoint of questionnaire development. From the standpoint of empirical testing, MRG's semi-ipsative assessment instruments have been used extensively with hundreds of thousands of individuals in thousands of organizations. The results show that this questionnaire design has exceptional construct and predictive validity, and has produced feedback that is extremely useful to both organizations and individuals in improving their role-related performance.

About the Authors:

Cornelia Sawyer, is 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