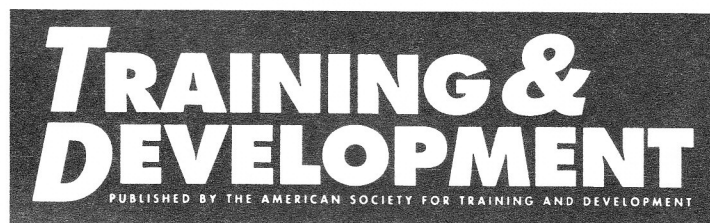


Everything You Always Wanted To Know About Employee Surveys

BY KAREN B. PAUL AND DAVID W. BRACKEN

Reprinted from



JANUARY 1995

AMERICAN SOCIETY FOR TRAINING AND DEVELOPMENT
Alexandria, Virginia

Everything You Always Wanted To Know About Employee Surveys

BY KAREN B. PAUL AND DAVID W. BRACKEN

Here's a way to design effective employee surveys by answering some commonly asked questions.

More than 70 percent of organizations survey their employees either annually or biannually. Yet, there doesn't seem to be a lot of specific, practical, and documented information about creating employee surveys. It isn't easy to construct surveys that elicit accurate data from employees to determine organizational needs. The following strategy can help you design winning surveys.

Q & A

Here are answers to 12 of the most commonly asked questions about designing surveys, based on our own research and consulting experience.

Question 1. Is there a best time to conduct surveys? It depends on the required response rates. Obviously, it's best not to conduct surveys that require high response rates during times when many employees are

unavailable—such as vacations and holidays.

The timing also depends on how the survey data will be used. Survey administrators may want to schedule surveys during regular business-planning cycles so that the survey results can be used in strategic planning and budgeting. Or, the survey's administrator may want to schedule a survey so that the results will be published just before performance appraisals are conducted.

Quarterly sample surveys that target only a small segment of the employee population can do spot checks on employee attitudes. The results can be compared over time with results from other employee surveys and measurement instruments. The quarterly surveys may include one-time items on issues currently affecting employees.

Overall, survey administrators should ensure that employees view participation in surveys as an important business event.

Question 2. Which are best: sample surveys or census surveys? Surveys of a small, selected segment of the employee population sometimes generate results that are just as reliable as results acquired through more expensive census surveys, which involve the entire organization. But when the employee population is small, sample surveys aren't as reliable as census surveys.

When using a sample approach, the survey administrator must know how to calculate the required sample size to represent the population adequately. To calculate sample size, the administrator must have an accurate list of the demographic groups that will participate. Using results from

past surveys, the administrator may want to predict response rates so that the survey design is more likely to produce results that are reliable statistically.

The survey administrator also must be able to accept the large margins of error (confidence intervals) typically generated by sample surveys. Such intervals can have a direct effect on various comparisons among groups—such as demographic and trend comparisons.

Sample sizes are calculated according to a margin of error that the organization deems acceptable, typically plus or minus 5 percent. The smaller the margin of error is, the larger the required sample size. If the sample becomes too large, the survey may be less cost-effective than a census survey.

Reports on survey results must note the margin of error or confidence interval to indicate the degree of confidence with which one could say that the results accurately reflect the truth. It's important to report the confidence interval, to avoid misinterpretation or misapplication of the data. When the confidence interval is not taken into account, differences that appear statistically significant may, in fact, be due only to chance.

As a communication tool, sample surveys provide only some employees with the opportunity to give input. Consequently, sampling may dilute organization-wide communication and decrease employees' commitment to accepting the survey results and any ensuing actions.

Employees selected to take part in a sample survey may wonder why they were chosen, and they may worry about the confidentiality of their responses. They may also think—out of loyalty—that their responses should reflect the views of co-workers who aren't participating. Employees not selected may worry about being excluded.

Sampling places some limitations on reporting. Reports should include only the groups specified in the sample. In interpreting survey results, one shouldn't assume that the data apply to similar groups within the employee population or to groups that are new to the population since the survey.

Some surveys select participants based on job classification or other organizational structures. The results of such surveys may not be valid if the organization restructures after the survey and respondents are assigned to other work areas. Still, people will request survey reports on work groups and trends, to fit new organizational charts.

To make results adaptable, surveys should include demographic coding for group functions, departments, and locations—even though it's difficult to design coding that will ease respondents' suspicions about confidentiality and accuracy.

Question 3. How long should surveys be? This question is open to debate. Some studies show that

AN EVEN
NUMBER OF
RESPONSE OPTIONS,
WITH NO NEUTRAL
MIDPOINT, TENDS
TO FORCE
RESPONDENTS TO
TAKE A STAND

shorter surveys get higher response rates than longer ones. Other studies suggest that the number of questions is irrelevant, as long as respondents are interested in or committed to the aims and outcomes of the survey. Interestingly, different demographic groups may prefer different-length surveys. For example, managers tend to like surveys of three to five pages; nonmanagerial employees tend to prefer surveys of 10 to 12 pages.

Generally, experts say that a survey should take about 30 minutes to complete. Given time and cost constraints, 80 to 100 questions—excluding questions about demographics—is an ideal length for most surveys.

Question 4. Does it matter how many response-scale points are used? Different surveys use different response formats. The typical format is a rating scale with response points

arranged from low to high, along equal intervals—for example, a scale of 1-2-3-4-5. The five-point Likert scale probably is the most widely used rating scale.

The number of points on the scale depends on the kind of information required. Some rating scales are designed to assess people's opinions on certain topics in a positive or negative direction—for example, "agree or disagree" and "satisfied or dissatisfied." On such scales, an odd number of response points may encourage equivocal responses, depending on the label at the midpoint. A neutral midpoint label such as "not sure" lends itself to uncertainty. Our study shows that 20 percent of respondents choose a neutral option if one is offered.

An even number of response options with no neutral midpoint tends to force respondents to take a stand, though people with strong attitudes typically give the same ratings with or without a neutral midpoint. In fact, overall response rates are sometimes lower on surveys that don't have neutral midpoints. The choice of whether to include a midpoint also depends on the specific content of the question. Sometimes, it just doesn't make sense to allow respondents to be uncertain.

To help interpret respondents' use of the neutral midpoint to express uncertainty, the survey should ask a prescreening or "filter" question: "Do you know enough about the topic to have an opinion?" Or, the survey can offer a "don't know" option outside of the range of the rating scale, especially to identify respondents' levels of knowledge about various organizational issues.

If the survey author determines that the target audience is able to make fine distinctions between various options, more response points may yield results that can be used to interpret or track data across time to detect trends.

Question 5. Does it matter what the rating-scale anchors are or how many scales are used? It's usually best to label a scale's high end with the most positive anchor. That's because most organizational surveys measure positive concepts such as effective communication, competitive edge, and

teamwork. When analysts calculate means or summary scores, high numbers will immediately identify positive attributes or attitudes. This approach helps simplify analysis.

An example of a scale with the most positive anchor at the high end is “5—highly satisfied,” “4—satisfied,” “3—neither satisfied nor dissatisfied,” “2—dissatisfied,” and “1—highly dissatisfied.”

It’s important to choose an accurate label to attach to the midpoints of three-, five-, and seven-point scales. Respondents might apply different meanings to the same options—particularly when the rating scale is supposed to represent different degrees of agreement or satisfaction. The midpoint should be part of a continuum, not a completely different concept.

The number of response points on the rating scale helps determine the anchors or labels. On surveys with more than six response numbers, the survey author might label only the end points—for example, “strongly agree” and “strongly disagree” or “always” and “never.” But when the survey results will be reported as averages, survey designers should label each point.

The labels should correspond to the psychological values respondents are likely to assign. Depending on what those values are, the type of rating scale can skew results. One study of an upward-feedback instrument compared results from identical questions using two different scales—a frequency scale and a satisfaction scale. The different rating scales generated very different response patterns on several of the same items.

It’s best to use only one type of rating scale in a survey. Switching rating scales tends to confuse respondents and cause errors and frustration. Changing scales may hamper the reporting of results. And too many scales can make a survey too lengthy.

Sometimes the wording of items dictates which type of scale to use. Here are some examples.

- ▶ “Our benefits are as good as or better than those offered by other companies in this area.” (agreement)
- ▶ How satisfied are you with your

benefits plan? (satisfaction)

- ▶ How would you rate the benefits plan? (grade)

▶ To what extent does the benefits plan provide you and your family with adequate coverage? (extent)

- ▶ “My benefits cover my medical requirements.” (frequency)

Last, the survey instructions should tell respondents to leave an item blank if they think they don’t know enough about it to have an opinion.

Question 6. Do the demographics belong at the beginning or end of a survey? Surveys with the demographics at the end tend to produce higher response rates. Once respondents have spent time answering the other survey questions, they’re likely to complete the demographic section at the end because of a “completion tendency.” In addition, respondents may be tired by the end of a survey. They may be more likely to complete the survey if the final items are the typically unchallenging demographic questions.

Generally, demographic questions about age, gender, and education aren’t threatening. But such questions may intimidate certain respondents (such as some members of minority groups) or may seem threatening when combined with other demographic questions. When controversial demographic questions are placed at the beginning of a survey, respondents may react negatively and not complete the other questions.

In organizations in which trust is an issue, achieving full participation on a survey is more important than getting demographic information. In such situations, it’s best to keep the demographic questions to a minimum. Fortunately, survey information without the demographic information is still usable.

Another way to reduce respondents’ anxiety about demographic questions is to make such questions optional or to explain how the information will be used. Many people worry about being identified. They may not know that most survey administrators don’t have the time, energy, or resources to track people’s opinions and identities.

For example, the survey can read: “In this section, we are asking for

some information about you to help us understand the survey data. We will not use the information to identify you. We will include your answers with those of other employees so that your responses remain anonymous.”

Question 7. Is it better to combine items from various topics or to group all of the items about a single topic in the same place on a survey?

Studies show that responses fall into “sets” when survey authors cluster certain items under category headings. Response sets imply that respondents are answering all of the items similarly, based on their attitudes toward the category heading. But categories don’t completely determine the types of responses received. A well-designed survey clusters items that have a common basis, sometimes derived from statistical analysis.

Evidence suggests that survey developers can avoid response-set bias by carefully selecting the category headings. For example, using “shaping excellence” instead of “rewards and recognition,” may remove negative implications for some people.

TO REDUCE
ANXIETY ABOUT
DEMOGRAPHIC
QUESTIONS, MAKE
THEM OPTIONAL
OR EXPLAIN HOW
THE INFORMATION
WILL BE USED

Surveys that group items by categories can be effective in communicating an organization’s priorities—especially when the headings are based on the organization’s vision or values and when the items define desired behaviors. Category-type surveys also help employees see a direct relationship between the survey instruments and the survey reports.

Question 8. Is it important to have some negatively worded items on a survey? A form of response bias occurs when respondents consistently choose one response option regardless of the question, even drawing a line through one answer column. In such cases, one has to wonder whether the respondents even read the questions. Including negatively worded items helps deflect this blatant type of response bias by forcing the respondents to choose different options to express their opinions.

An example of a negatively worded item: "Our benefits are worse than those offered in most companies."

But negatively worded items don't always work as intended. Respondents may view such items as having certain meanings due solely to their negative phraseology. And negative wording doesn't work well with some types of scales—such as satisfaction scales.

When the survey is to be a two-way communication tool, negatively worded questions might imply that top-level managers think that employees are harboring negative attitudes. Respondents may feel as if they're being tricked, and that the survey is some kind of test. Such feelings can reduce response rates. Negatively worded items may cause some people to resist management actions that are based on survey results.

The survey administrator should monitor negatively worded items closely during data collection, analysis, reporting, and interpretation. If survey analysts are using summary scoring—which aggregates or adds up the ratings of similar items within a category—negatively worded items may require them to reverse the scores of negative items before adding those scores to the scores of positively worded items. Some administrators find this kind of scoring and analysis difficult to do.

When a survey contains negatively worded items, the survey administrator should

- ▶ decide early on how results will be presented in reports
- ▶ request a sample report of survey results and ask someone outside of the survey process to interpret it
- ▶ request item frequencies on the

raw data from negatively worded items and from some positively worded items surrounding the negative ones

▶ ensure that the users of the survey results handle data according to the survey administrator's instructions and that reports on the results explain how to interpret negatively worded items.

One way to minimize response bias is to keep surveys short, explain to respondents why surveys are important, and share information about how the organization will use the data. It makes sense that when people understand how survey data will be used—and when they aren't fatigued by an over-long survey—they will respond more carefully and completely.

Question 9. Are norms useful? Norms are descriptive statistics that enable survey administrators to compare mean survey results with the scores of people in a defined population. A major consideration in using norms on surveys is the quality of the information on which the norms are based. Do the norms represent the types of companies with which the survey organization wants to be compared? Is the data base updated and purged of old information at least every two years?

By definition, a norm is an average. Using norms for the purpose of comparison may imply that an organization is satisfied with average parameters. Many survey administrators think that norms give managers a false sense of security when employees achieve or exceed averages. A more proactive approach is to establish a benchmark for continuous improvement, based on past survey results. The ultimate goal should be to achieve the organization's vision, as determined through past surveys.

Sometimes, the external suppliers of survey norms require that the norms be used without modification. In such cases, survey authors may have to word items in specified formats that don't meet their needs. They may also have to use whole sets of normative items, including some that are irrelevant to the objectives of their organizations.

It's best to use the survey organization's own vision to define and

establish benchmarks. If normative data are still required, the data provider should describe and ensure the high quality of the normative data base.

Question 10. Which is better: custom surveys or standardized surveys? Many standardized surveys yield reliable and valid information; customized surveys can yield invaluable organization-specific information. The primary considerations are fit, cost, and the way survey results will be used.

Fit refers to the extent a standardized or off-the-shelf survey is consistent with an organization's priorities. Many standardized surveys require a company to buy into someone else's definition of an effective organization.

When a survey is used mainly as a communication tool, its message and results should convey the organization's own values. Some standardized surveys can be modified. But if the survey aims to communicate an organization's vision, then a customized version is best. If the organization is concerned about job stress or job satisfaction, a standardized survey may be more useful.

With standardized surveys, users pay only a part of the development costs, compared with paying in full for the development of customized surveys. Small organizations often find the development of custom surveys to be too expensive, unless the surveys are used repeatedly.

Question 11. What are the differences between mail-in and on-site surveys? Generally, the benefits of on-site administration outweigh the seemingly high cost. Still, opponents of on-site surveys point to the cost of taking employees off their jobs. They may mistakenly assume that employees don't complete mail-in surveys during work hours.

The main consideration against using mail-in administration is a typically lower response rate—about 60 to 75 percent lower than the response rate for surveys administered on site, which tend to produce about a 90 percent response rate. Yet, it may be impractical to administer surveys on site in organizations that have several sites or that lack meeting rooms large enough to accommodate all respondents.

Remember: Employees look for management's commitment in the survey process. Asking employees to fill out surveys on their own time is hardly a sign of management support. On-site administration presents an opportunity to communicate the commitment of senior-level managers. One company showed employees a video featuring the CEO giving his commitment to the survey and promising to take action on the results.

Another benefit of on-site administration is a faster turnaround time. Sessions can take place in just a few days, with survey administrators present to answer respondents' questions.

Question 12. Is a minimum number of respondents required to produce reliable results? On surveys of a particular work group—employees who report to the same supervisor—reliability is partly determined by the size of the group. Large groups tend to generate more reliable and consistent results. The results from small groups

are more affected by the answers of individual members. As group membership or personal attitudes change in small groups, results can vary widely.

Confidentiality is a big concern, especially in organizations with limited experience in conducting surveys. In such situations, it's best that an organization initially survey a large group until employees trust that their confidentiality will be protected in future surveys.

Traditionally, survey experts consider 10 respondents to be the minimum size for reliable reporting of a particular work or functional group. One drawback is that most organizations have some work groups of fewer than 10 employees. Problems can arise when the supervisors of such groups don't get survey results—especially in the areas of career development and performance appraisal. Consequently, an unofficial industry norm has emerged: a minimum of five respondents.

There are many decisions to make along the way when constructing employee surveys. What works well is often learned the hard way. Knowing some answers to commonly asked questions about surveys can help ease the process. ■

Karen Paul is an industrial and organizational psychologist in organizational learning services at 3M Human Resources, 3M Center 225-1N-13, St. Paul, MN 55144-1000;

OrgVitality

David W. Bracken, Ph.D.
Vice President



PO Box 88491
Atlanta, GA 30356
Phone: (402) 617-1316
DavidBracken@orgvitality.com
<http://dwbracken.wordpress.com/>