

Program Evaluation: Issues in Survey Design

By **Sharon Gross, Ph.D.**, Attitude Research Litigation & Organization Consultants

Academic Citation: Sharon Gross, "Program Evaluation: Issues in Survey Design," Kravis Leadership Institute, *Leadership Review*, Spring, 2004.

About the Author: **Sharon Gross** received her Ph.D. in Social Psychology from the University of California where she is a Research Assistant Professor. She conducts study and questionnaire design, survey implementation, statistical analyses, and jury consultation for Attitude Research Litigation and Organization Consultants (www.attituderesearch.com). Attitude Research designs and conducts research on organizational, marketing, political, or litigation issues.

Contrary to popular belief, program evaluations can be a win-win proposition for all parties involved -- students, lecturers, and program designers. The "trick" to achieving valid (true) and reliable (repeatable) results lies in the questions themselves, their order, and the scale used for responses.

There are two circumstances in which students give their evaluation of a program: those in which they receive a grade (schools, colleges, universities), and those in which they do not (conferences, seminars). Herein these will be called Schools and Conferences respectively. It is well known that when students evaluate a course for which they receive a grade, there is a high correlation between the grade they expect to receive and their evaluation of the program. That is, the higher the grade they expect, the better their evaluation and vice versa. It is usually for this reason that professors, lecturers, and guest speakers are often reluctant to participate in surveys. They think that the results do not reflect a true picture of the value of their teaching, particularly in difficult classes. There are, however, ways to design a survey so that this pitfall is avoided. This will be addressed below in the section on Survey Questions.

SURVEY DESIGN

As indicted, the goal of any survey is to yield responses that accurately reflect the views of the participants (validity) such that similar results would be found if the survey were re-administered (reliable).

At Schools, those interested in the evaluations are the departments (as a way of evaluating their professors), the professors (as a way of evaluating themselves) and the students (as a way of choosing classes for the next semester). Most schools, however, do not share their information with the students. This has caused student bodies to finance evaluations on their own, but with the approval of the participating instructors.

At Conferences, it is the organizers who are most interested in the evaluation of their leadership Conferences. In this way they are able to monitor the good, the bad, and the ugly to make improvements for the future. But, just because some event received high praises at one time, this does not guarantee continuing praise. Life is a dynamic process and times change the importance of various factors. Thus, it is important to continue to conduct evaluations even if all is well.

As stated, validity (truthfulness) of the results is critical. Thus, it is important to be a careful reader of reported outcomes. For example, if a report says, "The participants rated the program very highly," can you take this at face value? The unequivocal answer is no, you cannot. These words must be backed up with the questions asked and the statistics that tell you the range of the responses (e.g., 1=extremely poor and 5=extremely good), the number of respondents, and either the frequency or average of the numerical answers. In this way, you can see for yourself just what "very highly" means.

The best way to design a survey is to think backwards. That is, ask yourself, "What do I want to know?" Conferences might be interested in participants' perceptions of the brochure and advertising, the facility used and its accommodations, as well as the degree to which they enjoyed the social events and field trips. For both schools and conferences, topics of interest include those about the course itself, the professor or lecturers, and the students or participants who took part in the survey. The latter is important in answering the question, "who is giving this evaluation?"

There are, however, trade-offs in everything, and survey design is no exception. One constraining factor is the length of the survey. In addition to increased costs (due to extra paper, copying, and increased analyzing), you don't want it to be so long that your participants become fatigued or turned-off.

Fatigue can cause participants to get "lazy" in their responses by simply agreeing (or disagreeing) with all survey statements from the point where they became tired. There is a way, however, to design your survey so that you can remedy the potentially biased effects caused by fatigue or laziness. It requires that you make two forms of the survey -- one forward and one backward such

that the first question of one form is the last question of the other form. This allows you to test to see if there is a reliable difference between the responses to any question. If there is no difference, then the results can all be pooled. If there is a difference, you can eliminate the responses from the questions that occurred in the second half of the questionnaire and keep the ones placed in the first half. This does, of course, reduce your number of data points by half.

Another reason for invariant responding is described by the Halo Effect¹ or its reverse, the devil effect (e.g., Cook, G. I., Marsh, R. L., & Hicks, J. L., 2003). If a participant experiences an especially positive or negative event during the program, feelings from this specific circumstance can spread across the entire event causing the participant to answer all survey questions in kind (either positively, or negatively). Of course, if the response is consistently positive due to the halo effect, no one would complain, but if it occurs as the reaction of a disgruntled participant, the result pulls down the average response of the group -- lowers the mean. A solution for this occurrence is to analyze the data for outliers. Outliers are responses that deviate from the others by a great degree and there are statistical ways to eliminate this bias².

Once you have figured out what you want to know, you have to write questions (and/or statements) that will allow you to unambiguously get the answers.

SURVEY QUESTIONS

Most people have heard the computer slogan "Garbage In, Garbage Out" or GIGO for short. This warning aptly applies to the design of survey questions (e.g., Krosnik, 1999; Schwartz, 1999). There are a number of pitfalls that must be avoided or taken into consideration when writing the questions or statements.

¹ "The tendency to see individuals as possessing all positive characteristics and none that are negative is termed the halo effect." (Gergen & Gergen, 1981, p.56).

² Outliers are points that lie perhaps four or more standard deviations from zero. A remedy is to winsorize the data by replacing the deviant response with the next highest (or next lowest) response. One degree of freedom is lost (Dixon & Tukey, 1968).

As stated, when students expect a good grade, they give high evaluations and vice versa. To mitigate this circumstance, we always suggest that the statement "I have learned a lot in this class/course" be included³. In this way even the students who expect a poor grade can demonstrate the effectiveness of the teaching.

Another pitfall to avoid in question construction is use of compound sentences. For example, "The professor was fair and knowledgeable." Because this sentence requests a response to two characteristics of the professor, the answers will always be ambiguous.

Also, heavily polarized words need to be avoided. For instance, in a study by Rugg (1941), people were less likely to "forbid" anti-American speech than they were to "not allow" it. That is, it was easier for people to say that something should not be allowed than to say it should be forbidden. Thus, heavily laden words within the questions/statement should be carefully considered.

QUESTION ORDER

The order in which questions are placed within the survey can keep results from being valid and/or reliable (e.g., Bartels, 2002). One must be cognizant of the possibility that an answer to a former question can affect the response to a latter one. For example, after reporting that you "strongly agree" with the general statement that the professor facilitated your learning, it may be difficult to "disagree" with the specific statement that the professor often gave positive feedback. Thus, special attention must be placed on the order of the questions.

SURVEY SCALES

Another critical factor in survey design is response options. In order to quantify responses to questions, numbers must be assigned to each possible answer. For example, if you ask a yes/no question, you might assign "1" to the yes answer and "2" to the no answer. Likewise, you might use a scale with responses of "strongly agree," "agree," "neutral," "disagree," and "strongly disagree." For these answers you might assign the numbers 1 through 5 respectively, or you might assign the numbers +2, +1, 0, -1, and -2 respectively. A common concern when using this type of numerical assignment is that there might be non-equivalency between numbers. For example, it is possible that a response of "Strongly Disagree" (-2) should have been more heavily weighted than a response of "Strongly Agree" (+2) meaning that the strength of disagreement feeling placed it farther below the neutral midpoint (e.g., -3.2). Likert (1932) suggested that this problem goes away when evaluating a

³ By including this question, you avoid the pitfall of previous surveys of students' evaluations (e.g., Greenwald & Gillmore, 1998; APA Monitor, 1998). Namely, many evaluations are skewed by the grades the students expect to receive. This question is the "great equalizer" in that it addresses a critical aspect of students' perceptions of the course regardless of the grade they expect.

person's position on a single topic by summing all of the responses with the positive/negative direction held constant in the assignment of response numbers (e.g., "5" means a pro-topic position and "1" means a con-topic position).

A Semantic Differential scale (Osgood, Suci, & Tanenbaum, 1957) is also commonly used. Usually a 7-point or 9-point bi-polar scale is used with each end of the scale labeled with polar opposites (e.g., good/bad, liked/disliked). This type of scale is useful when you have a large number of participants because you can see if your responses form a normal distribution or are skewed.

A common mistake in many surveys is to label the midpoint of the scale "don't know." The problem that this raises is that it erroneously presupposes that people do not maintain a neutral position even when they are well informed. Therefore, there is no accurate response choice for those participants. Thus, the midpoint of a scale should be labeled "neutral" or "moderate." "Don't Know" should never be an option for an answer. Instead, this response can be inferred from the missing data.

It is also possible to have no middle position. That is, it might be desirable to force people to show their leaning. This can be accomplished by using an even number of options (e.g., 1=Strongly Disagree, 2=Disagree, 3=Agree, 4=Strongly Agree). In this way you eliminate all middle-of-the-roads by forcing them to choose a response direction.

Sometimes it is useful to have a series of items rank ordered. In this case, it is necessary, for example, for participants to place a number between 1 and 5 next to five items to be ranked. Directions warn them that they must be sure to use each number only once. However, even with college students, a large percentage does not do this correctly. Many will invariably use the 1-5 as a good/bad scale rather than a ranking scale. One way to avoid this confusion is to place ranking questions first, before respondents are introduced to a 1-5 agree/disagree scale.

COMMENTS

One of the amazing findings from student evaluations is that what one person finds wonderful (e.g., interaction with students), another student will detest. This is most obvious when comments are requested. For this reason, it is important to solicit comments to capture the essence of the groups feeling about the course/sessions/instructor.

SUMMARY

As you can see, designing a survey to accurately capture the feelings and perceptions of participants is not an easy task. It is, however, an extremely useful tool when appropriately conducted.

References

- Bartels, L. M. (2002). Question order and declining faith in elections. *Public Opinion Quarterly*, 66, 67-79.
- Cook, G. I., Marsh, R. L., & Hicks, J. I. (2003). Halo and devil effects demonstrate valenced-based influences on source-monitoring decisions. *Consciousness & Cognition: an International Journal*, 12, 257-278
- Dixon, W. J., & Tukey, J. W. (1968). Approximate behavior of the distribution of winsorized t (trimming/winsorization 2). *Technometrics*, 10, 83-98
- Gergen, K. J., & Gergen, M. M. (Eds) (1981). *Social Psychology*, Harcourt Brace Janovich, Inc.
- Krosnik, J. A. (1999). Survey Research. *Annual Review of Psychology*, 50, 537-567.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 140, 5-53
- Osgood, C. E., Suci, G. J., & Tanenbaum, Ph. D. (1957). *The Measurement of Meaning*. Urbana: University of Illinois Press.
- Rugg, D. (1941). Experiments in wording questions: II. *Public Opinion Quarterly*, 5, 91-92.
- Schwartz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist* 54, 93-105.