Sense and Nonsense about Surveys

Howard Schuman

Contexts 2002 1: 40
DOI: 10.1525/ctx.2002.1.2.40

sense and nonsense about surveys

Understanding surveys is critical to being an informed citizen, but popular media often report surveys without any guidance on how to interpret and evaluate the results. Some basic guidelines can promote more sophisticated readings of survey results and help teach when to trust the polls.

Surveys draw on two human propensities that have served us well from ancient times. One is to gather information by asking questions. The first use of language around 100,000 years ago may have been to utter commands such as "Come here!" or "Wait!" Questions must have followed soon after: "Why?" or "What for?" From that point, it would have been only a short step to the use of interrogatives to learn where a fellow hominid had seen potential food, a dangerous animal, or something else of importance. Asking questions continues to be an effective way of acquiring information of all kinds, assuming of course that the person answering is able and willing to respond accurately.

The other inclination, learning about one's environment by examining a small part of it, is the sampling aspect of surveys. A taste of something may or may not point to appetizing food. A first inquiry to a stranger, a first glance around a room, a first date—each is a sample of sorts, often used to decide whether it is wise to proceed further. As with questions, however, one must always be aware of the possibility that the sample may not prove adequate to the task.

sampling: how gallup achieved fame

Only within the past century—and especially in the 1930s and 1940s—were major improvements made in the sampling process that allowed the modern survey to develop and flourish. A crucial change involved recognition that the value of a sample comes not simply from its size but also from the way it is obtained. Every serious pursuit likes to have a morality tale that supports its basic beliefs: witness Eve and the apple in the Bible or Newton and his apple in legends about scientific discovery. Representative sampling has a marvelous morality tale also, with the additional advantage of its being true.

The story concerns the infamous *Literary Digest* poll prediction—based on 10 million questionnaires sent out and more than two million received back—that Roosevelt would lose decisively in the 1936 presidential election. At the same time, George Gallup, using many fewer cases but a much better method, made the more accurate prediction that FDR would win. Gallup used quotas in choosing respondents in order to represent different economic strata, whereas the *Literary Digest* had worked mainly from telephone and automobile ownership lists, which in 1936 were biased toward wealthy people apt to be opposed to Roosevelt. (There were other sources of bias as well.) As a result, the *Literary Digest* poll disappeared from the scene, and Gallup was on his way to becoming a household name.

> The percentage of people who refuse to take part in a survey is particularly important. In some federal surveys, the percentage is small, within the range of 5 to 10 percent. For even the best non-government surveys, the refusal rate can reach 25 percent or more, and it can be far larger in the case of poorly executed surveys.

Yet despite their intuitive grasp of the importance of representing the electorate accurately, Gallup and other commercial pollsters did not use the probability sampling methods that were being developed in the same decades and that are fundamental to social science surveys today. Probability sampling in its simplest form calls for each person in the population to have an equal chance of being selected. It can also be used in more complex applications where the chances are deliberately made to be unequal, for example, when oversampling a minority group in order to study it more closely; however, the chances of being selected must still be known so that they can later be equalized when considering the entire population.

intuitions and counterintuitions about sample size

Probability sampling theory reveals a crucial but counterintuitive point about sample size: the size of a sample needed to accurately estimate a value for a population depends very little on the size of the population. For example, almost the same size sample is needed to estimate, with a given degree of precision, the proportion of left-handed people in the United States as is needed to make the same estimate for, say, Peoria, Illinois. In both cases a reasonably accurate estimate can be obtained with a sample size of around 1,000. (More cases are needed when extraordinary precision is called for, for example, in calculating unemployment rates, where even a tenth of a percent change may be regarded as important.)

The link between population size and sample size cuts both ways. Although huge samples are not needed for huge populations like those of the United States or China, a handful of cases is not sufficient simply because one's interest is limited to Peoria. This implication is often missed by those trying to save time and money when sampling a small community.

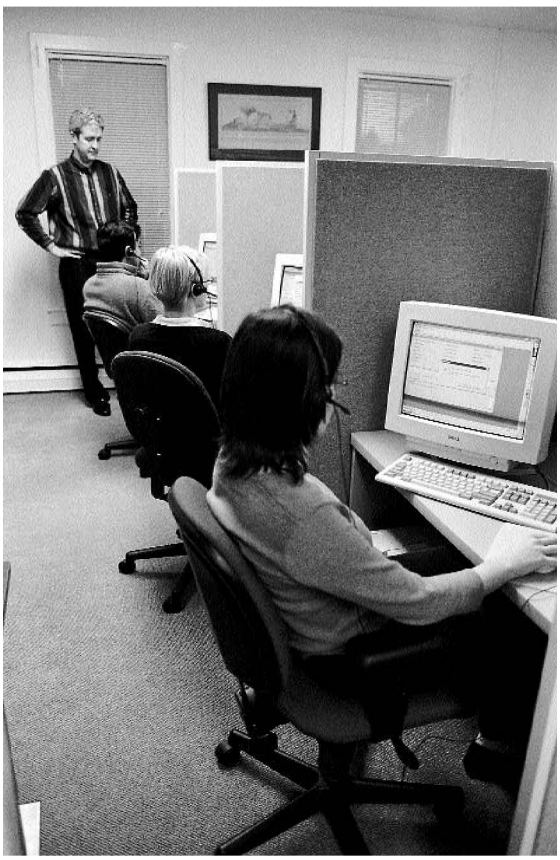


Photo courtesy of Denise Applewhite, Princeton University

Telephone interviewers and survey research supervisor.

Moreover, all of these statements depend on restricting your interest to overall population values. If you are concerned about, say, left-handedness among African Americans, then African Americans become your population, and you need much the same sample size as for Peoria or the United States.

who is missing?

A good sample depends on more than probability sampling theory. Surveys vary greatly in their quality of implementation, and this variation is not captured by the "margin of error" plus/minus percentage figures that accompany most media reports of polls. Such percentages reflect the size of the final sample, but they do not reveal the sampling method or the extent to which the targeted individuals or households were actually included in the final sample. These details are at least as important as the sample size.

When targeted members of a population are not interviewed or do not respond to particular questions, the omissions are a serious problem if they are numerous and if those missed differ from those who are interviewed on the matters being studied. The latter difference can seldom be known with great confidence, so it is usually desirable to keep omissions to a minimum. For example, sampling from telephone directories is undesirable because it leaves out those with unlisted telephones, as well as those with no telephones at all. Many survey reports are based on such poor sampling procedures that they may not deserve to be taken seriously. This is especially true of reports based on "focus groups," which offer lots of human interest but are subject to vast amounts of error. Internet surveys also cannot represent the general population adequately at present, though this is an area where some serious attempts are being made to compensate for the inherent difficulties.

The percentage of people who refuse to take part in a survey is particularly important. In some federal surveys, the percentage is small, within the range of 5 to 10 percent. For even the best non-government surveys, the refusal rate can reach 25 percent or more, and it can be far larger in the case of poorly executed surveys. Refusals have risen substantially from earlier days, becoming a major cause for concern among serious survey practitioners. Fortunately, in recent years research has shown that moderate amounts of nonresponse in an otherwise careful survey seem in most cases not to have a major effect on results. Indeed, even the *Literary Digest,* with its abysmal sampling and massive nonresponse rate, did well predicting elections before the dramatic realignment of the electorate in 1936. The problem is that one can never be certain as to the effects of refusals and other forms of nonresponse, so obtaining a high response rate remains an important goal.

Calling Spirits from the Vasty Deep

Two characters in Shakespeare's *Henry IV* illustrate a pressing problem facing surveys today:

Glendower: I can call spirits from the vasty deep.
Hotspur: Why, so can I, or so can any man; But will they come when you do call for them?

New impediments such as answering machines make contacting people more difficult, and annoyance with telemarketing and other intrusions discourages people from becoming respondents. The major academic survey organizations invest significant resources in repeatedly calling people and also in trying to persuade people to be interviewed. Thus far response rates for leading surveys have suffered only a little, but other organizations more limited by time and costs have seen rates plummet.

Fortunately, research about the effect of nonresponse on findings has increased. Two recent articles in *Public Opinion Quarterly* report surprisingly small differences in results from surveys with substantial differences in response rates. One study focuses on the University of Michigan's Survey of Consumers and finds that the number of calls required to complete a single interview doubled from 1979 to 1996. However, controlling for major social background characteristics, the authors also report that stopping calls earlier and making fewer attempts to convert refusals would have had little effect on a key measure, the Index of Consumer Sentiments. In a second study researchers conducted two basically similar surveys: one accepted a 36 percent response rate to conserve time and money; the other invested additional time and resources to obtain a 61 percent response rate. On a wide range of attitude items, the researchers found few noteworthy differences in outcomes due to the large difference in response rates.

It is important to keep in mind that bias due to nonresponse will occur only if non-respondents differ from respondents on the measures of interest and in ways that cannot be controlled statistically. Thus, while high response rates are always desirable in principle, the actual effects of nonresponse call for careful empirical research, not dogmatic pronouncements.

questions about questions

Since survey questions resemble the questions we ask in ordinary social interaction, they may seem less problematic than the counterintuitive and technical aspects of sampling. Yet survey results are every bit as dependent on the form, wording and context of the questions asked as they are on the sample of people who answer them.

No classic morality tale like the *Literary Digest* fiasco highlights the question-answer process, but an example from the early days of surveys illustrates both the potential challenges of question writing and the practical solutions.

In 1940 Donald Rugg asked two slightly different questions to equivalent national samples about the general issue of freedom of speech:

- Do you think the United States should forbid public speeches against democracy?
- Do you think the United States should allow public speeches against democracy?

Photo courtesy of Stony Brook Center for Survey Research

Telephone survey interviewer using headset and computer for data entry.

Taken literally, forbidding something and not allowing something have the same effect, but clearly the public did not view the questions as identical. Whereas 75 percent of the public would not allow such speeches, only 54 percent would

forbid them, a difference of 21 percentage points. This finding was replicated several times in later years, not only in the United States but also (with appropriate translations) in Germany and the Netherlands. Such "survey-based experiments" call for administering different versions of a question to random subsamples of a larger sample. If the results between the subsamples differ by more than can be easily explained by chance, we infer that the difference is due to the variation in wording.

In addition, answers to survey questions always depend on the form in which a question is asked. If the interviewer presents a limited set of alternatives, most respondents will choose one, rather than offering a different alternative of their own. In one survey-based experiment, for example, we asked a national sample of Americans to name the most important problem facing the country. Then we asked a comparable sample a parallel question that provided a list of four problems from which to choose the most important; this list included none of the four problems mentioned most often by the first sample but instead provided four problems that had been mentioned by fewer than 3 percent of the earlier respondents. The list question also invited respondents to substitute a different problem if they wished (see Table 1). Despite the invitation, the majority of respondents (60 percent) chose one of the rare problems offered, reflecting their reluctance to go outside the frame of reference provided by the question. The form of a question provides the "rules of the game" for respondents, and this must always be kept in mind when interpreting results.

Photo by Sheryl Sinkow, Gerontology Institute, Ithaca College

Simulated interview for survey workers in training.

Other difficulties occur with survey questions when issues are discussed quite generally, as though there is a single way of framing them and just two sides to the debate. For example, what is called "the abortion issue" really consists of different issues: the reasons for an abortion, the trimester involved and so forth. In a recent General Social Survey, nearly 80 percent of the national sample supported legal abortion in the case of "a serious defect

table 1
Experimental Variation Between Open and Closed Questions

A. Open Question	B. Closed Question
"What do you think is the most important problem facing this country today [1986]?"	"Which of the following do you think is the most important problem facing this country today [1986] – the energy shortage, the quality of public schools, legalized abortion, or pollution – or, if you prefer, you may name a different problem as most important." 1. Energy shortage. 2. Quality of public schools. 3. Legalized abortion. 4. Pollution.

Adapted from: H. Schuman and J. Scott, "Problems in the Use of Survey Questions to Measure Public Opinion," *Science* v. 236, pp. 957-959, May 22, 1987.

In a survey experiment, less than 3% of the 171 respondents asked the question on the left volunteered one of the four problems listed on the right. Yet, 60% of the 178 respondents asked the question on the right picked one of those four answers.

in the baby," but only 44 percent supported it "if the family has a low income and cannot afford any more children." Often what is thought to be a conflict in findings between two surveys is actually a difference in the aspects of the general issue that they queried. In still other cases an inconsistency reflects a type of illogical wish fulfillment in the public itself, as when majorities favor both a decrease in taxes and an increase in government services if the questions are asked separately.

solutions to the question wording problem

All these and still other difficulties (including the order in which questions are asked) suggest that responses to single survey questions on complex issues should be viewed with considerable skepticism. What to do then, other than to reject all survey data as unusable for serious purposes? One answer can be found from the replications of the forbid/allow experiment above: Although there was a 21 percentage points difference based on question wording in 1940 and a slightly larger difference (24 percentage points) when the experiment was repeated some 35 years later, both the forbid and the allow wordings registered similar declines in Americans' intolerance of speeches against democracy (see Figure 1). No matter which question was used—as long as it was the same one at both times—the conclusion about the increase in civil libertarian sentiments was the same.

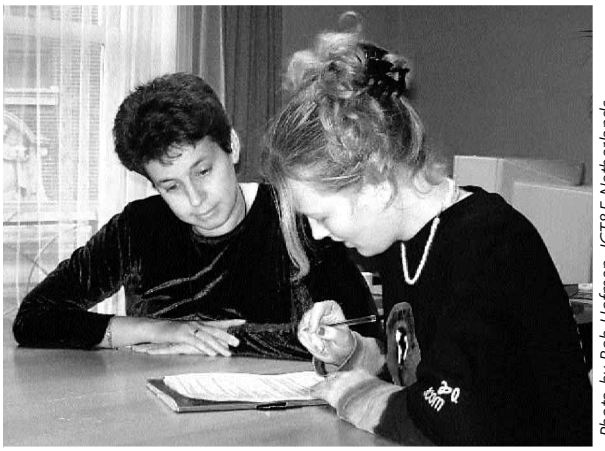
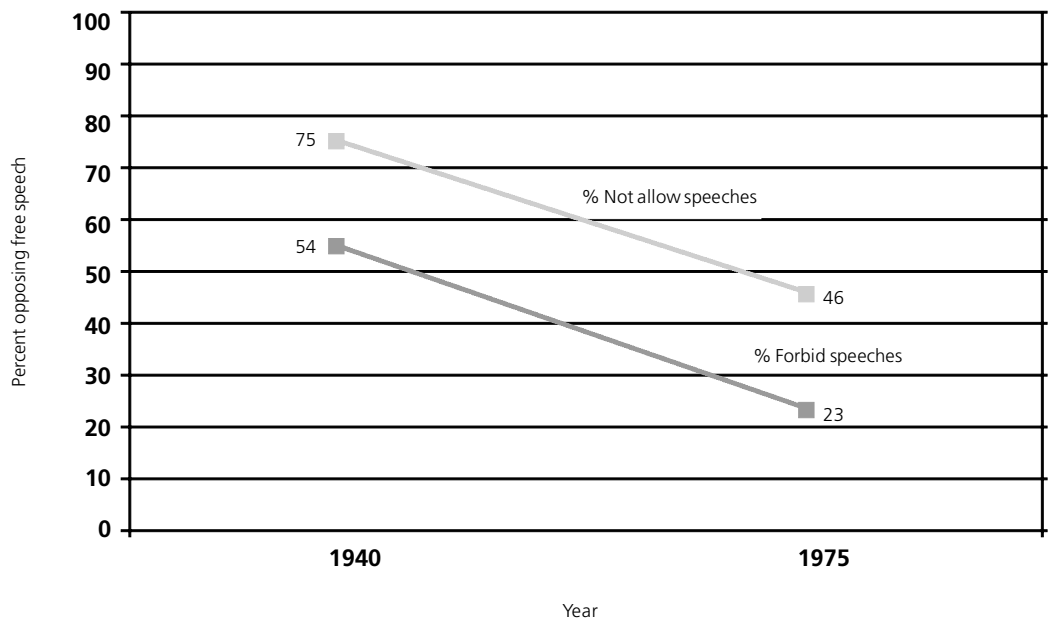


Photo by Bob Hofman, ICT&E Netherlands

Interview on the digital divide. In a departure from conventional protocol, interviewer (on right) is sitting next to rather than across from interviewee.

More generally, what has been called the "principle of form-resistant correlations" holds in most cases: if question wording (and meaning) is kept constant, differences over time, differences across educational levels, and most other careful comparisons are not seriously affected by specific question wording. Indeed, the distinction between results for single questions and results based on comparisons or associations holds even for simple factual inquiries. Consider, for example, a study of the number of rooms in American houses. No God-

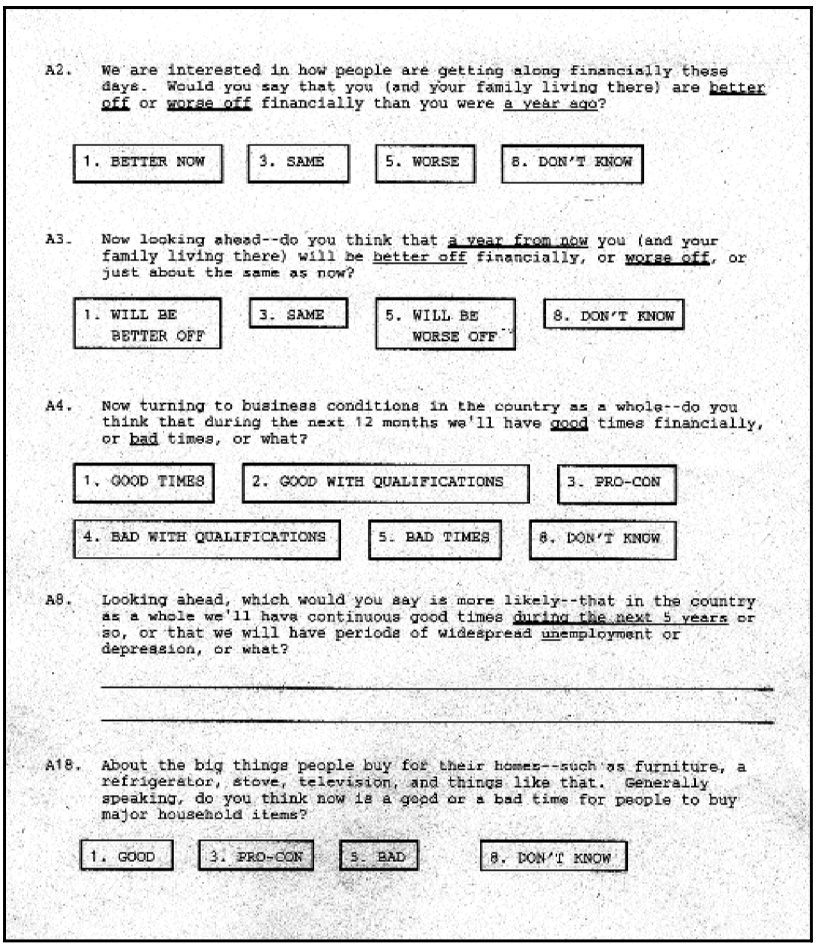
figure 1
Attitudes Toward Free Speech Against Democracy



given rule states what to include when counting the rooms in a house (bathrooms? basements? hallways?); hence the average number reported for a particular place and time should not be treated as an absolute truth. What we can do, however, is try to apply the same definitions over time, across social divisions, even across nations. That way, we gain confidence in the comparisons we make—who has more rooms than who, for example.

We still face the task of interpreting the meaning of questions and of associations among questions, but that is true in all types of research. Even an index constructed from a large number of questions on the basis of a sophisticated statistical calculation called factor analysis inevitably requires the investigator to interpret what it is that he or she has measured. There is no escaping this theoretical challenge, fundamental to all research, whether using surveys or other methods such as field observations.

Survey researchers should also ask several different questions about any important issue. In addition to combining questions to increase reliability, the different answers can be synthesized rather than depending on the angle of vision provided by any single question. A further safeguard is to carry out frequent experiments like that on the forbid/allow wordings. By varying the form, wording, and context of questions, researchers can gain insight into both the questions and the relevant issues. Sometimes variations turn out to make no difference, and that is also useful to learn. For example, I once expected support for legalized abortion to increase when a question substituted *end pregnancy* for the word *abortion* in the phrasing. Yet no difference was found. Today, more and more researchers include survey-based experiments as part of their investigations, and readers should look for these sorts of safeguards when evaluating survey results.



A2. We are interested in how people are getting along financially these days. Would you say that you (and your family living there) are better off or worse off financially than you were a year ago?

1. BETTER NOW | 3. SAME | 5. WORSE | 8. DON'T KNOW

A3. Now looking ahead--do you think that a year from now you (and your family living there) will be better off financially, or worse off, or just about the same as now?

1. WILL BE BETTER OFF | 3. SAME | 5. WILL BE WORSE OFF | 8. DON'T KNOW

A4. Now turning to business conditions in the country as a whole--do you think that during the next 12 months we'll have good times financially, or bad times, or what?

1. GOOD TIMES | 2. GOOD WITH QUALIFICATIONS | 3. PRO-CON

4. BAD WITH QUALIFICATIONS | 5. BAD TIMES | 8. DON'T KNOW

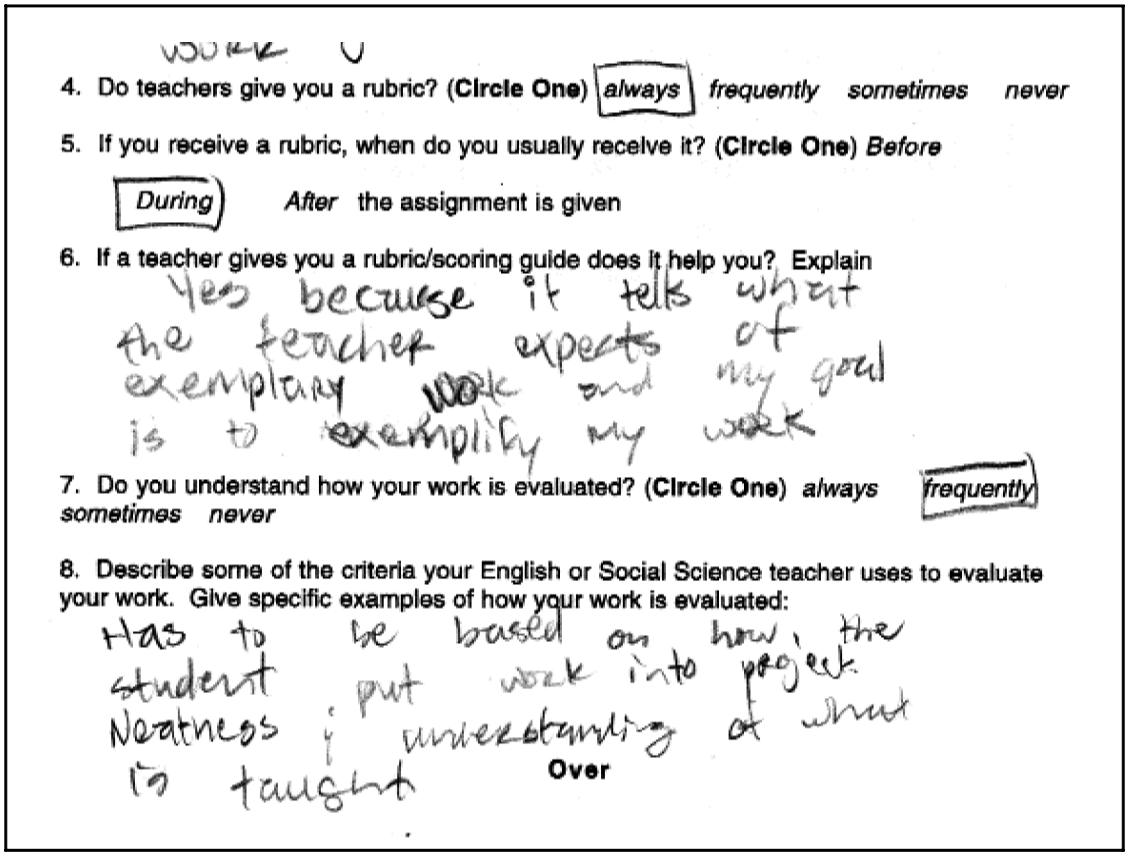
A8. Looking ahead, which would you say is more likely--that in the country as a whole we'll have continuous good times during the next 5 years or so, or that we will have periods of widespread unemployment or depression, or what?

A18. About the big things people buy for their homes--such as furniture, a refrigerator, stove, television, and things like that. Generally speaking, do you think now is a good or a bad time for people to buy major household items?

1. GOOD | 3. PRO-CON | 5. BAD | 8. DON'T KNOW

Courtesy of Survey Research Center, University of Michigan.

Section of interview form used in the Surveys of Consumers conducted by the Survey Research Center, University of Michigan.



4. Do teachers give you a rubric? (Circle One) *always* *frequently* *sometimes* *never*

5. If you receive a rubric, when do you usually receive it? (Circle One) *Before*

During *After* the assignment is given

6. If a teacher gives you a rubric/scoring guide does it help you? Explain

Yes because it tells what the teacher expects of exemplary work and my goal is to exemplify my work

7. Do you understand how your work is evaluated? (Circle One) *always* *frequently* *sometimes* *never*

8. Describe some of the criteria your English or Social Science teacher uses to evaluate your work. Give specific examples of how your work is evaluated:

Has to be based on how the student put work into project. Neatness & understanding of what is taught

Over

Page from completed, self-administered questionnaire used to study high school students' views of grading.

the need for comparisons

To interpret surveys accurately, it's important to use a framework of comparative data in evaluating the results. For example, teachers know that course evaluations can be interpreted best against the backdrop of evaluations from other similar courses: a 75 percent rating of lectures as "excellent" takes on a quite different meaning depending on whether the average for other lecture courses is 50 percent or 90 percent. Such comparisons are fundamental for all survey results, yet they are easily overlooked when one feels the urge to speak definitively about public reactions to a unique event.

Comparative analysis over time, along with survey-based experiments, can also help us understand responses to questions about socially sensitive subjects. Experiments have shown that expressions of racial attitudes can change substantially for both black and white Americans depending on the interviewer's race. White respondents, for instance, are more likely to support racial intermarriage when speaking to a black than to a white interviewer. Such self-censoring mirrors variations in cross-race conversations outside of surveys, reflecting not a methodological artifact of surveys but rather a fact of life about race relations in America. Still, if we consider time trends, with the race of interviewer kept constant, we can also see that white responses supporting intermarriage have clearly increased over the past half century (see Table 2), that actual intermarriage rates have also risen (though from a much lower level) over recent years, and that the public visibility of cross-race marriage and dating has also increased. It would be foolish to assume that the survey data on racial attitudes reflect actions in any literal sense, but they do capture important *trends* in both norms and behavior.

Surveys remain our best tool for learning about large populations. One remarkable advantage surveys have over some other methods is the ability to identify their own limitations, as illustrated by the development of both probability theory in

table 2

Percent of White Americans Approving or Disapproving of Racial Intermarriage, 1958-1997

"Do you approve or disapprove of marriage between blacks and whites?"

Year	Approve	Disapprove
1958	4	96
1978	34	66
1997	67	33

Source: Gallup Poll

sampling and experiments in questioning. In the end, however, with surveys as with all research methods, there is no substitute for both care and intelligence in the way evidence is gathered and interpreted. What we learn about society is always mediated by the instruments we use, including our own eyes and ears. As Isaac Newton wrote long ago, error is not in the art but in the artificers. ■

recommended resources

Converse, Philip E. "The Nature of Belief Systems in Mass Publics." In *Ideology and Discontent,* ed. D. E. Apter. New York: The Free Press, 1964. A profound and skeptical exploration of the nature of public attitudes.

Groves, Robert M. *Survey Errors and Survey Costs.* New York: Wiley, 1989. A sophisticated consideration of the sources of error in surveys.

Kalton, Graham. *Introduction to Survey Sampling.* Thousand Oaks, Calif.: Sage Publications (Quantitative Applications in the Social Sciences), 1983. A brief and lucid introduction to sampling.

Page, Benjamin I., and Robert Y. Shapiro. *The Rational Public: Fifty Years of Trends in Americans' Policy Preferences.* Chicago: University of Chicago Press, 1992. In part, a persuasive reply to Converse's skepticism.

Schuman, Howard, and Stanley Presser. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context.* San Diego, Calif.: Academic Press, 1981 (Reprint edition with new preface, Thousand Oaks, Calif.: Sage Publications, 1996). Several experiments discussed in the present article are drawn from this volume.

Stouffer, Samuel A. *Communism, Conformity, and Civil Liberties,* with introduction by James A. Davis. New York: Doubleday, 1955; New Brunswick, N.J.: Transaction Publishers, 1992. Stouffer's keen awareness of both the possibilities and the limitations of survey data is reflected in this classic investigation. Also relevant to today's political climate.

Sudman, Seymour, Norman M. Bradburn, and Norbert Schwarz. *Thinking About Answers: The Application of Cognitive Process to Survey Methodology.* San Francisco: Jossey-Bass, 1996. A clear discussion of survey questioning by three well-known researchers.

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. *The Psychology of Survey Response.* Cambridge: Cambridge University Press, 2000. A comprehensive account of response effects, drawing especially on ideas from cognitive psychology.