

In my opinion...

Good intentions don't count

by Alan J. Parker and
Marcel Escoffier

While simple guest surveys can be poorly constructed with little negative consequence, often surveys are used in making important policy decisions. Researchers and policy makers must carefully construct their research instruments in order to avoid biases which may result in muddled or incorrect responses. The authors review the process of creating, administering, and analyzing surveys with an eye toward reducing survey bias to a minimum. Reliable results require a rigorous and careful approach when creating and using instruments.

With the rise of the Internet, information is more available, more widely disseminated, and, potentially, more unreliable than ever before. The idea of information being more readily available is one of the greatest contributions of the Internet, especially for students and faculty in academic settings. A problem arises because the information is not evaluated by some reputable or knowledgeable source prior to its posting to the web. While academic journals have their peer

review, and even popular journals have some editor review articles before they go to press, the net allows just about anyone to post just about anything. Students are particularly susceptible to accepting as truth most information obtained through the Internet unless it clearly comes from some obviously ridiculous source (e.g., www.lies.com). Even then, if the information is relevant to their needs (i.e., term paper due in the morning), they may use it.

In the hospitality industry surveys and opinion polls are frequently produced and widely disseminated via the Internet. Unfortunately, many studies are flawed, some irreparably so. Even though a researcher may be "pure of heart" and driven by the highest motives, unless a study conforms to the basic tenants of good research statistics, it may be useless.

As a case in point, a study was put on the HOTEL-ONLINE site as a special report. Obvious errors

were committed by the investigators, but, perhaps most maddening, the report's conclusions are of great importance to the hospitality industry. The results showed that 90 percent of those in the United Kingdom (UK), 70 percent of Australians, and nearly a third of all Americans believe that tourism development is in danger of destroying the environment.¹ If the report is true, such an important finding would have important ramifications in the industry. But in order for such a conclusion to be valid, certain steps must be taken to ensure the reliability of any study. These issues will be discussed.

Methodology is an issue

At issue in the methodology is if the survey instrument (the questionnaire) makes sense as to its content (construct validity, understanding by those who are asked to fill it out, etc.), length, and internal consistency (i.e., do the questions asked seem to relate to one another and are the responses similar to similar questions?) Perhaps more importantly, a prime source of research bias is the incorrect choice of a sample of subjects.

One obvious problem with any study is that those asking the questions (the interviewers, or the ones who wrote the survey) usually have much more knowledge of the subject matter than do those answering the questions. A group interested in pandas and their preservation probably know a wealth of facts concerning panda

physiology, range of habitat, eating preferences, etc., that would be well beyond the level of even a moderately interested member of the general population. It becomes too easy to ask respondents very detailed questions which they know little or nothing about. Most people tend to be helpful, friendly, or interested, so they answer these types of questions as best they can (i.e., trying to satisfy perceived expectations of the person who is administering the survey).

Just this past month, one author experienced such an event in dramatic detail. A major university is building married student dorms in the community. University surveys showed no strong community opposition to the project. The questionnaire was clearly written, short, to the point, and personally administered by a university official who knocked on every door in the neighborhood. Aroused neighbors formed a neighborhood group which conducted its own survey of all neighbors in the immediate area which found very strong negative feelings toward the proposed project. The same sample group is responding strongly both positively and negatively to the same issue. The only difference was the group administering the survey and how they phrased their questions.

It seems apparent that if researchers are to produce surveys that fairly reflect the attitudes of those surveyed, they must guard against such biases lest

they cloud the results to the extent that the survey becomes useless.

Broad biases exist

Survey questionnaires are subject to numerous biases which can be broadly classified as concerning the creation, the administration, and the evaluation of the instrument. Biases at the point of creation may be further broken down into those concerning the constructs to be measured, those concerning the selection of methodology, and those which arise due to poor or unprofessional instrument layout. When the time comes to actually administer the survey, choices made concerning when and where it will be administered, how it will be administered, and who will administer it all contain a potential for biases to significantly affect survey results. Finally, once the surveys have all been returned, there are biases which may creep into the analysis phase, such as choice of statistics, comparison to other surveys, and inferences made concerning the results of the survey. While no instrument can be completely free of bias, a survey can be made as bias-free as possible and residual biases can be accounted for in the final analysis stage.

Biases identified early

Generally, the process begins with the initial idea that there is something of interest to be studied. Scientists refer to this idea as a hypothesis, but in most surveys

done in the hospitality industry it might be closer to the truth if the term "idea" is used. Numerous disciplines have studied theories relevant to the hospitality industry. Management, psychology, sociology, anthropology, and many other fields have literally dozens or hundreds of theories which may be used to explain the behaviors seen in guests, employees, investors, and, in fact, everyone associated with the industry.

A researcher about to conduct a survey probably will have trained in one or more of these fields of research. Each approaches any human behavior from a different perspective. This inherent research bias is probably not a bad thing, but it exists nonetheless and can become a problem when a survey is done by someone in a field where the issue being studied is tangential to his or her field of expertise, for example, say a sociologist studying the choice of retirement investments, an issue that may be more relevant to the field of economics or finance.

Furthermore, construct biases occur when the researcher is a strong advocate of one position or another. Certainly a survey of attitudes concerning swimming as a recreation could be constructed in a very different way by a researcher who is a non-swimmer and who is afraid of the water than if the survey had been created by someone who enjoys swimming, diving, and other water sports.

It's convenient and relatively cheap for college professors to

survey college students. But one wonders if the convenience sample is appropriate for some studies. Similarly, the choice of a mail survey versus interviews, or a telephone survey versus the Internet, may impart a certain bias to any survey.

Sample is important

A decision must then be made as to the number of those to be surveyed. Since the hospitality industry is so large, it might be argued that meaningful surveys must adequately sample a very broad spectrum of the population. Even a survey done for one airline or one hotel company may require a sample size of a thousand or more people to be statistically significant. It is at this point of creating the instrument that a decision must be made concerning the types of measures to be used. By far most industry surveys use a Likert-like scale of one sort or another. Research into this form of question has shown that it works best when there are five or seven choices, with "neutral" not being one of them. The center choice should say, "neither ___ nor ___" with the blanks filled in with the terms used at either end of the response field. Alternative response choices include faces (smile for approval, frown for disapproval) or other questioning methods. Obviously, how the question gets written greatly effects the responses. One could imagine how the personality tests of the 1930s that asked, "I like to go to gay

parties," might engender a very different response today.

Finally, the actual construction or layout of the instrument may contain hidden biases. A typed, poorly organized, and poorly laid out instrument can tell those being surveyed that this is not very important and that they shouldn't think too much about it. Along the same lines, an interview survey done while people are rushing to take a train or to get a taxi might have respondents spending a few hurried moments to complete. It is this last bias that caused lawmakers in most states to enact legislation which limits the number of words that can be used for a ballot question. The theory is that voters are rushed to cast their votes and make way for the next voter. In their rush they may not read and understand fully what they are voting on if presented with a verbose proposal full of legalese. This principle of getting to the point and asking the respondent a direct question can be good advice to anyone wishing to conduct a survey.

Finally, research shows that the choice of administration method and the construction of the instrument itself are not interchangeable. That is, a written survey instrument cannot just be scanned into a computer and used as is as an Internet survey. It is advisable for professional surveys to be laid out by a graphic artist so as to ensure a professional looking instrument. A professional survey helps assure careful, thoughtful

responses. On mailed surveys, the response rate has been shown to be higher when the instrument looked really professional.

Bias can be reduced

A consideration which needs to be made is how to go about assuring that the survey accurately reports the beliefs and attitudes of a representative sample of the population of interest. Convenience samples are probably the norm. But, statistically, a good survey should draw its sample of respondents randomly from the population in question. For example, if the purpose of a survey is to address the special needs of seniors when they travel, then seniors are the population of interest. Once the term "senior" is defined, a cursory analysis of relevant demographic data would indicate that there are more females than males. Further data analysis might also indicate that seniors who are female travel more often than those who are male. If this were true, then the survey sample should reflect these demographic realities.

For surveys of the general population, a sample must be drawn that more or less reflects the general population. Statisticians agree that for a sample to be valid anyone in the population of interest should have an equal chance of being selected to participate in the survey. This may place too great a burden on most surveys, but some attempt should be made to ensure that at least a quasi-random sample has been selected.

Administering a survey to a group of people on one flight, say to Australia, would certainly produce questionable results.

Determining the other size and composition is tied into the question of how, when, and where to administer the survey. A survey to determine airline seat preferences logically could be done at an airport. The results might vary if respondents were about to leave on a flight versus those just arriving, especially from a very long flight. Obviously, such a survey would probably be less reliable were it to be administered to people traveling on Amtrack, or to those who had just driven to Disneyland. It could be argued that a survey of hotel guest attitudes might logically be administered to those staying at a hotel. If such a survey were to be administered at an airport, one might wish to pre-qualify the respondent to be sure that he or she actually plans to stay at a hotel, rather than flying to visit family or some other destination.

In order to capture as broad a range of respondents as possible, survey administrators often use telephone surveys, mass mailings, and the Internet. Each method has its strengths and weaknesses. Telephone surveys are probably statistically viable now; they were not some years back before universal phone service. They would still be suspect in many third world situations. Mass mailings work well, but the mailing list can be subject to bias. A mailed survey of attitudes toward the environment sent to members of

the Sierra Club might engender a different response than one sent to the members of the National Rifle Association. Because of the industry's high turnover, hospitality surveys often go out to hotel or restaurant managers and are undeliverable. A third approach is the use of the Internet. Two biases come to mind. First, Internet use is very closely associated with a person's age and socio-economic status. Second, some Internet users delight in sending in multiple responses to a survey.

What works?

There is no method that is unbiased, but the direct interview is very common. Provided certain conditions are met, the interview method produces the least biased results. Obviously, the interviewer must be trained so as not to bias the response. Professor Escoffier participated in an interview where the interviewer let out a giant sigh when the response was not as anticipated. Such oral and visual cues can greatly influence the results, as any marketing textbook can detail when explaining how a focus group works. But, given appropriate training, a skilled interviewer can avoid these gross biases and operate as a good bias eliminator. If he or she sees someone is hurried, exasperated, fatigued, or just plain irritable, the interviewer can thank the respondent and move on to someone else. He or she can detect other biases such as not understanding the language, puzzling over termi-

nology, collusion, and the like. The use of an interviewer also provides more uniform and complete responses than when subjects are left to their own devices.

Finally, administering a good survey includes the ability to assure that one respondent does not see the completed surveys of other respondents. Mailed, telephone, and direct interview methods can assure this; mass mailings where no attempt is made to avoid duplicate addresses and Internet surveys do not.

Results can show bias

Assuming the investigator has produced a relatively error-free instrument and has administered it so as to reduce biases inherent in that phase of the investigation, the task now begins of analyzing the results. Perhaps the first step in this procedure should be to analyze the reliability of the results. Comparing similar questions for similar responses (for example, asking a respondent how many times he or she has been on a cruise and then, later, asking how often he or she cruises and when was the first cruise) can help detect respondents who fabricated their responses.

Comparing people's responses to expected norms is another reliability verification technique (for instance, first class airline passengers may respond more favorably to a question concerning the quality of the food offered on a flight than would those flying

coach). Responses that approach the limits (almost all saying "no" to a question) may indicate either very strongly held beliefs (which should be observed elsewhere in the survey) or a question that was poorly understood by respondents at large.

Finally, one should not ignore the issue of external reliability. If the survey results are at significant variance with similar surveys conducted in the past, the investigator is well advised to look for possible weaknesses with his own procedures before proclaiming to the world that he has proven the other research wrong. The National Restaurant Association has, for many years, surveyed restaurant patrons concerning what factors they consider in making a decision to dine at a restaurant. The results are always similar; food and service top the list, with price, location, and ambiance being other significant factors. A researcher whose survey said that some sixth factor was most important, or whose survey otherwise significantly contradicted the NRA surveys, should realize that something went horribly wrong and that the survey needs to be re-considered and re-conducted. Assuming no significant reliability issues arise, the investigator can go on to conduct further analysis.

Pilot survey necessary

At this point, the wise researcher would have done everything mentioned so far in a

pilot survey of a few actual subjects. Assuming the results indicate that those few respondents were able to successfully complete the survey, and assuming that "eyeballing" the responses shows no glaring abnormalities, the researcher would then undertake a large survey. It should be stressed that many books err when describing the survey method in implying that a pilot study is conducted without considering the issues addressed so far. This is untrue. The pilot study needs to mirror as many conditions of the real survey as possible. If the pilot surveys seem to reflect responses that the researcher had anticipated, and assuming there is an indication that the instrument is reliable, the researcher can safely go on to invest the time and money needed to conduct a large scale survey.

Once the survey has been completed, and assuming that the reliability tests on the actual study closely resemble the results of the pilot survey, the investigator may begin the analysis phase. Choice of statistical methods to use are an on-going concern. Obviously, if simple descriptive statistics are all one wishes, the survey can be constructed without much effort.

But good surveys compare responses using more sophisticated statistical tests. At the least, the researcher should analyze outliers (responses that were significantly outside the

norm.) While many such cases are statistical anomalies, some offer profound insight. Studies sometimes show that the outliers are the subjects with special needs or prejudices, or are the start of major trends. A survey of computer users in 1976 would have shown that those owning a personal computer were outliers from the norm. Yet close examination of their reasons for pioneering the personal computer revolution helped companies like IBM and others get a head start on the others. Those investigating emerging trends might investigate correlations among variables. A correlation between two or more variables may be irrelevant; there is a high correlation between gum chewing and crime rate in large cities, but it can often indicate that issues need further study. One common fallacy is to assume that correlation assumes causality. While there is a high correlation between chickens and eggs, the age old question of which came first cannot be answered by a correlation statistic.

Other analysis possible

A more sophisticated analysis is in the use of various cross-tabulation techniques (the chi-square statistic, for example.) Frequently responses showing strong leanings on one question also show strong leanings on another question in a survey. Demographic responses such as gender or age often relate highly

to other survey responses. The chi-square statistic frequently can show if such a relationship is statistically significant.

While surveys are often more exploratory in nature, and, hence, do not attempt to verify a hypothesis, they can be used in hypothesis creation and testing. In these cases, the instrument is frequently constructed so as to allow the use of the Analysis of Variance (ANOVA) or Analysis of Co-Variance (ANCOVA) statistics. Quasi-experimental design is outside the purview of this article, but initial survey results often lead researchers in the direction of more scientific experimentation.

A last caveat concerning the use of statistics is to be very careful when doing statistical tests. Most programs set the level for statistical significance at 95 percent (usually entered as .05). This means on a 20-question survey, one question will be significant by chance alone (one out of 20 questions will erroneously be reported as significant at the 95 percent level). This is known as a Type II error. A Type I error is when a research hypothesis which is true is rejected. Type II error is accepting a research hypothesis that is untrue. One wag once reported the possibility of a Type III error, solving the "wrong" or sub-optimal problem.

A survey utterly free of any research bias is probably an impossible goal. Like service perfection, it can never be achieved, but

constantly striving toward that goal is what every researcher should be doing. Surveys probably run the gamut of incredibly well executed to so poorly constructed as to be an embarrassment to all concerned.

Biases can converge

A recent study reported in *Hotel Online* illustrates how many of these biases may converge to create a survey whose usefulness for decision making may be questioned.³

The survey was conducted by an interest group which wished to convince hoteliers and the public at large that a stronger commitment to ecological issues is both a good thing to do and good for business. Accordingly, the group constructed a survey which they conducted at airports in three countries, Australia, the United Kingdom, and the United States. The reported results were very mixed. Respondents had trouble being consistent in their responses; there were variations between countries, and there was a problem within the combined results.

Researchers made several questionable assumptions when creating the instrument. First, they assumed that respondents would be relatively knowledgeable concerning environmental issues. National differences in the veracity of this assumption are clearly evident in the results. Next, the researchers assumed that questions phrased in good

English grammar in the United Kingdom were as understandable and meaningful in the United States and Australia. This assumption, too, is suspect. Finally, the researchers assumed that their passion for the topic would be shared by respondents so they created an instrument that was quite long and which demonstrated "response fatigue."

The instrument was administered in passport lounges at Heathrow airport in London, at LaGuardia in New York, and at an airport in Australia. While one may assume that passengers awaiting international flights are demographically similar, this assumption may be flawed. Worse, there is every indication that people awaiting international flights may very well not reflect the travel public norm. Finally, given the tens of millions of passengers flying in a given year, the selection of 300 subjects probably reflects too small a sample to even represent the international flying traveler. Interviews were conducted by people who had some training, but no method was used to select subjects other than to ask people if they were willing to participate in the survey. It is likely that responses were biased for these reasons and others.

The responses to several questions should have signaled to the investigators that something was wrong. When the British were asked if they ever inquired of a hotel what that hotel's environ-

mental policies might be, none responded "very often." A majority said "never." Yet these respondents who are so unlikely to ask a hotel what its policies are, professed to be deeply concerned with how hotels treat the environment. If concern never translates into behavior, a researcher might well ask if there was a purpose to this whole exercise. The researcher might then look at the question concerning the use of recycled toilet paper or the questions where similar questions failed to generate similar responses and might come to the realization that something had gone wrong. At the least, inconsistent results so frequently found in this survey should have inspired the researchers to try administering the survey to another set of groups. Compar-

isons of the two sets of surveys would have helped establish the reliability or non-reliability of the instrument.

References

¹ "Consumer Attitudes Toward the Role of Hotels in Environmental Sustainability," *Hotel Online* (July 23, 2002).

² Volkema, "Managing the problem formulation process: guidelines for team leaders and facilitators," *Human Systems Management* 16(1997): 27-34.

³ It is important to note that the authors agree with the goals and objectives of those who conducted the survey. But this recently published report should be reviewed for possible methodological errors (we call them biases) before the results can be relied on for policy decision making. The authors hope that this article will help improve the methods used in subsequent studies in this area.

Alan Parker is a professor and Marcel Escoffier is an associate professor in the School of Hospitality and Tourism Management and chair of the Faculty Senate at Florida International University.