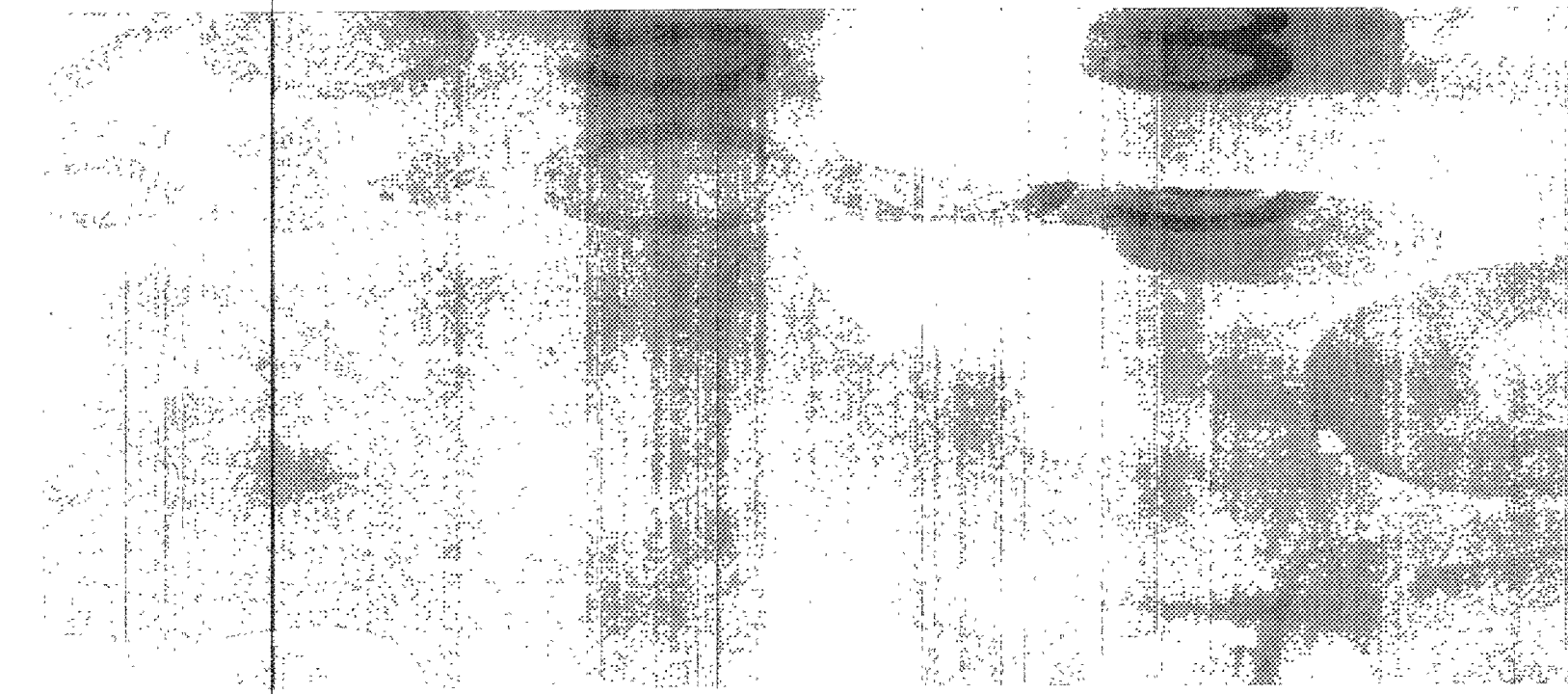


CAN WE TRUST THE DA

The Web has changed the face of research, b

Three umpires are asked to justify their calls. The first, an empiricist by persuasion, says, "I call them as I see them." The second, with the authority of a philosophical realist, says, "I call them as they are." Not to be outdone, the third umpire, an operationist, says, "The way I call them—that's the way they are."

Working with people, with measures of values, attitudes, behavior, and with weak relationships, we researchers do our best to describe consumer and business buyer behavior, explain market response, and estimate market share. But, like the umpires in the story, we are left to our own devices when asked to justify the calls we are making. The prediction of future sales, the go/no-go recommendation for a new product, the identification of a target segment—we would love to have confidence about these things. But, in truth, we remain uncertain about the validity of our measures and methods. The recent growth in online research has done little to reduce this uncertainty.

A OF ONLINE RESEARCH?

raditional cautions still apply. By Thomas W. Miller

Exhibit 1 shows online survey results from "Studies of Information, Research, and Consulting Activities" (SIRCS) cooperative research with Modalis Research Technologies, supported by the Institute for International Research. An expert panel of information users (66 research clients) and providers (239 research suppliers) identified types of online research conducted and used within their organizations. Attitude and behavior surveys, customer and employee satisfaction studies, advertising research, and concept tests were seen as the most common types of online research. Among panelists whose organizations were using online research, 93% predicted their organizations would be using online research more extensively in the future.

Exhibit 2 shows SIRCS ratings of online research relative to traditional research. Across 18 research characteristics, a rating of 1 implied online research was "much worse" than traditional research, and a rating of 5

EXECUTIVE SUMMARY

Online research is a research growth area. More than just a new technology, it represents a change in the way we conduct and think about research. With raw data just a few clicks away, most of us understand the time and cost advantages of an online medium. But do we understand the limitations of online data? This article describes a multimethod approach for learning about the data of online research and compares online and traditional research methods

implied online research was "much better." Information users were somewhat more favorable to online research than information providers Many panelists saw online research as being faster and less costly than traditional research, and many also thought it was less susceptible to interviewer bias In a number of other areas, however, online research was seen as being inferior to traditional research. Of particular concern to panelists were issues of measurement validity and sample representativeness

Measurement Matters

In marketing research, we never eliminate uncertainty But through the use of appropriate models and methods, we can at least measure uncertainty In order to measure uncertainty about online data, we need to conduct experimental and observational multimethod research, examining respondent, stimulus, and method effects.

To understand fully the meaning of online data, we must understand the context of online research What attitudes and experiences do people bring to the online medium? How do they react to technology? Marketing stimuli and measurement instruments appear differently when presented on a computer screen Respondents use keystrokes and mouse clicks in place of spoken words and pencil marks, and they may act and describe themselves differently when working online

The size and type of monitor, the resolution and color pallet—these affect the appearance of online surveys Each Web browser renders Web pages in its own way We usually don't know enough about respondent hardware and software to know that online surveys are displayed the way we want them displayed The image of a rating scale, especially a horizontally oriented scale, can be distorted by the Web browser or by the graphical user interface (GUI), changing distances between scale points

Scrolling is an issue, especially for Web survey pages with many items or many item alternatives Some respondents may view survey pages in their entirety, while others are forced to use scrollbars or cursors to view them Recognizing difficulties associated with scrolling, developers of survey tools suggest the use of short survey pages. Some online survey tools require respondents to click on a "continue" button at the bottom of each page before moving to the next page Scrolling, viewing additional pages, and using extra key clicks to move safely from one page to the next can make survey tasks longer and more tedious

Web-based surveys, like traditional mail surveys, are self-administered Respondents can resize and reposition windows and choose tiled or cascaded windows They also can run concurrent applications, split their attention between surveys and other tasks, walk away from surveys, or ask others to respond to surveys.

Online research presents technical challenges Message length, bandwidth, and network traffic affect the time it takes to transmit a message across a network. Wealthier, more highly educated, professional, and technically oriented respondents are likely to have faster, more powerful computers and wider bandwidth connections Even with sufficient bandwidth on the link between a respondent's workstation and his/her Internet service provider (ISP), heavy network traffic can cause long delays. Online survey transmission times, like other Internet transmission times, vary by season, day of the week, and time of day

Concerned about a lack of standardization in computer hardware, software, and network configurations, online research providers encourage the use of standardized software and minimum hardware configurations Digital Marketing Services relies upon scheduled, periodic software upgrades to the customers of America Online. Knowledge Networks has taken the extreme approach of providing survey respondents with Web-TV devices and cable-TV-based Internet connections, enforcing its own form of standardization

We have many questions to answer about online measures—questions best answered within a framework for multimethod research. Exhibit 3 illustrates how marketing data arise from reactive and nonreactive measures and methods (See page 30.) When we raise concerns about the reliability and validity of marketing measures, we are usually talking about reactive measures, such as respondent self-reports and ratings

We distinguish among four types of reactive multimethod research, defined by whether the research is experimental or observational and whether the research involves a between- or within-subjects design Experimental research involves randomization, observational research does not For between-subjects experimental research, subjects are randomly assigned to methods (e g , online and traditional measurement methods), with each subject receiving only one method. For within-subjects experimental research, each subject receives more than one method with the order of methods being randomized When we run experiments, we can make direct comparisons between methods because we have "equated" participant groups through random assignment of participants to methods

Within-subjects observational research builds upon a paradigm popularized by Donald T Campbell and Donald W Fiske in their 1959 *Psychological Bulletin* paper "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix " To learn about the data of online research within a Campbell-Fiske paradigm, researchers make concurrent online and traditional measurements Within-subjects observational studies permit a direct comparison of online and traditional measures because each subject is serving as his/her own control

Between-subjects observational research examines the reliability and validity of measures by posing the same study questions to separate online and traditional respondent groups. For observational between-subjects research, we have two ways of

demonstrating comparability—with and without adjustment for demographic differences between groups. When we observe similar response distributions, either before or after adjusting for demographic differences, we say that online and traditional methods are comparable.

Studies in Comparability

Are online measures comparable to traditional measures? In his tutorial at the 2000 EXPLOR Forum, Jeff Miller of Burke Inc presented results from experimental between-subjects multimethod studies. With survey participants assigned at random to online survey and telephone interview conditions, significant differences emerged in the patterns of survey response on 5-point scales of purchase intent. People responding to online survey questions about their likelihood of purchasing household consumer products were less likely to use the endpoints of scales, labeled "definitely will buy" and "definitely will not buy," than people responding to telephone interview questions about the same products. How people respond to an item when they see scale point descriptions (as in mail or online surveys) may be different from how people respond when they listen to scale point descriptions (as in a telephone interview).

In "Exploring the Possibilities of Online Focus Groups," a between-subjects experimental study of participant behavior in focus groups, we used common discussion guides and moderators across three modalities: face-to-face, telephone, and online. We observed differences in speech segment length and total transcript length across modalities. Additional cooperative research with The Gallup Organization indicated fundamental differences in the ways people express themselves in face-to-face, telephone, and online focus groups. Using computer-assisted thematic content analysis to examine the transcripts, we observed that people use a higher proportion of strong words (both positive and negative) when working online. We might explain these response-to-media differences by noting that typing is different from speaking and that people are less susceptible to social pressure (or social graces) when working online. People also may be more honest online because they have a feeling of anonymity.

In "Reliability and Comparability of Choice-Based Measures," we asked a simple comparability question: "Are measures obtained from the online administration of choice tasks comparable to measures obtained from paper-and-pencil administration?" University student subjects made choices between pairs of hypothetical job offers, with each student making choices for 24 pairs of jobs in test and retest phases. This experimental research involved between- and within-subjects components. Students assigned to comparability test conditions received both online and paper-and-pencil surveys, with the order of surveys randomized. We used the proportion of choices in agreement between online and paper-and-pencil surveys as an index of comparability between surveys. Results indicated high levels of comparability for this choice task.

The ACNielsen BASES parallel testing program, as reported by Joseph Wilke at the 2000 Advanced Research Techniques Forum, was a between-subjects research program involving 50 parallel tests in 1998 and 25 in 1999 and 2000. Sixteen consumer products manufacturers participated in studies designed to examine the comparability of online and traditional mall-intercept methods. ACNielsen BASES researchers observed high correlations between aggregate measures of new product purchase intent from online and mall-intercept groups. Scores for purchase intent may differ between online and mall-intercept groups (with online groups often yielding lower levels of purchase intent), but online methods can supply valid sales forecasts when researchers calibrate forecasts against appropriate norms.

EXHIBIT 1 Types of online research conducted

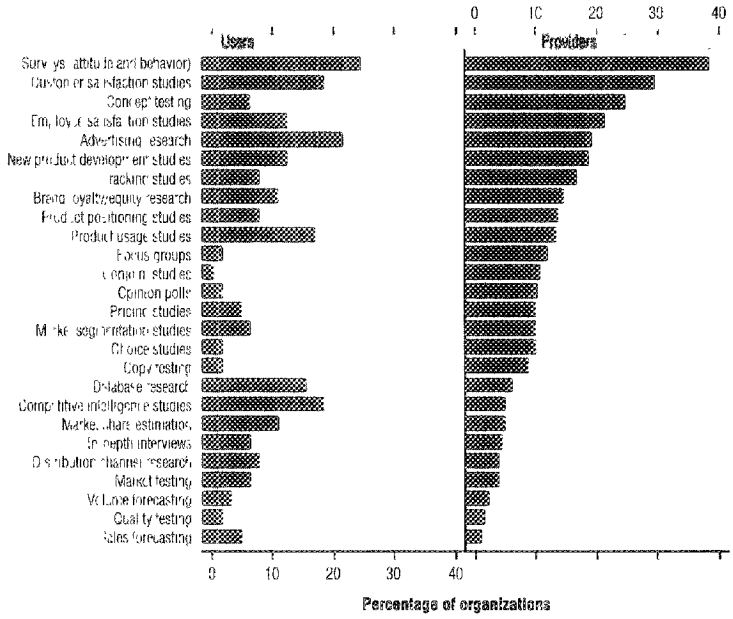
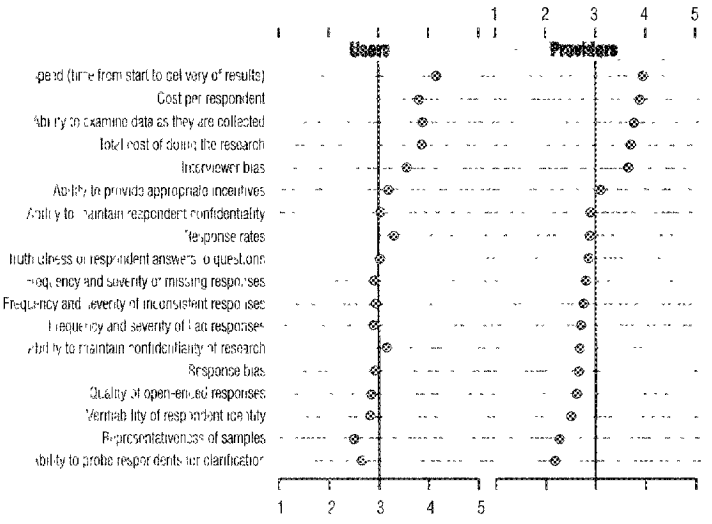
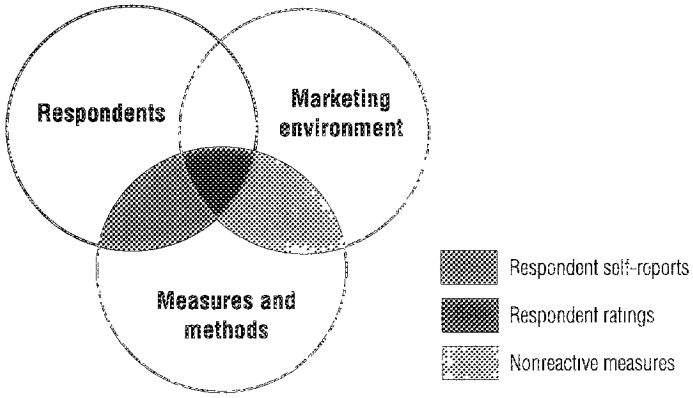


EXHIBIT 2 Online research vs. traditional research



Mean rating on online research: (1) much worse, (2) worse, (3) neither better nor worse, (4) better, (5) much better.

EXHIBIT 3 Framework for multimethod research



Can differences between online and traditional survey results be explained by demographic differences between online and traditional respondent groups? Despite our best efforts to build parallel groups of respondents, online and traditional groups often differ on key demographic variables. Propensity scoring helps us adjust for demographic differences between groups. We can use propensity scoring in between-subjects observational studies. We let an individual's online propensity score be his/her estimated probability of being in the online group. We develop a propensity scoring formula by logistic regression, predicting the method of administration (online or traditional) from demographic variables.

In cooperative research with the Chicago office of Mintel International Group Ltd., we worked with data from online and telephone panels. Common questions and demographic categories were used in the panel surveys. For example, one survey question asked, "When you go to the movie theatre, how much would you expect to spend in total on a ticket and refreshments per person?" When we examined data from online and telephone respondents, we observed slightly higher levels of movie spending in the telephone panel. Recognizing that online and telephone groups, as well as per-person movie spending, could differ by geography and household composition, we used propensity scoring to adjust for demographic differences between the online and telephone panels. We divided the combined group of online and telephone participants into quintiles (five blocks of approximately equal size) on the basis of propensity scores. For the movie-spending question, propensity scoring suggested that differences between online and telephone means could be explained by demographic differences between the two panels.

Similar patterns of response for online and traditional groups within propensity score quintiles demonstrate comparability between methods. When different patterns emerge, propensity scoring sometimes helps identify the sources of those differences. What should researchers do when observational studies yield different results for online and traditional groups after adjustment for demographic differences between groups?

Until household access to the Internet is more widespread, there may be good reasons to adjust for demographic differences between online and traditional groups. Many online research providers employ statistical case-weighting methods. While they continue with their development of online purchase intent norms, ACNielsen BASES researchers may be adjusting online results to match traditional mall-intercept results. Analysts like George Terhanian of HarrisInteractive use propensity scoring as an adjustment methodology to bring online survey results in line with traditional survey results.

There are also good reasons not to use case-weighting methods. Researchers like Donna Wydra of Market Facts and Jeff Miller of Burke Inc. have argued against case-weighting methods. Karl Irons of The NPD Group suggests that adjusting online results to match mall intercept, phone, or mail results sends the wrong message to the users of online research data. A goal of marketing research is to understand consumer and business buyer behavior. We want to predict marketplace behavior, not mall, phone, or mail survey-taking behavior. Recall the maxim: Listen to what people do, not to what they say. When online measures do a better job of predicting marketplace behavior, then online measures are preferred.

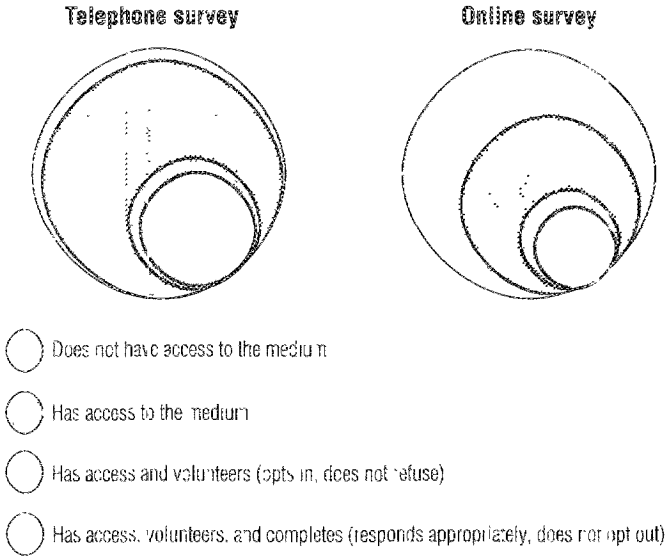
Adjusting for participant group differences and weighting to make online respondent groups look more like traditional respondent groups or more like the general population may seem appropriate. As experimental studies show, however, there can be response-to-media differences, including systematic differences in respondent behavior across face-to-face, mail, telephone, and online media. Furthermore, propensity scoring and case-weighting methods are only as good as the data we use. If self-selection into online vs. traditional survey groups were merely a matter of demographics (or the limited demographics for which we have measures), then we might be able to equate groups by statistical adjustment. But self-selection into online and traditional groups is also likely to be affected by attitudes, experiences, and personal preferences.

Coping with Nonsampling Errors

Lack of measurement control and standardization, limited access to online media, respondent refusals, nonresponse, inaccurate response, and incomplete survey response represent sources of nonsampling error. Coverage is a particular problem for online consumer surveys because many households have no chance of being included in online samples. If we want to use online samples to make inferences about the general population, then we must recognize that at the present time fewer households have access to the Internet than to mail or telephone.

Exhibit 4 provides a summary of nonsampling errors as they might apply to telephone and online surveys. Suppose we were interested in making inferences about the general population of the United States as represented by the outer circles in the exhibit. Inner circle areas are suggestive of proportions of the general population that might have access to a research medium (e.g., telephone or online) and are included in household samples. It is reasonable to think of household telephone access as being high (90% to 95%) relative to household Internet access (around 50%). Voluntarism (nonrefusal, agreeing to participate or opt in) often differs across survey methods and is highly influenced by incentives provided to respondents. Obtaining logically correct and complete responses is also likely to vary across sur-

EXHIBIT 4 Nonsampling errors and coverage



vey methods. The innermost circles of the exhibit suggest the combined effect of these nonsampling errors. Because of access problems, most online samples can be expected to cover a smaller proportion of the general population (shown here as 10%) than telephone surveys (shown as 20%). The exhibit's innermost circle area for a telephone survey may not be too far from the truth. Bickart and Schmittlein described research showing that 20% to 23% of adults accounted for all survey responses in the United States.

If we were willing to settle for the 10% or so of people with both household access to the Internet and a willingness to participate in online research, we would have a difficult time reaching them directly through the Internet. There is no central listing of e-mail addresses and no online analogue of the telephone book or random-digit dialing. Furthermore, sending unsolicited e-mail to potential research participants is inappropriate, and is illegal in some jurisdictions.

Coverage errors arise from problems of access and control. Undercoverage results when population units have zero probability of selection, improper coverage occurs when population units outside the sample frame are mistakenly included in the sample. Because online surveys are self-administered, researchers need to define rigorous controls regarding access to surveys. Targeted respondents may try to access an online survey more than once or give access instructions to family members, friends, or colleagues.

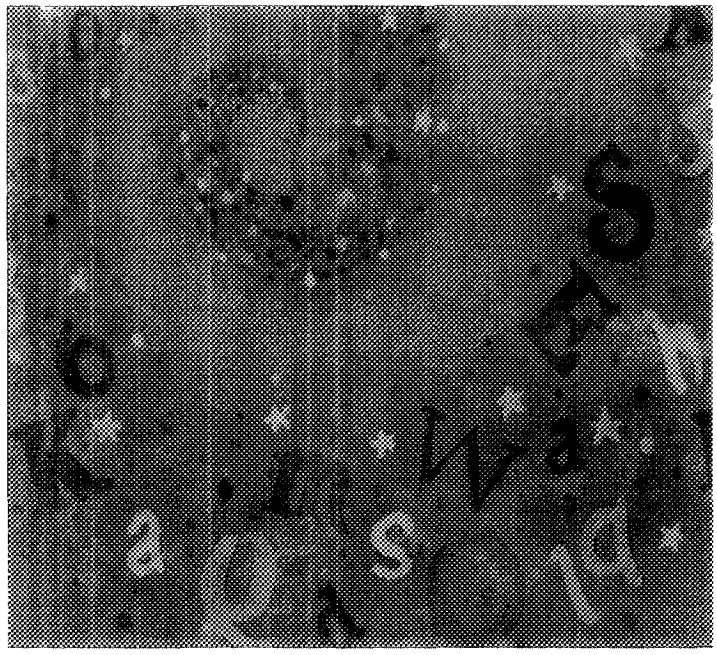
To increase sample representativeness, research providers often employ stratified and quota sampling plans. To increase response rates, they can use mixed-mode surveys, gathering data across a variety of media. Phone recruitment and online survey administration can be a useful, although expensive, mixed-mode approach. Incentives may help increase response rates across all research media. Knowledge Networks' way of dealing with coverage problems was to build its panel through traditional telephone recruiting with random-digit dialing, providing data entry stations to all members, many of whom were neither computer nor Internet users.

Developers of large online panels strive for demographic balance with the general population. Their panels, though not probability samples from the general population, have been used extensively in marketing research. Sample Surveying Inc. has developed volunteer panels of online respondents in specific product categories, such as auto, communications, computers, education, finance, health, and travel. For many research problems a specialized panel may be more appropriate than a probability sample from the general population.

The economics of research changes with technology. Although variable costs associated with sample recruitment and panel maintenance are often similar for online and traditional research, variable costs for data collection are much lower for online research. Larger samples are less expensive with online methods, and larger samples translate into lower sampling errors. Accordingly, online research can be of great benefit when nonsampling errors are under control and when there is adequate coverage of the target population.

Making the Call

Lacking appropriate external criteria for the evaluation of research, it is often difficult to choose among alternative research methods. Some argue the method of research makes no difference, that managers will make the same business decisions whether guided by data from online or traditional methods. When the objective of research is to get a ranking of marketing stimuli or products from most to least preferred, as we might obtain from a product concept test, then we would expect online and traditional methods to support similar decisions. But when the objective is to obtain an accurate forecast of sales or market share, then the method of research may well make a difference. Given our concerns about control and standardization in Web-browser-based surveys, we argue that online measures require special care in their development. We also urge caution in the



interpretation of online data, especially in light of the risk of non-sampling errors.

Can we trust the data of online research? Like the baseball umpires in our opening story, marketing researchers will be asked to call ball or strike, fair or foul, safe or out. Better to make these calls with our eyes open, focused on the data at hand, than to say, "I'm sorry, my head was turned when marketing research went through its transformation to online media, measures, and methods." The use of online research is growing quickly, and we still have much to learn about the data it produces. ●

Additional Reading

Bickart, Barbara, and David Schmittlein (1999), "The Distribution of Survey Contact and Participation in the United States: Constructing a Survey-Based Estimate," *Journal of Marketing Research*, 36, 286-294

Brewer, John, and Albert Hunter (1989), *Multimethod Research: A Synthesis of Styles*. Newbury Park, CA: Sage Publications

Doyle, Kenneth O. (1998), "Research Note: Reliability and Dimensionality of Stimulus Ratings," *Journal of Advertising Research*, (March-April), 45-50

Esipova, Neli, Thomas W. Miller, Michal D. Zarnecki, Jason Elzaurdia, and Sanjay Ponnaiya (2001), *Exploring the Possibilities of Online Focus Groups*. Madison, WI: University of Wisconsin-Madison and ACNielsen Center for Marketing Research

Miller, Thomas W., and Peter R. Dickson (2001), "On-line Market Research," *International Journal of Electronic Commerce*, 5 (3), 139-167

Miller, Thomas W., and Abhilasha Gupta (2001), *Studies of Information, Research, and Consulting Services* (SIRCS) *Fall 2000 Survey of Organizations*. Madison, WI: University of Wisconsin-Madison and ACNielsen Center for Marketing Research

Miller, Thomas W., David Rake, Takashi Sumimoto, and Peggy S. Hollman (2001), *Reliability and Comparability of Choice-Based Measures: Online and Paper-and-Pencil Methods of Administration*. Madison, WI: University of Wisconsin-Madison and ACNielsen Center for Marketing Research

Thomas W. Miller is director of the ACNielsen Center for Marketing Research at the University of Wisconsin-Madison. He may be reached at tmiller@bus.wisc.edu