Public Opinion Quarterly Advance Access published December 1, 2009

Public Opinion Quarterly 2009, pp. 1–38

NATIONAL SURVEYS VIA RDD TELEPHONE INTERVIEWING VERSUS THE INTERNET
COMPARING SAMPLE REPRESENTATIVENESS AND RESPONSE QUALITY

LINCHIAT CHANG
JON A. KROSNICK

Abstract In a national field experiment, the same questionnaires were administered simultaneously by RDD telephone interviewing, by the Internet with a probability sample, and by the Internet with a nonprobability sample of people who volunteered to do surveys for money. The probability samples were more representative of the nation than the nonprobability sample in terms of demographics and electoral participation, even after weighting. The nonprobability sample was biased toward being highly engaged in and knowledgeable about the survey's topic (politics). The telephone data manifested more random measurement error, more survey satisficing, and more social desirability response bias than did the Internet data, and the probability Internet sample manifested more random error and satisficing than did the volunteer Internet sample. Practice at completing surveys increased reporting accuracy among the probability Internet sample, and deciding only to do surveys on topics of personal interest enhanced reporting accuracy in the nonprobability Internet sample. Thus, the nonprobability Internet method yielded the most accurate self-reports from the most biased sample, while the probability Internet sample manifested the optimal combination of sample composition accuracy and self-report accuracy. These results suggest that Internet data collection from a probability sample yields more accurate results than do

LINCHIAT CHANG is an independent contractor in San Francisco, CA, USA. JON A. KROSNICK is Professor of Communication, Political Science, and Psychology, Stanford University, 434 McClatchy Hall, 450 Serra Mall, Stanford, CA 94305, USA, and University Fellow at Resources for the Future. This research was funded by a grant from the Mershon Center of the Ohio State University to Jon Krosnick and was the basis of a Ph.D. dissertation submitted by the first author to The Ohio State University. The authors thank Mike Dennis, Randy Thomas, Cristel deRouvray, Kristin Kenyon, Jeff Stec, Matt Courser, Elizabeth Stasny, Ken Mulligan, Joanne Miller, George Bizer, Allyson Holbrook, Paul Lavrakas, Bob Groves, Roger Tourangeau, Stanley Presser, Michael Bosnjak, and Paul Biemer for their help and advice. Address correspondence to Jon A. Krosnick; e-mail: krosnick@stanford.edu.

doi:10.1093/poq/nfp075

telephone interviewing and Internet data collection from nonprobability samples.

During the history of survey research, the field has witnessed many transitions in the uses of various modes of data collection for interviewing nationally representative samples of adults. Initially, face-to-face interviewing was the predominant method of data collection, yielding high response rates, permitting the development of good rapport between interviewers and respondents, and allowing the use of visual aids that facilitated the measurement process. But the cost of face-to-face interviewing increased dramatically since the 1970s (Rossi, Wright, and Anderson 1983; De Leeuw and Collins 1997), prompting researchers to explore alternative methods, such as telephone interviewing, self-administered paper-and-pencil mail questionnaires (Dillman 1978), audio computer-assisted self-interviewing (ACASI), telephone audio computer-assisted self-interviewing (T-ACASI), Interactive Voice Response (IVR) surveys (Dillman 2000), and more. Among these alternatives, telephone interviewing of samples generated by random digit dialing became a very popular method during the last 30 years, an approach encouraged by studies done in the 1970s suggesting that telephone data quality was comparable to that obtained from face-to-face interviews (e.g., Groves and Kahn 1979).

Recent years have seen a surge in the challenges posed by telephone interviewing. It has become increasingly difficult to maintain response rates, causing the costs of data collection to rise considerably. It is possible to achieve response rates nearly as high as those observed 20 years ago, but doing so costs a great deal more. So holding budget constant over time, the response rate that can be obtained today is considerably lower than that which was obtainable in 1980 (Lavrakas 1997; Holbrook, Krosnick, and Pfent 2007).

Against this backdrop, Internet surveys appeared as a promising alternative about 10 years ago. Some survey firms that had concentrated their efforts on telephone interviewing shifted to collecting a great deal of data over the Internet, including the Gallup Organization and Harris Interactive. And other firms were newly created to take advantage of the Internet as a data collection medium, including Knowledge Networks and Greenfield Online.

Resistance to new modes of data collection is nothing new in the history of survey research, and it is as apparent today as it has been in the past. Just as the survey industry was reluctant to embrace the telephone when it emerged decades ago as an alternative to face-to-face interviewing, some researchers today are hesitant about a shift to internet-based data collection when the goal is to yield representative national samples. This skepticism has some basis in reality: there are notable differences between face-to-face, telephone, and mail surveys on the one hand and Internet surveys on the other in terms of sampling and recruitment methods, most of which justify uncertainty about the quality of data obtained via Internet surveys (e.g., Couper 2000).

Nonetheless, practical advantages of Internet surveys are obvious. Computerized questionnaires can be distributed easily and quickly via web sites postings or hyperlinks or attachments to emails. No travel costs, postage or telephone charges, or interviewer costs are incurred. Respondents can complete questionnaires on their own whenever it is convenient for them. Turn-around time can be kept short, and the medium allows easy presentation of complex visual and audio materials to respondents, implementation of complex skip patterns, and consistent delivery of questions and collection of responses from respondents. Therefore, it is easy to understand why many survey practitioners today find the Internet approach potentially appealing in principle. But for the serious research community, practical conveniences are of limited value if a new methodology brings with it declines in data quality. Therefore, to help the field come to understand the costs and benefits of Internet data collection, we initiated a project to compare this method to one of its main competitors: telephone surveying.

We begin below by outlining past mode comparison studies and compare Internet and telephone survey methodologies in terms of potential advantages and disadvantages. Next, we describe the design of a national field experiment comparing data collected by Harris Interactive (HI), Knowledge Networks (KN), and the Ohio State University Center for Survey Research (CSR) and detail the findings of analyses comparing sample representativeness and response quality.

The present investigation assessed sample representativeness by comparing demographic distributions to benchmarks obtained from the Current Population Survey (CPS). We also studied mode differences in distributions of key response variables and assessed data quality using regression coefficients, structural equation model parameter estimates, and systematic measurement error documented by experimental manipulations of question wording.

Internet and Telephone Survey Methodologies

Two primary methodologies have been employed by commercial firms conducting surveys via the Internet. One method, employed by companies such as Harris Interactive (HI), begins by recruiting potential respondents through invitations that are widely distributed in ways designed to yield responses from heterogeneous population subgroups with Internet access. Another approach entails probability samples reached via Random Digit Dialing who are invited to join an Internet survey panel; people without computer equipment or Internet access are given it at no cost. Two firms taking this approach in the United States (Knowledge Networks (KN) and the RAND Corporation) have given WebTV or MSNTV equipment and service to respondents who needed them. These two approaches to recruit panel members have been outlined in detail elsewhere (Couper 2000; Best et al. 2001; Chang 2001; Berrens et al. 2003), so we describe briefly the methods used at the time when our data collection occurred.

Harris Interactive Methodology. Harris Interactive recruited more than three-quarters of their panel members from one of the most popular Internet search engines: www.excite.com. On the main page of excite.com, a link appeared inviting visitors to participate in the poll of the day. Respondents who voted on the day's issue then saw a link inviting them to become panel members for the Harris Poll Online (HPOL). The second source of panel members was the website of Matchlogic, Inc., an online marketing company and a subsidiary of Excite@Home. Matchlogic posted banner advertisements on the Internet to attract consumers with promises of incentives such as free merchandise and sweepstakes. When a person registered for those incentives, he or she was invited to become a panel member for HPOL. At the time when our data collection occurred, Excite and Matchlogic accounted for about 90 percent of all panel members; the others were recruited by invitations on other websites.

People visiting the Harris Poll Online (HPOL) registration site were asked for their email addresses and demographic information and were told that HPOL membership would allow them to influence important decision-makers in government, nonprofit organizations, and corporations, could help to shape policies, products, and services, would have access to some survey results prior to publication in the media, and might win cash, free consumer products, or discount coupons, or receive other tangible incentives.

Harris Interactive's database has contained more than 7 million panel members, and subsets of these individuals were invited to participate in particular surveys. A panel member who was invited to do a survey could not be invited to do another survey for at least 10 days, and each panel member usually received an invitation at least once every few months. Survey completion rates ranged from a low of 5 percent to a high of 70 percent, with an average of 15–20 percent.

To generate a national sample, panel members were selected based on demographic attributes (e.g., age, gender, region of residence) so that the distributions of these attributes in the final sample matched those in the general population. Each selected panel member was sent an email invitation that described the content of the survey and provided a hyperlink to the website where the survey was posted and a unique password allowing access to the survey. Respondents who did not respond to an email invitation or did not finish an incomplete questionnaire were sent reminder emails.

Harris Interactive weighted each sample using demographic data from the Current Population Survey (CPS) and answers to questions administered in Harris Poll monthly telephone surveys of national cross-sectional samples of 1,000 American adults, aged 18 and older. The goal of their weighting procedure was to adjust for variable propensity of individuals to have regular access to email and the Internet to yield results that can be generalized to the general population.

Knowledge Networks Methodology. Beginning in 1998, Knowledge Networks recruited panel members through RDD telephone surveys and provided people with WebTV equipment and Internet access in exchange for their participation in surveys. Knowledge Networks excluded telephone numbers from their RDD samples that were not in the service area of a WebTV Internet service provider, leading to exclusion of about 6–7 percent of the general population. Knowledge Networks attempted to obtain mailing addresses for all sampled telephone numbers and succeeded in doing so for about 60 percent of them. These households were then sent advance letters stating that they had been selected to participate in a survey panel, that they would not pay any cost, that confidentiality was assured, and that a Knowledge Networks staff member would call them within a week. A \$5 or \$10 billion was included with the letter to encourage cooperation.

Telephone interviews were attempted with all households that received an advance letter. Telephone interviews were also attempted with one-third (selected randomly) of the telephone numbers for which an address could not be obtained. During the telephone interviews, respondents were told they had been selected to participate in an important national study using a new technology and that they would be given a WebTV receiver that would allow them free access to the Internet and opportunities to answer brief surveys on their television. Respondents were told that their participation was important and were asked about the extent to which household members were experienced with the Internet and proficient with computers and about some demographics of household members. Arrangements were then made to mail the WebTV equipment to the respondent.

After households received the WebTV equipment and installed it (with assistance from a technical support office via telephone if necessary), respondents completed “profile” surveys that measured many attributes of each adult household member. Each adult was given a free email account and was asked to complete surveys via WebTV. Whenever any household member had a new email message waiting to be read, a notification light flashed on the WebTV receiver (a box located near the television set). Panel members could then log into their WebTV accounts and read the email invitation for the survey, which contained a hyperlink to the questionnaire. Panel members were usually sent one short survey per week, typically not exceeding 15 minutes. When a panel member was asked to complete to a longer questionnaire, he or she was then given a week off or offered some other form of incentive or compensation.

Typically, about 85 percent of respondents who were asked to complete a KN survey did so within two weeks of the invitation, and few responses were received after that. Respondents who failed to respond to eight consecutive surveys were dropped from the panel, and the WebTV receiver was removed from their homes.

Thus, households that intended to provide data for any given survey could fail to do so because of dropout at several stages throughout the recruitment process. Some households were excluded because they did not live in an area covered by a WebTV provider. Some households were not contacted to complete the initial RDD telephone interview. Other households were contacted but refused to join the panel. Of the households that signed up initially, some failed to install the WebTV equipment in their homes. And some people who had the equipment installed either failed to complete a questionnaire for a particular survey or dropped out of the panel altogether after joining it.

Hypotheses About Differences Between Methods

ADVANTAGES OF THE TELEPHONE OVER THE INTERNET

There are many reasons why the quality of survey responses collected via the Internet might differ from those collected by telephone. One potential strength of telephone surveying is the presence of interviewers, who can provide positive feedback to respondents in order to encourage effortful engagement in the response process (Cannell, Miller, and Oksenberg 1981). Likewise, interviewers can project interest and enthusiasm, which may be unconsciously contagious (Chartrand and Bargh 1999), and respondents' moods can be unconsciously enhanced by the emotions in the interviewer's voice (Neumann and Strack 2000). Thus, if interviewers' voices transmit interest and enthusiasm about a survey, they may inspire increased respondent engagement. Such processes cannot occur when respondents complete self-administered questionnaires.

Interviewers can also create a sense of accountability among respondents due to "the implicit or explicit expectation that one may be called on to justify one's beliefs, feelings, and actions to others" (Lerner and Tetlock 1999). In past research, participants who reported their judgments aloud to another person recognized that their judgments were linked directly to them in the eyes of the individual with whom they were interacting, resulting in high accountability (e.g., Price 1987; Lerner, Goldberg, and Tetlock 1998). When the audience's views on the issues in question are not known, accountability generally leads people to devote more careful and unbiased effort to making judgments (for a review, see Lerner and Tetlock 1999). Although survey responding via the Internet to HI and KN is not anonymous, the palpable phenomenology of accountability under those circumstances may be considerably less than when a respondent is conversing with an interviewer. Therefore, this may increase the precision of survey responses provided over the telephone.

Telephone surveys have another potential advantage over Internet surveys: respondents do not need to be literate or be able to see clearly enough to read words, because all questions and answer choices are read aloud to them. Telephone respondents also do not need to be proficient at using a computer or to be knowledgeable about how to navigate the Internet. Thus, telephone surveys may be more manageable than Internet surveys.

ADVANTAGES OF THE INTERNET OVER THE TELEPHONE

Just as interviewers may be advantageous, they may also entail drawbacks, so the absence of interviewers might be a strength of Internet surveys. Interviewers are known to create errors and biases when collecting data (Kiecker and Nelson 1996). Due to misunderstandings, bad habits, or biased expectations, some interviewers occasionally provide inappropriate cues (van der Zouwen, Dijkstra, and Smit 1991) or change the wordings of questions (Lyberg and Kasprzyk 1991). None of this can occur in an Internet survey.

Some studies suggest that people are more concerned about presenting a favorable self-image during oral interviews than when completing self-administered questionnaires (Fowler, Roman, and Di 1998; Acree et al. 1999). If self-administered questionnaires do indeed decrease concern about impression management, people may be less likely to conform to socially desirable standards and more likely to provide honest answers to questions on threatening, anxiety-arousing, or otherwise sensitive questions (e.g., Tourangeau and Smith 1996; Wright, Aquilino, and Supple 1998).

Pauses can feel awkward during telephone conversations, which may induce interviewers and respondents alike to rush the speed of their speech, making it difficult for respondents to understand questions and to calmly reflect on the meaning of a question or think carefully to generate an accurate answer. In contrast, Internet respondents can set their own pace when completing a survey, pausing to deliberate about complex questions and moving quickly through questions that are easy to interpret and answer. In addition, Internet respondents can take breaks when they are fatigued and return refreshed. These factors may facilitate better efficiency and precision in answering by Internet respondents.

Internet respondents have the flexibility to complete a questionnaire at any time of day or night that is convenient for them. Telephone interviewing organizations allow for call scheduling at times that are convenient for respondents, but their flexibility in call scheduling falls short of the 24-hour accessibility of Internet surveys. Thus, Internet respondents can choose to complete a survey when they are most motivated and able to do so and when distractions are minimized, perhaps causing improved response quality.

Telephone respondents need to hold a question and its response options in working memory in order to answer accurately. Because Internet respondents can see questions and response categories, they need not commit them to memory before generating answers. If respondents fail to remember the details of a question after reading it once, they can read the question again. And when long checklists or complex response scales are used, Internet respondents are not especially challenged, because the response options are fully displayed. This may reduce the cognitive burden of the response process and may thereby improve reporting accuracy.

PRACTICE EFFECTS

RDD telephone surveys typically involve respondents who have some experience responding to questionnaires,[1] but KN and HI respondents were members of long-term panels and therefore had regular practice at survey responding.

1. The 2003 Respondent Cooperation and Industry Image Survey conducted by the Council for Marketing and Opinion Research (CMOR) suggested that 51 percent of their respondents had participated in surveys within the past year, an average of five times (Miller and Haas 2003).

A great deal of psychological research shows that practice at cognitive tasks improves performance on them, so regular experience answering survey questions may enhance the accuracy of Internet panel members' responses (Smith, Branscombe, and Bormann 1988; Donovan and Radosevich 1999). Also, panel members may become especially self-aware and introspective about their thoughts, attitudes, emotions, and behaviors, further improving their ability to later report on those phenomena accurately (Menard 1991). Consistent with this reasoning, research on panel surveys has shown that people's answers to attitude questions become increasingly reliable as they gain more experience responding to them (Jagodzinski, Kuhnel, and Schmidt 1987).

POTENTIAL DRAWBACKS OF INTERNET PANELS

A potential drawback of repeated interviewing is "panel conditioning," whereby accumulating experience at doing surveys makes panel members less and less like the general public they are intended to represent. A number of studies exploring this possibility have found either no evidence of panel conditioning effects or very small effects. For example, Cordell and Rahmel (1962) found that participating in Nielsen surveys on media use did not alter later reports of media use. Likewise, Himmelfarb and Norris (1987) found that being interviewed on a wide range of topics did not alter people's subsequent reports of mental health, physical health, self-esteem, social support, or life events experienced (see also Sobol 1959; Clinton 2001). Willson and Putnam (1982) found in a meta-analysis that answering questions caused attitudes toward objects to become slightly more positive, but these effects were quite small and inconsistent across studies.

Some studies that documented conditioning effects tested the "stimulus hypothesis" (Clausen 1968): the notion that interviewing people on a particular topic may induce them to become more cognitively engaged in that topic subsequently. Some studies found support for this notion (e.g., Bridge et al. 1977; Granberg and Holmberg 1991), though others did not (e.g., Mann 2005). Other studies have documented how asking people just one question about their behavioral intentions could impact on subsequent behavior (e.g., Sherman 1980; Greenwald et al. 1987; Fitzsimons and Morwitz 1996). Thus, this literature clearly suggests that panel conditioning effects can occur (see also the literature on pretest sensitization; e.g., Bracht and Glass 1968).

Another potential drawback of panel studies involves respondent attrition: Some of the people who provide data during the first wave of interviewing do not participate in subsequent waves. If a nonrandom subset of respondents drop out, then this would compromise sample representativeness. However, the literature on panel attrition is actually quite reassuring on this point. Although a few past studies have documented instances in which people who were and were not reinterviewed differed from one another in some regard (e.g., Lubin, Levitt, and Zuckerman 1962; Groves, Singer, and Corning 2000), most studies found little or no sample composition changes attributable to panel attrition (e.g., Fitzgerald, Gottschalk, and Moffitt 1998a, 1998b; Falaris and Peters 1998; Zagorsky and Rhoton 1999; Clinton 2001). Further, this literature is based mostly on classical panel designs in which respondents are interviewed repeatedly on the same topic; panel attrition effects could be even less pronounced on Internet panels covering diverse topics over time.

COVERAGE AND NONRESPONSE ERROR

HI samples entailed coverage error, because they included only people who had pre-existing access to computers and the Internet, thus probably over-representing urban residents, men, wealthier, more educated, younger, and White people (Flemming and Sonner 1999; Rohde and Shapiro 2000). Although KN provided Internet access to its panel members, its sampling technique brought with it the same coverage error inherent in all RDD telephone surveys, excluding about 5 percent of the country's population because their households were without working telephones. If respondents who already had Internet access in their homes were more likely to reject the offer of free Internet access via WebTV, then the KN samples would under-represent regular Internet users.

TOPIC INTEREST

The method used by most nonprobability Internet firms to invited respondents to complete a questionnaire may create sample composition bias driven by interest in the topic. In the email invitations sent to selected HI respondents, a one-sentence description informed people about the content of the survey (e.g., telecommunications, entertainment, or politics). People interested in the topic may have been more likely to participate than people not so interested. Although the HI weighting procedure adjusted for demographic attributes and other variables, the adjustment procedure has not usually corrected for interest in a particular topic.

SUMMARY

In sum, if interviewers bring positive reinforcement, enthusiasm, and accountability into the survey process, and literacy is a significant problem, then response quality may be advantaged by the presence of interviewers and oral presentation in telephone surveys. But if a greater sense of privacy, self-pacing, flexibility to complete surveys at any time of day or night, an ability to see questions and response options, and practice effects advantage the Internet mode, then response quality may be better in such surveys than in telephone surveys. Furthermore, coverage and non-response error may advantage RDD surveys

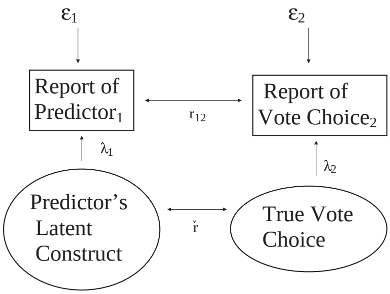

Figure 1. Model of Criterion Validity.

over KN, and may advantage KN over HI. In the present investigation, we did not set out to explicitly gauge the impact of each of the factors outlined above. Rather, we set out to ascertain whether mode differences existed in sample representativeness and response quality, and the extent and direction of such differences if they existed.

The National Field Experiment

For our study, HI, KN, and the OSU CSR collected data in two waves, once before the 2000 U.S. Presidential election campaign began, and then again after election day. During the pre-election survey, respondents predicted their presidential vote choice in the elections, evaluated the leading presidential candidates, and reported a wide range of attitudes and beliefs that are thought to drive vote choices. During the postelection survey, respondents reported whether they had voted, for whom they had voted, and again evaluated the leading presidential candidates. (See online Appendix 1 for the question wordings and variable codings.)

Our comparisons focused on the demographic composition of the samples in terms of education, age, race, gender, and income; the samples' interest in the survey topic; concurrent validity of the measures (i.e., their ability to distinguish between people on a criterion measured at the same point in time; e.g., Leary 1995); predictive validity of the measures (i.e., their ability to predict a criterion measured some time in the future; e.g., Aronson et al. 1990; Leary 1995); the extent of survey satisficing (Krosnick 1991, 1999), reliability, and social desirability response bias.

To assess concurrent and predictive validity, we conducted regressions predicting people's pre-election predictions of their vote choice and their post-election reports of their actual vote choices using a plethora of predictors of presidential candidate preferences. The logic underlying these criterion validity analyses is displayed in figure 1. Each determinant of vote choice (shown in the lower left corner of figure 1) is expected to be associated with true vote choice (shown in the lower right corner of figure 1), and the true magnitude of this association is $\check{r}$. Self-reports of these two constructs appear at the top of figure 1. By correlating self-reports measured pre-election (shown in the upper left of figure 1) with postelection reports of vote choice (shown in the upper right of figure 1), we obtain r_{12}. The lower the validity of the items (represented by λ_1 and λ_2) and the more measurement error in reports (represented by ε_1 and ε_2), the more r_{12} will be weakened in comparison to $\check{r}$. Thus, the observed strength of the relation between a measure of a vote choice determinant and vote choice is an indicator of response quality. The more respondents are willing and able to precisely report vote choice and its determinants, the stronger the relation between these two manifest variables will presumably be.

To assess survey satisficing, which occurs when respondents do not engage in careful and thorough thinking to generate accurate answers to questions (Krosnick 1991, 1999), we looked at nondifferentiation in answering batteries of questions using identical rating scales. Nondifferentiation occurs when respondents rate several target persons or issues or objects nearly identically on a single dimension because they do not devote effort to the reporting process. Although a set of identical ratings across objects may be the result of genuinely similar attitudes, nondifferentiation tends to occur under conditions that foster satisficing.

To test whether the data collection methods differed in terms of the amount of random measurement error in assessments, we made use of multiple measures of candidate preferences administered both pre-election and postelection to estimate the parameters of a structural equation model (see, e.g., Kenny 1979). This model posited that the multiple measures were each imperfect indicators of latent candidate preferences at the two time points and permitted those preferences to change between the two interviews.

Finally, we examined whether the two modes differed in terms of social desirability response bias. The survey questionnaire contained a question about whether the federal government should provide more, less, or the same amount of help for African Americans. Among White respondents, it is socially undesirable to express opposition to government programs to help Black Americans (see Holbrook, Green, and Krosnick 2003). Hence, we could assess the mode difference in social desirability response bias among Whites.

SAMPLES

OSU Center for Survey Research: Data collection was conducted by a group of fourteen supervisors and fifty-nine interviewers; both groups received formal training before working on the project and were continually monitored throughout the field period. Households were selected based on RDD within the forty-eight contiguous U.S. states, and one adult per household was randomly

Table 1. National Survey Samples, Field Periods, and Response Rates

	OSU Center For Survey Research	Knowledge Networks	Harris Interactive
Pre-election survey			
Eligible households	3,500	7,054	12,523
Participating respondents	1,506	4,933	2,306
Response rate	43%	25%[a]	NA
Cooperation rate	51%		
Completion rate		70%	18%
Start date	June 1, 2000	June 1, 2000	July 21, 2000
Stop date	July 19, 2000	July 28, 2000	July 31, 2000
Postelection survey			
Eligible households	1,506	4,143[b]	2,306
Participating respondents	1,206	3,416	1,028
Reinterview rate	80%	82%	45%
Start date	November 9, 2000	November 8, 2000	November 9, 2000
Stop date	December 12, 2000	November 21, 2000	November 26, 2000

[a]This figure is the product of 89% (the rate at which eligible RDD-sampled telephone numbers were contacted for initial telephone interviews) and 56% (the rate at which contacted households agreed to participate in the initial telephone interview and agreed to join the KN panel) and 72% (the rate at which households that agreed to join the KN panel had the WebTV device installed in their homes) and 70% (the rate at which invited KN panel respondents participated in the survey).

[b]Of the 4,933 who completed all of the first three instruments, 790 members were excluded from assignment to the follow-up survey for the following reasons: (a) temporarily inactive status (being on vacation, health problems etc.), (b) some individuals had been withdrawn from the panel, and (c) some individuals had already been assigned to other surveys for the week of the election.

sampled to be interviewed using the "last birthday" method (Lavrakas 1993). As shown in table 1, 1,506 respondents were interviewed pre-election between June 1 and July 19, 2000, and 1,206 of those respondents were interviewed postelection between November 9 and December 12, 2000, after the general elections. For the pre-election wave, AAPOR Response Rate 5 was 43 percent; the cooperation rate was 51 percent. Postelection, the number of completions divided by the number of Wave I respondents yielded a reinterview rate of 80.

Knowledge Networks: AAPOR Contact Rate 2 was about 89 percent for the initial telephone interview to recruit people to join the KN panel. Six to 8 percent of households were ineligible because they were outside of the WebTV service area. Of the interviewed eligible respondents, 56 percent agreed to join the KN panel, and 72 percent of these people eventually had WebTV installed in their homes.

The pre-election survey was conducted in July 2000, and the postelection survey was conducted in November 2000. The pre-election questionnaire was divided into three separate modules. Respondents were invited to complete the second module one week after they had been invited to complete the first module, and invitations to complete the third module were sent out two weeks after the invitations for the second module were sent out. Seven thousand fifty four people were invited to complete the first module. Four thousand nine hundred thirty three respondents completed all three modules within four weeks after the invitations for the first module were sent out, yielding a panel completion rate of 70 percent. Of the 4,933 respondents who completed the entire pre-election questionnaire, 790 were excluded from assignment to the postelection survey for varying reasons.[2] The remaining 4,143 people were invited to complete the postelection survey on November 8, 2000, and 3,416 did so within two weeks after receiving the invitation, yielding an 82 percent reinterview rate among those who completed the pre-election survey.

Harris Interactive: In June, 2000, 12,523 participants were pulled from the HPOL database stratified by gender, age, and region of residence (Northeast, South, Midwest, and West). The selected sample matched population parameters (from CPS data) in terms of distributions of age and region of residence, and there was an oversample of male respondents (because HI expected that nonrespondents were more likely to be male than female). Two thousand three hundred six respondents completed the pre-election questionnaire, yielding a completion rate of 18 percent.

After the election in November, 2000, these respondents were invited to complete the postelection survey, and 1,028 did so, yielding a reinterview rate of 45 percent among those who had completed the pre-election survey. No incentives were offered to respondents in exchange for their participation in this study.

DEMOGRAPHIC REPRESENTATIVENESS OF THE PRE-ELECTION SAMPLES

The demographics of the American adult population were gauged using the Annual Demographic Survey supplement of the Current Population Survey (CPS) conducted in March, 2000.[3] Table 2 displays these data and the

2. These included people who were temporarily on inactive status (e.g., on vacation, experiencing health problems, or too busy), people who had been dropped from the panel, and people who were assigned to complete other surveys instead.
3. The CPS is a monthly survey administered by the Census Bureau using a sample of some 50,000 households. Selected households participate in the CPS for four consecutive months, take eight months off, and then return for another four months before leaving the sample permanently.

demographics of the three pre-election samples. For each house, the left column shows the distributions for the unweighted samples, and the right column shows the distributions for the samples weighted using the weights provided to us by the data collection organizations. Under each column of percentages for a demographic variable is the average deviation of the results from the comparable CPS figures.

Focusing first on the unweighted samples, the CSR sample manifested the smallest average deviation for three variables (education, income, and age), whereas KN manifested the smallest deviations for two other variables (race and gender).[4] The HI sample consistently manifested the largest average deviations from the population. As shown in the bottom row of table 2, the average deviation for the unweighted samples was 4.0 percentage points for CSR, 4.3 percentage points for KN, and 8.7 percentage points for HI.

Education: The CSR sample under-represented the least educated individuals and over-represented individuals with college degrees or postgraduate degrees. A similar bias was present in the KN sample: people with high school education were under-represented, whereas people with more education were over-represented. The same bias was even more pronounced in the HI sample, which severely under-represented people with some high school education and high school graduates, and substantially over-represented people who had done postgraduate studies.

Income: The CSR sample under-represented the lowest income individuals; this bias was stronger in the KN sample and even more pronounced in the HI sample. All three samples over-represented the highest income individuals.

Age: The CSR sample under-represented individuals under age 25 and over age 75, but discrepancies from the population statistics were never large. Discrepancies were larger in the KN sample, which under-represented individuals under age 25 and over age 65. The same biases were most apparent in the HI sample, which substantially under-represented people over age 65.

Race: The CSR sample under-represented African-American respondents, and the KN and HI samples evidenced this same bias more strongly. The CSR sample under-represented White respondents, whereas the KN and HI samples over-represented Whites. All three samples over-represented people of other races, with the CSR sample doing so the most.

Participants in the CPS are 15 years old or older and are not institutionalized nor serving in the military. The questionnaire is administered via either telephone or face-to-face interviewing.

4. The initial sample of panel members invited to do the pre-election KN survey was very similar to the subset of those individuals who completed the survey, so discrepancies of the KN sample from the population were largely due to unrepresentativeness of the sample of invited people, rather than due to biased attrition among these individuals who declined to complete the questionnaire.

Table 2. Demographic Composition of Pre-election Samples Compared to CPS data

	OSU Center for Survey Research		Knowledge Networks		Harris Interactive		2000 CPS March Supplement
	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	
Education							
Some high school	7.0%	17.1%	6.7%	12.3%	2.0%	7.9%	16.9%
High school grad	31.3%	32.7%	24.4%	33.5%	11.8%	36.5%	32.8%
Some college	19.6%	19.8%	32.3%	28.5%	36.6%	26.9%	19.8%
College grad	30.1%	21.7%	26.0%	18.2%	25.8%	19.8%	23.0%
Postgrad work	12.0%	8.6%	10.6%	7.4%	23.7%	9.0%	7.5%
TOTAL	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
N	1504	1504	4925	4925	2306	2250	
Average error	4.6	0.5	7.4	3.8	13.9	4.9	
Income							
<$25,000	19.0%	19.0%	14.3%	18.0%	12.6%	24.8%	30.5%
$25–50,000	36.9%	37.1%	32.5%	35.3%	32.3%	29.8%	28.3%
$50–75,000	22.0%	22.4%	27.5%	25.8%	25.9%	20.6%	18.2%
$75–100,000	12.9%	13.4%	13.8%	11.9%	14.8%	11.6%	10.1%
>$100,000	9.2%	8.1%	11.9%	9.0%	14.5%	13.0%	12.5%
TOTAL	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
N	1138	1138	4335	4335	1976	1917	
Average error	6.0	6.4	6.8	6.5	8.6	2.3	
Age							
18–24	10.0%	13.5%	7.8%	9.8%	8.0%	14.0%	13.2%
25–34	17.9%	15.3%	19.1%	19.1%	21.2%	18.9%	18.7%
35–44	24.5%	22.7%	25.8%	22.8%	21.5%	21.8%	22.1%
45–54	20.7%	17.8%	23.0%	19.8%	27.9%	20.4%	18.3%

Continued

Table 2. Continued

	OSU Center for Survey Research		Knowledge Networks		Harris Interactive		2000 CPS March Supplement
	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	
55–64	12.1%	12.4%	12.4%	13.4%	15.5%	10.4%	11.6%
65–74	9.4%	12.5%	7.7%	9.7%	4.8%	12.3%	8.7%
75+	5.5%	5.8%	4.2%	5.5%	1.0%	2.2%	7.4%
TOTAL	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
N	1496	1496	4923	4923	2306	2250	
Average error	1.7	1.6	2.7	1.5	4.6	1.9	
Race							
White	78.5%	83.3%	86.4%	82.8%	89.6%	81.1%	83.3%
African American	9.7%	11.9%	6.9%	10.0%	3.6%	12.3%	11.9%
Other	11.8%	4.8%	6.7%	7.2%	6.8%	6.6%	4.8%
TOTAL	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
N	1490	1490	4721	4721	2183	2132	
Average error	4.7	0.0	3.3	1.6	5.5	1.5	
Gender							
Male	45.1%	46.9%	49.2%	49.2%	60.1%	48.2%	48.0%
Female	54.9%	53.1%	50.8%	50.8%	39.9%	51.8%	52.0%
TOTAL	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
N	1506	1506	4910	4910	2306	2250	
Average error	2.9	1.1	1.2	1.2	12.1	0.2	
Average error	4.0	1.9	4.3	2.9	8.7	2.2	

Note.—Average errors are expressed in percentage points.

Gender: The CSR sample over-represented women, whereas the HI sample over-represented men. The KN sample's gender composition closely matched the population, and the HI sample was most discrepant.

Impact of weighting: The CSR weights adjusted for probability of selection using number of voice telephone lines and number of adults in the household, poststratified using the March 2000 CPS using age, education, income, race, and gender. KN similarly adjusted for unequal probabilities of selection using number of voice telephone lines per household and several sample design features and used rim weighting (with ten iterations) to adjust according to the most recent monthly CPS figures. The HI weights used CPS data and answers to some questions administered in monthly telephone surveys of national cross-sectional samples of 1,000 adults, aged 18 and older as benchmarks in terms of gender, age, education, race, ethnicity, and a variable representing the propensity of an individual respondent to have regular access to the Internet. In table 2, the right column under each house's label shows the distributions of the demographics after the weights were applied. Not surprisingly, weighting shrunk the demographic deviations from the population considerably.

DEMOGRAPHIC REPRESENTATIVENESS OF THE POSTELECTION SAMPLES

Table 3 shows the distributions of the demographics of the postelection samples in the same format as was used in table 2. Among the unweighted samples, the CSR sample continued to manifest the smallest average deviations for education, income, and age, and the KN sample maintained the smallest deviations for race and gender. The HI sample showed the largest average deviations from the population on all five attributes. As shown in the bottom row of the table, the average deviations for the unweighted samples were 4.5 percentage points for the CSR sample, 4.3 percentage points for the KN sample, and 9.3 percentage points for the HI sample. Weighting had a similar effect here to that observed in table 2.

INTEREST IN THE SURVEY'S TOPIC

A number of indicators suggest that the HI respondents were considerably more interested in the topic of the survey (politics) than were the CSR and KN respondents (see table 4). The CSR respondents gave significantly fewer correct answers to the political knowledge quiz questions than the KN respondents (average percent correct answers given, unweighted: 53 percent versus 58 percent; $b = .09$, $p < .001$; weighted: 50 percent versus 62 percent, $b = .08$, $p < .001$) and HI respondents (unweighted: 53 percent versus 77 percent, $b = .24$, $p < .001$; weighted: 50 percent versus 70 percent,

Table 3. Demographic Composition of Postelection Samples Compared to CPS data

	OSU Center For Survey Research		Knowledge Networks		Harris Interactive		2000 CPS March Supplement
	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	
Education							
Some high school	6.6%	17.1%	7.0%	13.5%	1.1%	7.5%	16.9%
High school grad	29.1%	31.6%	25.9%	32.9%	10.9%	39.5%	32.8%
Some college	20.1%	21.1%	31.9%	28.2%	35.5%	27.1%	19.8%
College grad	31.6%	21.7%	24.9%	18.3%	26.8%	17.3%	23.0%
Postgrad work	12.6%	8.5%	10.3%	7.1%	25.8%	8.6%	7.5%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
N	1201	1201	3404	3404	1040	1040	
Average error	5.6	1.0	6.7	3.4	15.1	6.0	
Income							
<$25,000	17.1%	17.5%	15.0%	19.9%	10.0%	18.9%	30.5%
$25–50,000	36.9%	37.7%	33.4%	36.3%	32.1%	31.9%	28.3%
$50–75,000	22.4%	22.3%	27.6%	25.4%	27.1%	20.9%	18.2%
$75–100,000	14.4%	14.7%	13.1%	10.9%	15.9%	12.8%	10.1%
>$100,000	9.3%	7.8%	10.8%	7.5%	15.0%	15.5%	12.5%
TOTAL	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
N	917	917	3006	3006	882	882	
Average error	6.7	7.2	6.9	6.3	8.3	4.7	
Age							
18–24	8.1%	12.9%	5.9%	9.5%	6.3%	15.7%	13.2%
25–34	17.2%	15.9%	18.2%	20.6%	18.7%	17.5%	18.7%
35–44	24.6%	22.4%	24.3%	22.7%	19.6%	22.0%	22.1%
45–54	22.1%	18.2%	22.9%	19.1%	30.5%	19.3%	18.3%

	OSU Center For Survey Research		Knowledge Networks		Harris Interactive		2000 CPS March Supplement
	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	
55–64	12.1%	11.7%	14.0%	13.1%	17.6%	11.1%	11.6%
65–74	10.1%	13.1%	9.5%	9.2%	6.4%	12.7%	8.7%
75+	5.7%	5.8%	5.4%	5.7%	0.9%	1.6%	7.4%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
N	1197	1197	3408	3408	1040	1040	
Average error	2.4	1.4	2.8	1.6	5.2	2.2	
Race							
White	79.7%	83.2%	87.5%	81.9%	91.2%	81.4%	83.3%
African American	9.0%	11.9%	6.6%	10.3%	2.9%	12.7%	11.9%
Other	11.3%	4.8%	5.1%	7.9%	5.8%	5.8%	4.8%
TOTAL	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
N	1192	1192	4721	4721	1040	1040	
Average error	4.3	0.0	3.3	2.1	6.0	1.2	
Gender							
Male	44.6%	47.1%	49.8%	48.0%	59.8%	48.8%	48.0%
Female	55.4%	52.9%	50.2%	52.0%	40.2%	51.2%	52.0%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
N	1203	1203	4910	4910	1040	1040	
Average error	3.4	0.9	1.8	0.0	11.8	0.8	
Average error	4.5	2.1	4.3	2.7	9.3	3.0	

NOTE.—Average errors are expressed in percentage points.

Table 4. Indicators of Interest in Politics

		OSU Center for Survey Research		Knowledge Networks		Harris Interactive	
		Unweighted sample	Weighted sample	Unweighted sample	Weighted sample	Unweighted sample	Weighted sample
Political knowledge quiz							
Average percentage of correct responses per respondent		53.0%	50.0%	58.0%	62.0%	77.0%	70.0%
	N	1506	1506	4940	4935	2306	2250
Mid-point selection							
Average percentage of midpoint selections per respondent		43.2%	43.8%	39.4%	39.5%	34.0%	33.9%
	N	1506	1506	4940	4935	2306	2250
Party identification							
Percentage of independents		21.8%	23.3%	22.0%	23.6%	13.1%	13.6%
	N	1461	1458	4792	4803	2306	2250
Pre-election reports of electoral participation							
Will vote in presidential election?	Yes	86.2%	84.6%	81.5%	78.5%	94.8%	90.7%
	No	13.8%	15.4%	18.5%	21.5%	5.2%	9.3%
	Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
	N	1456	1452	4914	4915	2313	2250

		OSU Center for Survey Research		Knowledge Networks		Harris Interactive	
		Unweighted sample	Weighted sample	Unweighted sample	Weighted sample	Unweighted sample	Weighted sample
Postelection reports of electoral participation							
Usually voted in past elections?	Yes	78.7%	74.4%	76.5%	70.2%	90.8%	83.7%
	No	17.9%	21.0%	18.5%	22.4%	6.5%	13.3%
	Ineligible	3.2%	4.6%	5.0%	7.4%	2.7%	3.0%
	Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
	N	1206	1204	3408	3408	1040	1028
Voted in 2000 presidential election?	Yes	78.9%	76.5%	77.7%	72.2%	93.8%	90.9%
	No	21.1%	23.5%	22.3%	27.8%	6.3%	9.1%
	Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
	N	1206	1205	3408	3406	1040	1028

$b = .19$, $p < .001$).[5] And the KN respondents gave significantly fewer correct answers than the HI respondents (average percent correct answers given, unweighted: 58 percent versus 77 percent, $b = .16$, $p < .001$; weighted: 62 percent versus 70 percent, $b = .11$, $p < .001$ weighted). The same differences persisted after controlling for sample differences in demographics: the CSR respondents gave significantly fewer correct answers than the KN respondents ($b = .07, p < .001$ unweighted; $b = .07, p < .001$ weighted) and the HI respondents ($b = .18, p < .001$ unweighted; $b = .19, p < .001$ weighted), and the KN respondents gave significantly fewer correct answers than the HI respondents ($b = .11, p < .001$ unweighted; $b = .12, p < .001$ weighted).

Likewise, the rate at which respondents selected the midpoints of rating scales (thereby indicating neutrality in evaluations of politicians, national conditions, and government policies) was highest for the CSR respondents, a bit lower for the KN respondents, and considerably lower for the HI respondents. The CSR respondents manifested significantly more midpoint selections than the KN respondents (average percent of midpoint selections made, unweighted: 43.2 percent versus 39.4 percent, $b = -.04$, $p < .001$; weighted: 43.8 percent versus 39.5 percent, $b = -.04$, $p < .001$ weighted) and the HI respondents (unweighted: 43.2 percent versus 34.0 percent, $b = -.09, p < .001$; weighted: 43.8 percent versus 33.9 percent, $b = -.10$, $p < .001$). And the KN respondents manifested significantly more midpoint selections than HI respondents (unweighted: 39.4 percent versus 34.0 percent, $b = -.06, p < .001$; weighted: 39.5 percent versus 33.9 percent,; $b = -.06$, $p < .001$). The same differences persisted after controlling for sample differences in demographics: the CSR respondents manifested significantly more midpoint selections than the KN respondents ($b = -.04$, $p < .001$ unweighted; $b = -.03$, $p < .001$ weighted) and the HI respondents ($b = -.09, p < .001$ unweighted; $b = -.09, p < .001$ weighted), and the KN respondents manifested significantly more midpoint selections than the HI respondents ($b = -.05, p < .001$ unweighted; $b = -.06$, $p < .001$ weighted).

The CSR and KN samples contained comparable proportions of people who identified themselves as political independents (rather than identifying with a political party), whereas the proportion of independents in the HI sample was considerably lower. The KN and CSR respondents were not significantly different from one another (unweighted: 22.0 percent versus 21.8 percent, $p > .80$; weighted: 23.6 percent versus 23.3 percent, $p > .20$), whereas the HI respondents were significantly less likely to be independents than the CSR

5. The b coefficients in this paragraph test the statistical significance of the differences between the percents of correct quiz question answers given by the three survey firms' respondents. These coefficients are from ordinary least squares regressions (because the dependent variable is continuous) predicting the percent of quiz questions that a respondent answered correctly using two dummy variables representing the three survey firms. The same sorts of regressions were conducted to test differences between the firms in terms of the average number of midpoint selections made by the respondents.

respondents (unweighted: 13.1 percent versus 21.8 percent, $b = -.58, p < .001$; weighted: 13.6 percent versus 23.3 percent, $b = -.61$, $p < .001$ weighted) or the KN respondents (unweighted: 13.1 percent versus 22.0 percent, $b = -.59$, $p < .001$; weighted: 13.6 percent versus 23.6 percent, $b = -.63$, $p < .001$ weighted).[6] The same differences persisted after controlling for sample differences in demographics: a nonsignificant difference between the KN and CSR respondents ($p > .60$ unweighted; $p > .10$ weighted), whereas the HI respondents were significantly less likely to be independents than the CSR respondents ($b = -.20$, $p < .01$ unweighted; $b = -.22$, $p < .001$ weighted) or the KN respondents ($b = -.25$, $p < .01$ unweighted; $b = -.40$, $p < .001$ weighted).

The HI respondents were most likely to say pre-election that they intended to vote in the upcoming election, and the KN respondents were least likely to predict they would vote. The CSR respondents were more likely than the KN respondents to say they would vote (unweighted: 86.2 percent versus 81.5 percent, $b = -.35, p < .001$; weighted: 84.6 percent versus 78.5 percent, $b = -.40$, $p < .001$) and less likely than the HI respondents to predict they would vote (unweighted: 86.2 percent versus 94.8 percent, $b = 1.06, p < .001$; weighted: 84.6 percent versus 90.7 percent, $b = .58, p < .001$). The KN respondents were less likely than the HI respondents to predict they would vote (unweighted: 81.5 percent versus 94.8 percent, $b = 1.41$ $p < .001$; weighted: 78.5 percent versus 90.7 percent, $b = .98, p < .001$). The same differences persisted after controlling for sample differences in demographics: the CSR respondents were more likely than the KN respondents to predict they would vote ($b = -.60$, $p < .001$ unweighted; $b = -.63$, $p < .001$ weighted) and less likely than the HI respondents to predict they would ($b = .36$, $p < .05$ unweighted; $b = .31$, $p < .05$ weighted). The KN respondents were less likely than the HI respondents to predict they would vote ($b = .96$, $p < .001$ unweighted; $b = .94$, $p < .001$ weighted).

Postelection reports of voter turnout were about equal in the CSR and KN samples and considerably higher in the HI sample (see the bottom portion of table 4). The CSR and KN rates were not significantly different from one another unweighted (78.7 percent versus 76.5 percent, $p > .30$), but when the samples were weighted, the CSR respondents' reported turnout rate was significantly higher than that of the KN respondents (74.4 percent versus 70.2 percent, $b = -.28, p < .01$).[7] The CSR respondents reported significantly lower turnout

6. The b coefficients reported in this paragraph and in the next two paragraphs are from logistic regressions predicting dichotomous dependent variables (identification as a political independent and saying that one intended to vote or did vote). These coefficients test the significance of the differences between the firms in terms of the percentages of respondents selecting particular answers.

7. This result can be viewed as consistent with evidence to be reported later that telephone respondents are more likely than Internet respondents to distort their reports of attitudes and behavior in socially desirable directions.

than the HI respondents, both weighted and unweighted (unweighted: 78.7 percent versus 90.8 percent, $b = 1.39$, $p < .001$; weighted: 74.4 percent versus 83.7 percent, $b = 1.21$, $p < .001$). The KN respondents reported significantly lower turnout than the HI respondents (unweighted: 76.5 percent versus 90.8 percent, $b = 1.46$ $p < .001$; weighted: 70.2 percent versus 83.7 percent, $b = 1.35$, $p < .001$). After controlling for sample differences in demographics, the CSR respondents reported significantly higher turnout than the KN respondents ($b = -.26$, $p < .01$ unweighted; $b = -.38$, $p < .001$ weighted) and significantly lower turnout than the HI respondents ($b = .57$, $p < .05$ unweighted; $b = .56$, $p < .05$ weighted). The KN respondents reported significantly lower turnout than the HI respondents ($b = .83$, $p < .001$ unweighted; $b = .94$, $p < .001$ weighted).

CONCURRENT VALIDITY

Binary logistic regressions were conducted predicting vote choice (coded 1 for Mr. Gore and 0 for Mr. Bush) with a variety of predictors using only respondents who said they expected to vote for Mr. Bush or Mr. Gore.[8] All predictors were coded to range from 0 to 1, with higher numbers implying a more favorable orientation toward Mr. Gore.[9] Therefore, positively signed associations with predicted vote and actual vote were expected.

Concurrent validity varied substantially across the three houses (see table 5). As shown in the bottom row of table 5, the average change in probability that a respondent will vote for Gore instead of Bush based on the predictor measures in the CSR sample (unweighted: .47; weighted: .46) was weaker than the average change in probability for KN (unweighted: .56; weighted: .55), which in turn was weaker than the average change in probability for HI (unweighted: .63; weighted: .59). Concurrent validity was significantly lower for CSR than for KN for 22 of the 41 predictors, and concurrent validity was significantly lower for KN than for HI for 34 of the 41 predictors. Concurrent validity was significantly higher for CSR than for KN for none of the 41 predictors, and concurrent validity was significantly higher for KN than for HI for none of the predictors. Sign tests revealed significantly lower concurrent validity for CSR than for KN ($p < .001$), significantly lower concurrent validity for CSR than

8. A total of 26.8 percent of the CSR respondents, 27.3 percent of the KN respondents, and 13.5 percent of the HI respondents predicted that they would vote for someone other than Mr. Bush or Mr. Gore or said they would not predict for whom they would vote despite the follow-up leaning question. All regressions were run in STATA, which provides correct variance estimates from weighted analyses.

9. Respondents also reported their opinions on seven other policy issues, but the associations between opinions on these issues and vote choices were either zero or close to zero (logistic regression coefficients of .29 or less when the three samples were combined). Therefore, we focused our analyses on the issues that manifested substantial concurrent and predictive validity (logistic regression coefficients of 1.00 or more when the three samples were combined).

Table 5. Change in the Probability That the Respondent Will Vote for Mr. Gore instead of Mr. Bush (Pre-election Vote Choice) as the Result of Change from the Minimum to the Maximum Value of the Predictor

	Unweighted samples			Weighted samples		
Predictor	CSR	KN	HI	CSR	KN	HI
Clinton approval: Job	.73**	.85**	.88**	.71**	.84**	.88**
Clinton approval: Economy	.67**	.78**	.81**	.65**	.78**	.80**
Clinton approval: Foreign relations	.65**	.81**	.85**	.62**	.81**	.82**
Clinton approval: Crime	.54**	.79**	.87**	.56**	.79**	.85**
Clinton approval: Race relations	.61**	.80**	.84**	.58**	.81**	.86**
Clinton approval: Pollution	.46**	.78**	.85**	.47**	.78**	.86**
Past conditions: Economy	.50**	.67**	.73**	.48**	.68**	.71**
Past conditions: Foreign relations	.76**	.86**	.91**	.74**	.86**	.91**
Past conditions: Crime	.44**	.74**	.79**	.41**	.76**	.71**
Past conditions: Race relations	.45**	.84**	.87**	.42**	.83**	.81**
Past conditions: Pollution	.21**	.65**	.73**	.21**	.65**	.73**
Expectations: Economy	.52**	.53**	.51**	.54**	.52**	.48**
Expectations: Foreign relations	.44**	.45**	.40**	.48**	.44**	.36**
Expectations: Crime	.47**	.46**	.43**	.50**	.45**	.41**
Expectations: Race relations	.61**	.62**	.70**	.63**	.59**	.61**
Expectations: Pollution	.60**	.67**	.78**	.60**	.64**	.68**
Candidates' traits: Moral	.52**	.59**	.59**	.54**	.56**	.53**
Candidates' traits: Really cares	.54**	.60**	.68**	.57**	.59**	.63**
Candidates' traits: Intelligent	.52**	.59**	.67**	.55**	.55**	.57**
Candidates' traits: Strong leader	.30**	.32**	.29**	.37**	.31**	.29**
Evoked emotions: Angry	.56**	.56**	.59**	.58**	.53**	.52**
Evoked emotions: Hopeful	.41**	.51**	.54**	.45**	.50**	.50**
Evoked emotions: Afraid	.56**	.55**	.61**	.59**	.52**	.55**
Evoked emotions: Proud	.44**	.49**	.49**	.48**	.48**	.44**
Party identification	.94**	.91**	.95**	.91**	.91**	.91**
Political ideology	.76**	.93**	.94**	.74**	.91**	.90**
Military spending	.58**	.55**	.72**	.55**	.50**	.68**

Continued

Table 5. Continued

	Unweighted samples			Weighted samples		
Predictor	CSR	KN	HI	CSR	KN	HI
Welfare spending	.53**	.61**	.72**	.42**	.60**	.61**
Help for black Americans	.61**	.60**	.74**	.56**	.61**	.73**
Gun control	.58**	.52**	.63**	.55**	.52**	.64**
Pollution by businesses	.34**	.37**	.53**	.26**	.33**	.54**
Effort to control crime	.16*	.17**	.12*	.18**	.23**	.25**
Immigration restriction	.20**	.19**	.32**	.15*	.17**	.26**
Make abortion illegal	.31**	.34**	.48**	.31**	.29**	.40**
Make abortion legal	.40**	.36**	.51**	.42**	.32**	.45**
Help poor countries provide for people	.26**	.35**	.39**	.17**	.36**	.37**
Prevent people in other countries from killing each other	.27**	.29**	.45**	.24**	.19**	.40**
Prevent other governments from hurting their own citizens	.26**	.27**	.41**	.21**	.26**	.37**
Resolve disputes between other countries	.20**	.25**	.38**	.17*	.24**	.27**
Prevent other countries from polluting the environment	.21**	.37**	.50**	.18*	.36**	.43**
Build missile defense system	.31**	.38**	.52**	.29**	.32**	.42**
Average change in probability	.47	.56	.63	.46	.55	.59

* $p < .05$; ** $p < .01$.

for HI ($p < .001$), and significantly lower concurrent validity for KN than for HI ($p < .001$).[10]

Some of these differences between houses may be due to differences between the three samples in terms of demographics and political knowledge. To reassess the house effects after adjusting for those differences, we concatenated the data from the three houses into a single dataset and estimated the parameters of regression equations predicting anticipated vote choice with each substantive predictor (e.g., party identification), two dummy variables to represent

10. This sign test was computed by assigning a "+" to a predictor if one house had a stronger coefficient than the other and a "−" is assigned if the reverse was true and then computing the probability that the observed distributions of plusses and minuses occurred by chance alone.

the three houses, education, income, age, race, gender, political knowledge, political knowledge squared, and interactions of all of these latter variables with the substantive predictor. The interactions involving the demographics and knowledge allowed for the possibility that concurrent validity might vary according to such variables and might account partly for differences between the houses in observed concurrent validity. Our interest was in the two interactions of the house dummy variables with the substantive predictor; significant interactions would indicate reliable differences between houses in concurrent validity.

After controlling for demographics and political knowledge in concatenated regressions, sign tests again revealed significantly lower predictive validity for CSR than for KN ($p < .001$), significantly lower concurrent validity for CSR than for HI ($p < .001$), and significantly lower concurrent validity for KN than for HI ($p < .001$). Applying the sample weights weakened these differences a bit, but sign tests again revealed significantly lower concurrent validity for CSR than for KN ($p < .001$) and significantly lower concurrent validity for KN than for HI ($p < .05$), even when including the demographics and political knowledge and their interactions with the predictors in the equations.

PREDICTIVE VALIDITY

Table 6 shows change in probability estimates from equations predicting post-election vote choice with the 41 potential vote choice determinants. As shown in the bottom row of table 6, the average change in probability that a respondent will vote for Gore instead of Bush based on the predictor measures in the CSR sample (unweighted: .46; weighted: .45) was weaker than the average change in probability for KN (unweighted: .54; weighted: .53), which in turn was weaker than the average change in probability for HI (unweighted: .64; weighted: .57). Predictive validity was significantly lower for CSR than for KN for 24 of the 41 predictors, and predictive validity was significantly lower for KN than for HI for 32 of the 41 predictors. Predictive validity was significantly higher for CSR than for KN for none of the 41 predictors, and predictive validity was significantly higher for KN than for HI for none of the predictors. Sign tests revealed significantly lower predictive validity for CSR than for KN ($p < .001$), significantly lower predictive validity for CSR than for HI ($p < .001$), and significantly lower predictive validity for KN than for HI ($p < .001$).

After controlling for demographics and political knowledge in concatenated regressions, sign tests again revealed significantly lower predictive validity for CSR than for KN ($p < .05$), significantly lower predictive validity for CSR than for HI ($p < .001$), and significantly lower predictive validity for KN than for HI ($p < .001$). Applying the sample weights again weakened these differences, particularly the difference between KN and HI. Sign tests revealed

Table 6. Change in the Probability that the Respondent Voted for Mr. Gore Instead of Mr. Bush (Postelection Vote Choice) as the Result of Change from the Minimum to the Maximum Value of the Predictor

	Unweighted samples			Weighted samples		
Predictor	CSR	KN	HI	CSR	KN	HI
Clinton approval: Job	.77**	.87**	.93**	.77**	.88**	.86**
Clinton approval: Economy	.69**	.80**	.86**	.68**	.82**	.79**
Clinton approval: Foreign relations	.67**	.83**	.91**	.65**	.83**	.84**
Clinton approval: Crime	.58**	.80**	.92**	.62**	.78**	.85**
Clinton approval: Race relations	.61**	.81**	.88**	.59**	.78**	.85**
Clinton approval: Pollution	.44**	.78**	.89**	.42**	.76**	.81**
Past conditions: Economy	.50**	.70**	.74**	.47**	.72**	.70**
Past conditions: Foreign relations	.76**	.89**	.94**	.78**	.88**	.94**
Past conditions: Crime	.49**	.73**	.81**	.48**	.73**	.73**
Past conditions: Race relations	.44**	.84**	.93**	.48**	.82**	.94**
Past conditions: Pollution	.19**	.66**	.78**	.22**	.64**	.80**
Expectations: Economy	.47**	.45**	.46**	.45**	.44**	.46**
Expectations: Foreign relations	.40**	.39**	.35**	.40**	.37**	.35**
Expectations: Crime	.41**	.40**	.37**	.42**	.37**	.39**
Expectations: Race relations	.54**	.53**	.62**	.53**	.50**	.57**
Expectations: Pollution	.54**	.56**	.74**	.49**	.53**	.67**
Candidates' traits: Moral	.45**	.49**	.51**	.44**	.45**	.48**
Candidates' traits: Really cares	.47**	.49**	.57**	.47**	.46**	.57**
Candidates' traits: Intelligent	.45**	.50**	.59**	.43**	.45**	.54**
Candidates' traits: Strong leader	.28**	.31**	.25**	.30**	.28**	.27**
Evoked emotions: Angry	.49**	.47**	.49**	.48**	.44**	.54**
Evoked emotions: Hopeful	.37**	.40**	.41**	.37**	.38**	.44**
Evoked emotions: Afraid	.50**	.45**	.51**	.49**	.42**	.52**
Evoked emotions: Proud	.39**	.41**	.39**	.38**	.39**	.38**
Party identification	.90**	.91**	.96**	.88**	.90**	.94**
Political ideology	.81**	.94**	.96**	.79**	.91**	.96**
Military spending	.62**	.52**	.77**	.60**	.46**	.61**

Continued

Table 6. Continued

	Unweighted samples			Weighted samples		
Predictor	CSR	KN	HI	CSR	KN	HI
Welfare spending	.59**	.61**	.76**	.51**	.61**	.49**
Help for black Americans	.61**	.61**	.81**	.66**	.63**	.72**
Gun control	.61**	.59**	.71**	.59**	.61**	.61**
Pollution by businesses	.33**	.41**	.60**	.27*	.33**	.55**
Effort to control crime	.13*	.23**	.20**	.11	.3**	.32**
Immigration restriction	.21**	.19**	.33**	.15	.19**	−.01
Make abortion illegal	.39**	.37**	.56**	.39**	.32**	.48**
Make abortion legal	.41**	.36**	.61**	.40**	.33**	.55**
Help poor countries provide for people	.25**	.37**	.43**	.24**	.37**	.14**
Prevent people in other countries from killing each other	.31**	.32**	.51**	.31**	.31**	.36**
Prevent other governments from hurting their own citizens	.25**	.30**	.45**	.19**	.29**	.25*
Resolve disputes between other countries	.15*	.28**	.41**	.13*	.25**	.29**
Prevent other countries from polluting the environment	.22**	.42**	.57**	.17*	.43**	.51**
Build missile defense system	.35**	.36**	.55**	.38**	.32**	.30**
Average change in probability	.46	.54	.64	.45	.53	.57

*$p < .05$; **$p < .01$.

significantly lower predictive validity for CSR than for KN ($p < .001$), and marginally significantly lower predictive validity emerged for KN than for HI ($p < .10$).[11]

11. For both the pre-election and postelection surveys, the HI sample weights had an unconventionally wide range of values (from 0 to 26). As a result, variance estimates obtained from the weighted HI data were often much larger than those obtained from the other two samples, hence handicapping the ability to detect statistical significance of differences between HI data and the other two houses. The distribution of HI weights was examined for skewness and clumps. Although huge weights were assigned to some respondents, the majority of respondents received weights within the conventional range of less than 3. Furthermore, a sensitivity analysis on change in estimates before and after truncating the weights revealed little change in point estimates and variance estimates in the vote choice regression models presented in this paper. This is not surprising because the huge weights were assigned to very few respondents.

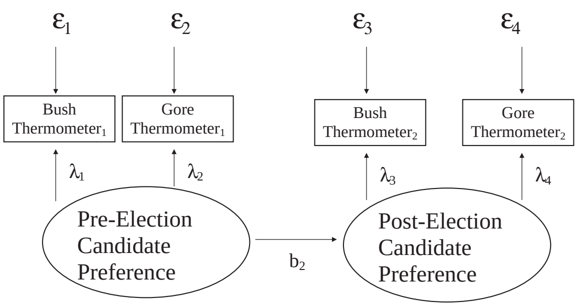

Figure 2. Structural Equation Model Used to Estimate Item Reliability.

SURVEY SATISFICING

The CSR respondents manifested more nondifferentiation than the KN respondents (unweighted: M = .40 versus .38, $b = -.02$, $p < .01$; weighted: M = .41 versus .38, $b = -.02$, $p < .001$), and the HI respondents manifested the least nondifferentiation (unweighted: M = .32, $b = -.06$, $p < .001$ compared with KN; weighted: M = .34, $b = -.05$, $p < .001$ compared with KN).[12] After controlling for differences between the samples in terms of demographics and political knowledge, the difference between KN and CSR was no longer statistically significant (unweighted $p > .20$; weighted $p > .50$), but HI continued to manifest the least nondifferentiation (unweighted: $b = -.04$, $p < .001$ compared with KN; weighted: $b = -.04$, $p < .001$ compared with KN).

RELIABILITY

To gauge the amount of random measurement error in answers using the pre-election and postelection feeling thermometer ratings of Mr. Bush and Mr. Gore, LISREL 8.14 was employed to estimate the parameters of the model shown in figure 2, which posited a latent candidate preference both pre-election and

12. To compute the nondifferentiation score for each respondent, we used the three pre-election feeling thermometer ratings and a formula developed by Mulligan et al. (2001):

$$\bar{x}_1 = \left(\frac{\sqrt{|therm1 - therm2|} + \sqrt{|therm1 - therm3|} + \sqrt{|therm2 - therm3|}}{3} \right).$$

Because thermometer ratings had been recoded to range from 0 to 1, scores on this index ranged from 0 to.804. A score of 0 indicated that all three thermometer ratings were identical, and a score of .804 indicated the highest level of observed differentiation among thermometer ratings. To yield an index where higher scores indicated more nondifferentiation, we subtracted .804 from each score and divided it by $-.804$, yielding a nondifferentiation index that ranged from 0 (indicating the least non-differentiation) to 1 (indicating the most differentiation).

Table 7. Structural Equation Model Parameter Estimates for Assessing Reliability

		Unweighted			Weighted		
Parameter	Indicator	CSR	KN	HI	CSR	KN	HI
Factor loadings	$Bush_1$	.61	.77	.85	.58	.74	.82
	$Gore_1$	−.73	−.78	−.86	−.69	−.74	−.80
	$Bush_2$	.72	.78	.88	.68	.78	.91
	$Gore_2$	−.79	−.80	−.86	−.76	−.79	−.83
Error variances	$Bush_1$	.62	.41	.29	.66	.46	.34
	$Gore_1$	.47	.40	.26	.53	.46	.36
	$Bush_2$	.48	.39	.23	.54	.39	.18
	$Gore_2$	.38	.37	.27	.43	.38	.32

postelection, measured by the feeling thermometer ratings. The stability of the latent construct is represented by a structural parameter, b_{21}. $\varepsilon_1 - \varepsilon_4$ represent measurement error in each indicator, and $\lambda_1 - \lambda_4$ are loadings of the manifest indicators on the latent factors. The larger $\lambda_1 - \lambda_4$ are, the higher the validities of the indicators; the smaller $\varepsilon_1 - \varepsilon_4$ are, the higher the reliabilities of the items are.

The parameters of the model were estimated separately for CSR, KN, and HI three times, first unweighted, then weighted using the weights supplied by the survey firms, and finally weighted using a set of weights we built to equate the samples in terms of demographics and political knowledge. Specifically, we weighted each sample to match the age, gender, education, and race benchmarks from the 2000 CPS March Supplement and to match the average political knowledge scores from all three samples combined.[13]

Consistently across all four indicators, the factor loadings were smallest for CSR, intermediate for KN, and largest for HI (see table 7). The error variances were consistently the largest for CSR, intermediate for KN, and smallest for HI. All of the differences between adjacent columns in table 7 are statistically significant ($p < .05$). Thus, these results are consistent with the conclusion that the CSR reports were less reliable than the KN reports, which in turn were less reliable than the HI reports.

SOCIAL DESIRABILITY RESPONSE BIAS

Among White respondents, it is socially undesirable to express opposition to government programs to help Black Americans (see Holbrook, Green, and Krosnick 2003). When asked whether the federal government should provide

13. This weighting was also done using income as well, and the results were comparable to those described in the text.

more, less, or the same amount of help for African Americans, the distributions of answers from White respondents differed significantly across the three houses. White KN respondents were more likely than White CSR respondents to say the government should provide less help to Black Americans (unweighted: CSR = 17.0 percent versus KN = 35.8 percent, $\chi^2 = 188.87, p < .001$; weighted: CSR = 16.1 percent versus KN = 34.1 percent, $\chi^2 = 189.41$, $p < .001$). And White HI respondents were more likely than White KN respondents to say the government should provide less help to Black Americans (unweighted: KN = 35.8 percent versus HI = 42.5 percent, $\chi^2 = 30.98$, $p < .001$; weighted: KN = 34.1 percent versus HI = 34.1 percent, $\chi^2 = 13.90$, $p < .001$). The same differences persisted when controlling for demographics and political knowledge: White CSR respondents gave significantly fewer socially undesirable answers than White KN respondents ($b = .88, p < .001$) and White HI respondents ($b = 1.02, p < .001$). And White KN respondents gave significantly fewer socially undesirable answers than White HI respondents ($b = .13, p < .05$).[14]

We also tested whether these differences persisted when controlling for vote choice in the 2000 Presidential election, party identification, and political ideology. The HI sample was more pro-Republican and more politically conservative than the other samples, so this may have been responsible for the HI sample's greater opposition to government help to Black Americans. And in fact, controlling for these additional variables made the difference in answers to the aid to Blacks question between White KN and HI respondents nonsignificant ($b = .10, p > .10$). However, even with these controls, White CSR respondents gave significantly fewer socially undesirable answers than did White KN respondents ($b = 1.00, p < .001$) and White HI respondents ($b = 1.11, p < .001$). Thus, the mode difference persisted.

PAST EXPERIENCE AND SELECTIVITY

The KN and HI data may have manifested higher response quality than the telephone data partly because the Internet respondents were panel members who had more practice doing surveys than the average telephone respondent. To test this notion, KN provided the number of invitations sent to each respondent and the number of surveys each respondent completed during the 3 months prior to our pre-election survey. HI provided the number of invitations sent to each respondent and the number of surveys each respondent ever completed.

We computed two variables: (a) "past experience," number of completed surveys in the past (recoded to range from 0 to 1 in both samples), and

14. These logistic regressions predicted socially undesirable responding (coded 1 = "less help for Black Americans" and 0 = "same" or "more help for Black Americans") with two dummy variables representing the three survey firms and main effects of education, income, age, gender, race, political knowledge, and political knowledge squared.

(b) "selectivity," the rate of responding to past invitations, which was the number of completions divided by number of invitations (also recoded to range from 0 to 1).[15]

To assess whether past experience or selectivity affected response quality, we repeated the binary logistic regressions predicting vote choice using each of the 41 predictors, controlling for the main effects of past experience and selectivity and the interactions between these two variables with each predictor. If having more experience with surveys improved response quality, a significant positive interaction between past experience and each predictor should appear. If being more selective about survey participation results in higher response quality on the surveys that a person completes, a significant negative interaction between selectivity and each predictor should appear.

These data uncovered many indications that past experience improved survey performance in the KN data. Past experience interacted positively with 37 of 41 predictors in the concurrent validity equations, meaning that concurrent validity was higher for people who had more past experience. Eleven of these interactions were significant ($p < .05$), and none of the interactions in the opposite direction were significant. In the predictive validity equation, past experience was positively associated with predictive validity for 33 of the 41 predictors in the KN data. Six of these effects were significant, and none of the interactions in the opposite direction were significant.

In contrast, the HI data showed very little evidence of practice effects. Past experience interacted positively with 23 of 41 predictors in the concurrent validity equations, just about the number that would be expected by chance alone. Only three of these interactions were significant, and none of those in the opposite direction were significant. In the predictive validity equations, 24 of the 41 predictors yielded positive interactions, only 3 of which were statistically significant, and none of the past experience effects in the opposite direction were significant. The absence of practice effects in the HI data may be because the range of practice in that sample was relatively small as compared to the KN sample.

Selectivity in past participation did not appear to be a reliable predictor of response quality in the KN sample. Selectivity interacted negatively with 15 of 41 predictors in the concurrent validity assessments (fewer than would be expected by chance), and none of the interactions was significant. Similarly, selectivity

15. Ten percent of HI respondents had never completed any HI survey before the pre-election survey in the present study, whereas only 0.3 percent of KN respondents had never completed any KN survey prior to ours. So the KN respondents were a bit more experienced with the survey platform than were the HI respondents. About 54 percent of KN respondents had completed all the surveys that KN had invited them to do during the prior three months, whereas only 2 percent of the HI respondents had a perfect completion rate since joining the HPOL panel. Thus, the HI respondents were apparently more selective than were the KN respondents, who were obligated to complete all surveys in order to keep their free WebTV equipment.

interacted negatively with predictive validity for 12 of the 41 predictors in the KN data, and none of these interactions was significant.

In contrast, selectivity was associated with improved response quality in the HI sample. Selectivity interacted negatively with 33 of 41 predictors in the concurrent validity equations; 15 of these interactions were significant, and none were significant in the opposite direction. In the predictive validity equations, 35 of the 41 predictors manifested negative interactions, 10 of which were significant, and none of the interactions in the opposite direction were significant.

All this suggests that at least some superiority in response quality of the KN sample over the CSR sample may be attributable to practice effects, and some of the superiority in response quality of the HI sample over KN sample may be due to strategic selectivity.

Discussion

These data support a series of conclusions:

(1) The probability samples were more representative of the nation's population than was the nonprobability sample, even after weighting.
(2) The nonprobability sample was biased toward individuals who were highly knowledgeable about and interested in the topic of the survey.
(3) Self-reports provided via the Internet were more accurate descriptions of the respondents than were self-reports provided via telephone, as manifested by higher concurrent and predictive validity, higher reliability, less satisficing, and less social desirability bias.
(4) The practice gained by participants in the KN panel enhanced the accuracy of their self-reports, but such practice did not enhance the accuracy of reports by members of the nonprobability Internet sample.
(5) The tendency of nonprobability sample members to choose to participate in surveys on topics of great interest to them made their self-reports more accurate on average than the self-reports obtained from the less selective KN respondents.

Our findings that practice effects enhance the quality of survey responses (and therefore advantage probability sample Internet surveys) are in harmony with the large literature in psychology showing that practice improves performance on complex tasks (e.g., Donovan and Radosevich 1999). And our findings are in line with other evidence suggesting that survey respondents provide more accurate reports after gaining practice by completing questionnaires (e.g., Novotny et al. 2001).

Although the response rate for the KN sample (25 percent) was considerably lower than the response rate for the CSR sample (43 percent), the average demographic representativeness of the KN sample was equal to that of the

CSR sample. This evidence is consistent with past findings suggesting that declines in response rates were not associated with notable declines in sample representativeness (Curtin, Presser, and Singer 2000; Keeter et al. 2000).

Conclusion

The results from the national field experiment suggest that the Internet offers a viable means of survey data collection and has advantages over telephone interviewing in terms of response quality. These results also demonstrate that probability samples yield more representative results than do nonprobability samples. We look forward to future studies comparing data quality across these modes to complement the evidence reported here and to assess the generalizability of our findings.

Supplementary Data

Supplementary data are available online at http://poq.oxfordjournals.org/

References

Acree, Michael, Maria Ekstrand, Thomas J. Coates, and Ron Stall. 1999. "Mode Effects in Surveys of Gay Men: A Within-Individual Comparison of Responses by Mail and by Telephone." *Journal of Sex Research* 36:67–75.

Aronson, Elliot, Phoebe C. Ellsworth, J. Merrill Carlsmith, and Marti Hope Gonzales. 1990. *Methods of Research in Social Psychology*. New York: McGraw-Hill.

Berrens, Robert P., Alok K. Bohara, Hank Jenkins-Smith, Carol Silva, and David L. Weimer. 2003. "The Advent of Internet Surveys for Political Research: A Comparison of Telephone and Internet Samples." *Political Analysis* 11:1–22.

Best, Samuel J., Brian Krueger, Clark Hubbard, and Andrew Smith. 2001. "An Assessment of the Generalizability of Internet Surveys." *Social Science Computer Review* 19:131–45.

Bracht, Glenn H., and Gene V. Glass. 1968. "The External Validity of Experiments" *American Educational Research Journal* 5:437–74.

Bridge, R. Gary, Leo G. Reeder, David Kanouse, Donald R. Kinder, Vivian T. Nagy, and Charles Judd. 1977. "Interviewing Changes Attitudes—Sometimes." *Public Opinion Quarterly* 41:57–64.

Cannell, Charles F., Peter V. Miller, and Lois Oksenberg. 1981. "Research on Interviewing Techniques." *Sociological Methodology* 12:389–437.

Chang, LinChiat. 2001. *A Comparison of Samples and Response Quality Obtained from RDD Telephone Survey Methodology and Internet Survey Methodology*. Doctoral Dissertation, Ohio State University.

Chartrand, Tanya L., and John A. Bargh. 1999. "The Chameleon Effect: The Perception-Behavior Link and Social Interaction." *Journal of Personality and Social Psychology* 76:893–910.

Clausen, Aage R. 1968. "Response Validity: Vote Report." *Public Opinion Quarterly* 32:588–606.

Clinton, Joshua D. 2001. *Panel Bias from Attrition and Conditioning: A Case Study of the Knowledge Networks Panel*. Unpublished manuscript, Stanford University.

Cordell, Warren N., and Henry A. Rahmel. 1962. "Are Nielsen Ratings Affected by Noncooperation, Conditioning, or Response Error?" *Journal of Advertising Research* 2:45–49.

Couper, Mick P. 2000. "Web Surveys: A Review of Issues and Approaches." *Public Opinion Quarterly* 64:464–94.

Curtin, Richard, Stanley Presser, and Eleanor Singer. 2000. "The Effects of Response Rate Changes on the Index of Consumer Sentiment." *Public Opinion Quarterly* 64:413–28.

de Leeuw, Edith D., and Martin Collins. 1997. "Data Collection Methods and Survey Quality: An Overview." In *Survey Measurement and Process Quality*, eds. Lars E. Lyberg, Paul Biemer, Martin Collins, Edith de Leeuw, Cathryn Dippo, Norbert Schwarz, and Dennis Trewin. New York: Wiley.

Dillman, Don A. 1978. *Mail and Telephone Surveys: The Total Design Method*. New York: Wiley.

Dillman, Don A. 2000. *Mail and Internet Surveys: The Tailored Design Method*. New York: Wiley.

Donovan, John J., and David J. Radosevich. 1999. "A Meta-Analytic Review of the Distribution of Practice Effect: Now You See It, Now You Don't." *Journal of Applied Psychology* 84:795–805.

Falaris, Evangelos M., and H. Elizabeth Peters. 1998. "Survey Attrition and Schooling Choices." *The Journal of Human Resources* 33:531–54.

Fitzgerald, John, Peter Gottschalk, and Robert Moffitt. 1998a. "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics." NBER Technical Working Papers, National Bureau of Economic Research, Inc.

———. 1998b. "An Analysis of the Impact of Sample Attrition on the Second Generation of Respondents in the Michigan Panel Study of Income Dynamics." *The Journal of Human Resources* 33:300–344.

Fitzsimons, Gavan J., and Vicki Morwitz. 1996. "The Effect of Measuring Intent on Brand Level Purchase Behavior." *Journal of Consumer Research* 23:1–11.

Flemming, Greg, and Molly Sonner. 1999. "Can Internet Polling Work? Strategies for Conducting Public Opinion Surveys Online." Paper presented at the annual meeting of the American Association for Public Opinion Research, St. Petersburg Beach, FL, USA.

Fowler, Floyd Jackson, Anthony M. Roman, and Zhu Xiao Di. 1998. "Mode Effects in a Survey of Medicare Prostate Surgery Patients." *Public Opinion Quarterly* 62:29–46.

Granberg, Donald, and Soren Holmberg. 1991. "Self-Reported Turnout and Voter Validation." *American Journal of Political Science* 35:448–59.

Greenwald, Anthony G., Catherine G. Carnot, Rebecca Beach, and Barbara Young. 1987. "Increasing Voting Behavior by Asking People if They Expect to Vote." *Journal of Applied Psychology* 72:315–8.

Groves, Robert M., and Robert L. Kahn. 1979. *Surveys by Telephone: A National Comparison with Personal Interviews*. New York: Academic.

Groves, R. M., E. Singer, and A. D. Corning. 2000. "Leverage-Saliency Theory of Survey Participation: Description and an Illustration." *Public Opinion Quarterly* 64:299–308.

Himmelfarb, Samuel, and Fran H. Norris. 1987. "An Examination of Testing Effects in a Panel Study of Older Persons." *Personality and Social Psychology Bulletin* 13:188–209.

Holbrook, Allyson L., Melanie C. Green, and Jon A. Krosnick. 2003. "Telephone versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias." *Public Opinion Quarterly* 67:79–125.

Holbrook, Allyson L., Jon A. Krosnick, and Alison M. Pfent. 2007. "Response Rates in Surveys by the News Media and Government Contractor Survey Research Firms." In *Telephone Survey Methodology*, eds. J. Lepkowski, B. Harris-Kojetin, P. J. Lavrakas, C. Tucker, E. de Leeuw, M. Link, M. Brick, L. Japec, and R. Sangster. New York: Wiley.

Jagodzinski, Wolfgang, Steffen M. Kuhnel, and Peter Schmidt. 1987. "Is There a "Socratic Effect" in Nonexperimental Panel Studies? Consistency of an Attitude toward Guestworkers." *Sociological Methods and Research* 15:259–302.

Keeter, Scott, Carolyn Miller, Andrew Kohut, Robert M. Goves, and Stanley Presser. 2000. "Consequences of Reducing Nonresponse in a National Telephone Survey." *Public Opinion Quarterly* 64:125–48.

Kenny, David A. 1979. *Correlation and Causality*. New York: Wiley.

Kiecker, Pamela, and James E. Nelson. 1996. "Do Interviewers Follow Telephone Survey Instructions?" *Journal of the Market Research Society* 38:161–76.

Krosnick, Jon A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5:213–36.

———. 1999. "Survey Methodology." *Annual Review of Psychology* 50:537–67.

Lavrakas, Paul J. 1993. *Telephone Survey Methods: Sampling, Selection, and Supervision*. Thousand Oaks, CA: Sage.

———. 1997. "Methods for Sampling and Interviewing in Telephone Surveys." In *Handbook of Applied Social Research Methods*, eds. Bickman Leonard, and Debra J. Rog. Thousand Oaks, CA: Sage.

Leary, Mark R. 1995. *Behavioral Research Methods*. Pacific Grove, CA: Brooks/Cole.

Lerner, Jennifer S., Julie H. Goldberg, and Philip E. Tetlock. 1998. "Sober Second Thought: The Effects of Accountability, Anger, and Authoritarianism on Attributions of Responsibility." *Personality and Social Psychology Bulletin* 24:563–74.

Lerner, Jennifer S., and Philip E. Tetlock. 1999. "Accounting for the Effects of Accountability." *Psychological Bulletin* 125:255–75.

Lubin, Bernard, Eugene E. Levitt, and Marvin Zuckerman. 1962. "Some Personality Differences between Responders and Nonresponders to a Survey Questionnaire." *Journal of Consulting Psychology* 26:192.

Lyberg, Lars, and Daniel Kasprzyk. 1991. "Data Collection Methods and Measurement Error: An Overview." In *Measurement Errors in Surveys*, eds. Paul Biemer, Robert M. Groves, Lars E. Lyberg, Nancy Mathiowetz, and Seymour Sudman. New York: Wiley.

Mann, Christopher B. 2005. "Unintentional Voter Mobilization: Does Participation in Pre-election Surveys Increase Voter Turnout?" *Annals of the American Academy of Political and Social Science* 601:155–68.

Menard, Scott. 1991. *Longitudinal Research*. Newbury Park, CA: Sage.

Miller, Jane S., and Shelly Haas 2003. *Council for Marketing and Opinion Research 2003 Respondent Cooperation and Industry Image Survey*. Cincinnati, OH: CMOR.

Mulligan, Ken, Jon A. Krosnick, Wendy Smith, Melanie Green, and George Bizer. 2001. *Non-differentiation on Attitude Rating Scales: A Test of Survey Satisficing Theory*. Unpublished Manuscript, Stanford University.

Neumann, Roland, and Fritz Strack. 2000. "Mood Contagion: The Automatic Transfer of Mood Between Persons." *Journal of Personality and Social Psychology* 79:211–23.

Novotny, Janet A., William V. Rumpler, Joseph T. Judd, Howard Riddick, Donna Rhodes, Margaret McDowell, and Ronette Briefel. 2001. "Diet Interviews of Subject Pairs: How Different Persons Recall Eating the Same Foods." *Journal of the American Dietetic Association* 101:1189–93.

Price, Kenneth H. 1987. "Decision Responsibility, Task Responsibility, Identifiability, and Social Loafing." *Organizational Behavior and Human Decision Processes* 40:330–45.

Rohde, Gregory L., and Robert Shapiro. 2000. *Falling Through the Net: Toward Digital Inclusion*. Washington, DC: U.S. Department of Commerce, Economics and Statistics Administration and National Telecommunications and Information Administration.

Rossi, Peter H., James D. Wright, and Andy B. Anderson. 1983. *Handbook of Survey Research*. Orlando, FL: Academic.

Sherman, Steven J. 1980. "On the Self-Erasing Nature of Errors of Prediction." *Journal of Personality and Social Psychology* 39:211–21.

Smith, Eliot R., Nyla Branscombe, and Carol Bormann. 1988. "Generality of the Effects of Practice on Social Judgement Tasks." *Journal of Personality and Social Psychology* 54:385–95.

Sobol, Marion G. 1959. "Panel Mortality and Panel Bias." *Journal of the American Statistical Association* 54(285):52–68.

Tourangeau, Roger, and Tom W. Smith. 1996. "Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context." *Public Opinion Quarterly* 60:275–304.

Van Der Zouwen, Johannes, Wil Dijkstra, and Johannes H. Smit. 1991. "Studying Respondent–Interviewer Interaction: The Relationship between Interviewing Style, Interviewer Behavior, and Response Behavior." In *Measurement Errors in Surveys*, eds. Paul Biemer, Robert M. Groves, Lars E. Lyberg, Nancy Mathiowetz, and Seymour Sudman. New York: Wiley.

Willson, Victor L., and Richard R. Putnam. 1982. "A Meta-Analysis of Pretest Sensitization Effects in Experimental Design." *American Educational Research Journal* 19:249–58.

Wright, Debra L., William S. Aquilino, and Andrew J. Supple. 1998. "A Comparison of Computer-Assisted Paper-and-Pencil Self-Administered Questionnaires in a Survey on Smoking, Alcohol, and Drug Use." *Public Opinion Quarterly* 62:331–53.

Zagorsky, Jay, and Pat Rhoton. 1999. *Attrition and the National Longitudinal Survey's Women Cohorts*. Columbus, OH: Center for Human Resource Research, Ohio State University.