Advance Access publication May 28, 2014

Political Analysis (2014) 22:304–320
doi:10.1093/pan/mpt049

Nonresponse and Mode Effects in Self- and Interviewer-Administered Surveys

Lonna Rae Atkeson
Department of Political Science, University of New Mexico, MSC05-3070, 1 University of New Mexico, Albuquerque, NM 87131-0001
e-mail: atkeson@unm.edu (corresponding author)

Alex N. Adams
Department of Political Science, University of New Mexico, MSC05-3070, 1 University of New Mexico, Albuquerque, NM 87131-0001
email: alexnadams@yahoo.com

R. Michael Alvarez
California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125
email: rma@hss.caltech.edu

Edited by Jasjeet Sekhon

We examine the quality of two probability-based polls, one interviewer administered (telephone) and one self-administered (Internet and mail mixed mode survey). The polls use the same sampling frame (registered voters) and the same questions. First, we examine the representativeness of both surveys using information known about the population, and although we find important differences between the two in terms of sampling and nonresponse bias, we also find that both surveys represent the underlying population despite low response rates. We also test for mode effects between surveys due to social desirability and how it influences nondifferentiation or satisficing. Using a variety of methods (*t*-tests, multivariate regression, and genetic propensity matching), we find evidence that the presence of an interviewer alters response patterns on ego-driven questions. The implications of our work are important, due to the increasing popularity of mixed mode surveys. Researchers need to be methodologically sensitive to these differences when analyzing surveys that allow for different response modes.

1 Introduction

Public opinion research is at a crossroads. Changes in survey methodology over the last fifteen years have transformed how researchers interact and collect data from respondents. The primary survey modes over the last half of the twentieth century were interviewer driven, in the form of in-person or telephone interviews. However, these methods have increasingly become problematic: gated communities and locked apartment buildings have made in-person interviewing difficult, and the rise of mobile phones and the associated disconnection of landlines have reduced coverage for phone surveys (Rivers ND; Carley-Baxter, Peytchev, and Black 2010; Blumberg and Luke 2013). Since the late 1990s, however, surveys have increasingly become more respondent driven with the rise of address-based sampling, Interactive Voice Response (IVR), and the Internet.

In addition, over the last fifteen years there has been an increase in the use of multiple modes to achieve greater-quality data and to compensate for survey error problems that plague each survey methodology (Dillman, Smyth, and Christian 2009). However, there is limited research on how different modes systematically influence response patterns, especially with political attitudes and behavior (though see Chang and Krosnick 2009, 2010). Importantly, our work extends previous

Authors' note: Supplementary materials for this article are available on the *Political Analysis* Web site.

© The Author 2014. Published by Oxford University Press on behalf of the Society for Political Methodology.
All rights reserved. For Permissions, please email: journals.permissions@oup.com

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt049 Published online by Cambridge University Press

research, as we use matching techniques to remove noncomparability between modes due to differences in sample composition. Our research also draws attention to this issue at a time when social scientists are beginning to combine survey modes more frequently. In this way, our research fills a growing methodological need in the discipline to be attentive to changes in methodology that might influence the quality of the data we generate and use in hypothesis-driven research. We examine the quality of the data by focusing on observable characteristics of the sample and selection to reduce or eliminate differences due to sample composition and survey nonresponse.

2 The Rapidly Changing Nature of Survey Research

Survey methodologists have long known of the trade-off between survey cost and survey error (Groves 1989). For many decades face-to-face (FTF) interviewing was the norm. In the 1970s, FTF interviewing gave way to telephone interviewing, which at the time provided adequate coverage and reduced data collection costs substantially. In the 1990s new technologies such as mobile phones, caller ID, and answering machines made traditional telephone interviewing less cost-effective because of high refusal rates and high no-answer rates and created new concerns about coverage. At the same time, increasing use of the Internet and other new technologies such as IVR created new opportunities for survey researchers, but many of the same problems existed. For example, there are problems with coverage and the lack of sampling frames in the case of the Internet as well as differences in hardware and software that alter the survey environment for different respondents. In the early 2000s, address-based sampling became possible as the USPS mail list became commercially available and created the largest database with near universal coverage of residential homes (Iannacchione 2011). Address-based sampling has become a critical component of many studies, including the American National Election Study and the General Social Survey (Davis, Smith, and Mardsen 2009; Lupia et al. 2009; O'Muircheartaigh et al. 2009; DeBell et al. 2011; Iannacchione 2011). These changes have resulted in a methodological shift from a survey environment that was dominated by a personal interaction between the interviewer and the respondent to a method that is highly impersonal and respondent driven (Dillman 2000).

Because different populations are likely attracted to different modes based upon familiarity and accessibility, survey researchers are increasingly using mixed modes to boost response rates and ensure adequate coverage (Dillman 2000; de Leeuw 2005). Mixed mode surveys use multiple methods to contact or receive information from respondents. Thus, mixed mode surveys involve both mixtures of data collection and communications with the respondent. For example, a mixed mode survey might contact sample members by phone or mail and then have them respond to a questionnaire over the Internet. Alternatively, a mixed mode survey might allow for multiple forms of response. For example, sample frame members may be able to complete the interview over the phone, by mail, or on the web. Thus, a respondent who does not respond over the Internet may in subsequent contact receive a phone call or a FTF visit or may be offered a choice of response mode on the initial contact.

A major reason for the rising interest in mixed mode surveys is the underlying concern that coverage in any survey mode may be problematic and therefore using multiple modes may compensate for problems that plague the use of any particular mode. The most recent report by the National Health Interview Survey (NHIS) estimates that nearly 38% of US households are mobile phone only and another 16% receive all or almost all of their calls on mobile phones even when they have a landline (Blumberg and Luke 2013).[1] This suggests that nearly half of all households are difficult or impossible to reach using a traditional RDD or landline sampling frame. The problem of coverage is exacerbated because mobile phone only households are not equally represented throughout the population: Younger adults, non-whites, renters, and poorer adults are much more likely to live in mobile phone only homes (Blumberg and Luke 2013). Indeed, over three in five adults ages 25–29 years, more than half of all renters, and over half of Hispanics live in mobile phone only households (Blumberg and Luke 2013, 2–3). With about half of the population relying, or mostly relying, on

[1] In early 2009, just after our surveys were in the field, these numbers were 20% wireless only households and another 14.5% receiving almost all of their calls on a mobile phone (Blumberg and Luke 2009).

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt049 Published online by Cambridge University Press

mobile phones, coverage issues for surveys that use landline-based methods may be a serious threat to one of the main goals of survey research—accurate population inference.

Mixed mode surveys, therefore, are a potential solution to this problem. They provide a means for reaching different subgroups of sample members (Groves and Kahn 1979; Day et al. 1995; Shettle and Mooney 1999) and as such they have been increasing in popularity. According to Dillman, Smyth, and Christian (2009, 11), one government administrator noted that the change in the survey environment means that "We are trying to give respondents what they want, but still do valid surveys. That means giving people a choice." For example, the American Community Survey first contacts potential respondents by mail; those who do not respond receive a telephone survey, and if that fails they attempt a FTF interview with a subsample of remaining nonrespondents (Alexander and Wetrogen 2000). Other government agencies, including the Bureau of Labor Statistics, with the Current Employment Statistics Survey, and the Centers for Disease Control, with the National Survey of Family Growth, utilize mixed mode surveys. Furthermore, many academic surveys (i.e., the American National Election Survey and the General Social Survey) are turning to mixed mode techniques (Iannacchione 2011). The 2012 ANES time series study includes both FTF surveys and online panelists.[2]

The underlying question inherent in using mixed mode survey methods is whether the data are reliable and consistent across survey modes. Specifically, are the observational errors associated with different modes such that they must be controlled for when using the data to answer substantive questions of interest? Our research seeks to address this question by examining the differences in nonresponse bias and mode effects between a self-administered questionnaire (SAQ) and an interviewer-administered questionnaire (IAQ). We select on observable characteristics of the sample to determine the extent and types of differences between survey modes.

2.1 *Nonresponse Bias*

Over the past two decades, response rates have been rapidly decreasing for government surveys (de Leeuw and de Heer 2002), academic surveys (Hox and de Leeuw 1994), and especially for market researchers (Balden 2004; Stoop 2005). With rising nonresponse rates have come concerns about nonresponse error. Although in all survey modes nonresponse error is a concern, there is hope that mixed mode surveys will reduce nonresponse error and increase response rates (de Leeuw 2005; Dillman 2000). Nonresponse error results when certain groups self-select out of the study, potentially creating a biased survey; for example, Burden (2000) argues that in the ANES "worsening presidential turnout estimates are the result mostly of declining response rates . . ." (389).[3] Mixed mode surveys may reduce nonresponse error if groups of respondents who may not have either the motivation or the ability to respond do so when provided with multiple or the right response options. For example, precontact with sample frame members in the form of an advance letter describing the study that creates legitimacy and trust between the respondent and the survey researcher has been shown to increase response rates with follow-up phone calls (de Leeuw et al. 2004).

Nonresponse error is a valid concern. Significant differences between respondents and nonrespondents on a sample statistic like the mean can lead to problems in survey inference (Peterson and Kerin 1981). Response bias may be especially important in mixed mode surveys where particular selection modes may over- or underrepresent some groups relative to others. Therefore, assessing how designs influence sample characteristics is an important consideration particularly in the current survey environment where it is unclear what the population of telephone or Internet users might look like. This may be especially true for mixed mode designs where different methods of interviewing may attract particular types of respondents. Internet users tend to be younger, whiter, and more male, and thus a design that relies heavily on the Internet may under-represent important subgroups in the population of interest (Zickur and Smith 2012).

[2]See ANES 2012 Time Series data page: http://www.electionstudies.org/studypages/anes_timeseries_2012/anes_timeseries_2012.htm (accessed January 11, 2013).
[3]See Martinez (2003), McDonald (2003), and Burden (2003) for further discussion of this issue.

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt049 Published online by Cambridge University Press

Likewise, mail surveys may attract older respondents (Atkeson and Tafoya 2008; Atkeson and Adams 2010).

Importantly, low response rates, which are nearly ubiquitous at this point, are not necessarily suggestive of problems of nonresponse and therefore are not a good measure of sample representativeness (Curtin, Presser, and Singer 2000; Keeter et al. 2000; Merkle and Edelman 2002; Groves, Presser, and Dipko 2004; Pew Research Center for the People and the Press 2004; Groves and Peytcheva 2008; Atkeson et al. 2011). This is because nonresponse error is not necessarily a function of the response rate, but is a function of the response rate interacting with the characteristics of the nonrespondent (Curtin, Presser, and Singer 2000). Thus, if survey topic salience, for example, leads to response bias, then nonresponse error becomes a problem (Groves, Singer, and Corning 2000). But low response rates, in general, may not be indicative of a core problem; indeed, they only offer indirect evidence of a potential problem (Dillman 2000). In several studies, tests of representativeness have showed little difference between surveys with high and low response rates (Curtin, Presser, and Singer 2000; Keeter et al. 2000; Merkle and Edelman 2002; Groves, Presser, and Dipko 2004; Pew Research Center for the People and the Press 2004; Groves and Peytcheva 2008; Atkeson et al. 2011).

In our study, the sample frames are the same, therefore any systematic difference between the samples must be due to differences in sampling bias and survey nonresponse. Moreover, because we know many observable characteristics of our population, we can test the representativeness of the SAQ and IAQ to determine how well each design represents the underlying population. In general, we hypothesize that differences will exist in demographic characteristics between survey modes due to the different survey formats and the digital divide. Given our knowledge of the digital divide and differences in households with connected landlines, we expect that interviewer-driven phone respondents will be older, less educated, and more likely to be minority than in the respondent-driven Internet/mail mixed mode survey.

2.2 *Mode Effects*

The presence or absence of an interviewer produces differences in how questions are cognitively considered (auditory or visually) and may affect responses. Specifically, evidence exists that the presence or absence of an interviewer influences answers to questions even when questions are identically worded (de Leeuw and Van Der Zouwen 1988; Dillman et al. 1996; Fowler, Roman, and Di 1998; Dillman and Christian 2005; Chang and Krosnick 2009, 2010). If we imagine that the survey process is similar to a conversation (Schwarz 1996), then the context provided by the survey either through the interviewer or through the presentation of question and answer scales may affect question interpretation and response. Importantly, survey mode likely influences respondent answers, potentially threatening the reliability of the data between survey modes (Campbell 1950; Peterson and Kerin 1981; Mensch and Kandel 1988).

Although there are many possibilities for differences in response associated with an SAQ versus IAQ (Schuman and Presser 1981; Schuman 1992; Sudman, Bradburn, and Schwarz 1996; Fowler, Roman, and Di 1998; Christian and Dillman 2004; Tourangeau, Couper, and Conrad 2004; Smyth et al. 2006; Chang and Krosnick 2009, 2010), social desirability effects in surveys where an interviewer is present may be especially strong and troubling (London and Williams 1990; Aquilino 1994). Research shows that SAQs result in less socially desirable responses than IAQs across a variety of issues.

Theoretically, the presence of an interviewer raises the concern of the respondent that his or her answers may be met with disapproval, leading respondents to provide more socially favorable and biased responses. Social desirability appears in the form of the overreporting of good behaviors and the underreporting of bad ones. Although voter turnout has been the most closely researched social desirability effect in political science (Holbrook and Krosnick 2010; Blair and Imai 2012), it is likely that social desirability invades other political attitudes as well. Sensitive questions that focus on the respondents' capability or ability often induce socially desirable responses that make the respondents seem healthier, more obedient, and more efficacious (Holbrook, Green, and Krosnick 2003; Tourangeau and Yan 2007; Kreuter and Tourangeau 2008; Gingerich 2010; Blair and Imai 2012).

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt049 Published online by Cambridge University Press

SAQs alternatively afford the individual greater privacy and anonymity, reducing or eliminating the socially desirable response. Thus, we should find evidence of social desirability in interviewer-driven methodologies on measures that make the respondent appear more confident and better able to handle the political world. We do not expect to find social desirability differences on behavioral measures such as media usage, but instead on attitudes about respondent performance in the election system and their ability to interact with it. We hypothesize that measures of efficacy including voter confidence at various levels (personal, county, and state), internal efficacy, overall rating of the vote experience, as well as trust in government will be more likely to display evidence of social desirability than behavioral measures such as watching the TV news, reading newspapers, discussing politics, the amount of time it took to vote, convincing others on political matters, giving money, the frequency of carrying government-issued identification, and statewide vote choices for Senate and President.

Satisficing may be another important problem that depends on mode response and theoretically may have a social desirability component. Satisficing occurs when respondents do not optimize their responses and instead answer questions with little motivation and with minimal cognitive effort (Krosnick 1991, 1999; Chang and Krosnick 2009, 2010). Satisficing leads respondents to choose satisfactory responses as opposed to optimized responses, where respondents carefully consider the question, retrieve relevant information from memory, make judgments about preferences, and then choose the best survey option (Cannell et al. 1981; Schwarz and Strack 1985; Tourangeau and Rasinski 1988). One way to measure satisficing is by analyzing the degree of nondifferentiation within a battery of survey questions; other methods include examining the quantity of open-ended responses or response times, all of which have the capacity to reduce the reliability of measures. Mixed mode surveys that combine both oral and self-administered surveys may produce different rates of satisficing due to the different visual patterns and the different cognitive effort involved in the survey. In particular, interviewer-driven phone surveys may motivate respondents to be attentive to the survey environment, and social desirability effects may reduce respondent incentives to respond quickly and with little effort to questions that probe different objects on the same scale (e.g., ideology or thermometer scores). For the respondent-driven interview the visual cues, for example an answer grid, may encourage identical response across different items due to reduced motivation. Some research shows that SAQs are more likely to display greater levels of nondifferentiation (Fricker et al. 2005), suggesting they have an increased level of satisficing. Other studies (Chang and Krosnick 2009, 2010), however, find that SAQs are less likely to display nondifferentiation than IAQs over an intercom. Based upon social desirability theory, we argue that the presence of an interviewer should diminish nondifferentiation, and therefore we hypothesize that we will find more evidence of satisficing in the respondent-driven survey mode.

A major problem in previous field studies comparing SAQs to IAQs is that they have difficulty differentiating mode from sample effects.[4] Thus, results often show differences between survey types, but those could be attributed to differences in sample composition and nonresponse bias. We solve this problem in our research design by using genetic propensity matching to hold the sample characteristics constant and then compare how mode interacts with key variables. Matching across samples takes account of nonresponse bias and sample differences between the surveys, leaving us with a data set that provides a valid comparison between respondents using different modes. We hypothesize that the IAQ will demonstrate more instances of response bias due to social desirability concerns across a variety of measures and more response differentiation. We argue that the presence of an interviewer and the conversational environment in which the interview takes place along with the psychological pressures that such a context inflicts on the individual are the key reasons for the behavior patterns we expect.

[4]An alternative approach by Chang and Krosnick (2010) uses an experimental design, which compares students who complete a survey on a computer with those who complete it over an intercom. The results of this study suggest that the computer instrument produced greater concurrent validity across a variety of measures than the intercom survey.

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt049 Published online by Cambridge University Press

3 Data

A key strength of our research design is that both the SAQs and IAQs are probability-based samples from the New Mexico voter file.[5] The substantive interest in our surveys was in assessing voters' election experience, which focused on voter confidence, voter satisfaction, poll worker helpfulness, experience with absentee ballots, experience at the polling place, and voter identification. We also asked about attitudes toward voter fraud, attitudes about the candidates, vote choice, political activity, partisanship, ideology, and demographic information (see Atkeson, Alvarez, and Hall 2010).[6] Because we could not obtain a sample frame of voters only until months after the election, we used registered voters from the November 2008 New Mexico voter file as our sample frame for both the self- and interviewer-administered surveys. Several months later, when the voter canvass was released, we merged this with the data provided by our survey respondents. We found that 92.1% of interviewer-administered respondents who indicated they voted were validated with county records. For the SAQ, the rate was 96.5%. The IAQ is on par with vote validation studies in ANES FTF surveys, which average between an 8% and 14% overreport, and the 3.5% error rate is very similar to the overreport estimates from the Current Population Studies (Traugott 1989; Belli, Traugott, and Beckman 2001).[7] We see this difference as our first test of social desirability effects between modes. Phone respondents are more likely to incorrectly indicate that they voted, the more socially desirable response, than Internet and mail respondents.

In the 2008 self-administered study, we sampled 6000 registered voters from the statewide voter registration file provided by the New Mexico Secretary of State. On November 10, we sent out 6000 first-class postcards to the sample of registered voters requesting their participation in our 2008 Election Administration Survey. The postcard explained the study and provided a web address (www.vote2008.unm.edu) that took respondents to a web page through which they could enter the survey (see Figure S1 in the supplementary material for a copy of the contact postcard). The postcard also explained that respondents could request a mail survey by contacting us via a toll-free number. Mail surveys were sent out immediately following requests to the respondent with a self-addressed and stamped envelope to return the survey to us. Subsequent "reminder" contacts were made via postcards to each nonrespondent to encourage response. Sample registered voters who did not respond were sent up to two reminder postcards with the same information. Respondents whose postcards were returned to us were labeled as unreachable.

The two types of SAQs (mail and Internet) were nearly identical, using similar styles for question presentation and order.[8] The survey software we used was Opinio 5.0. The web survey was partitioned across multiple pages and included a counter describing the percentage of the survey completed. The paper survey was on a 14" × 8" piece of paper and folded to create four sides of questions. Respondents colored in bubbles to answer questions. Eighty percent of respondents chose the Internet option and 20% chose the mail option. This number is consistent with our previous studies (Atkeson and Tafoya 2008; Atkeson and Adams 2010; Atkeson et al. 2011; Atkeson, Bryant, and Adams 2013). The self-administered response rate was 13.9% and was calculated as the number of completed or partially completed surveys divided by the number of persons in the sample that voted based upon government records (see Supplementary Table S2).[9,10]

[5]Atkeson, Lonna Rae; Adams, Alex N.; Alvarez, R. Michael. 2013. Replication data for: Nonresponse and Mode Effects in Self and Interviewer Administered Surveys. http://dx.doi.org/10.7910/DVN/23773 UNF:5:QH1dvPfRoJ+lbQk O32oeww== IQSS Dataverse Network [Distributor] V1 [Version].

[6]The frequency report can be found at http://polisci.unm.edu/common/documents/c-sved/papers/nm08pew.pdf (accessed January 11, 2014).

[7]We use the entire data sets for both surveys to test our hypotheses. We redid our analysis with only those who were identified as voters by the voter file, but the substantive interpretation and results were largely the same.

[8]Because of page constraints one question had to be moved elsewhere.

[9]This denominator represents the maximum number of eligible respondents, the total number of registered voters who voted in the sample, including voters who we were unable to contact and received a returned postcard from. A total of 13.9% represents the minimum response rate (RR2) as defined by the American Association of Public Opinion Research (AAPOR 2011). The cooperation rate was 2.6 points higher at 16.5%. The cooperation rate excludes voters whose mail was returned to us and were defined as unreachable. It represents the minimum cooperation rate (COOP2) according to AAPOR (2011).

[10]In this study, we do not consider differences between mail and Internet responses. Previous analysis of our data (Atkeson et al. 2011) shows few differences in response patterns. Given that both types of modes

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt049 Published online by Cambridge University Press

For the 2008 interviewer-administered telephone survey, we hired a prominent New Mexico firm, Research and Polling, Inc., to administer an almost identical survey.[11] Using the same sampling frame as the respondent-administered survey, a random sample of voters were interviewed by telephone. The self-reported telephone numbers from the current voter file were utilized when possible. A computerized telephone match was used for voter records without a phone number in the voter file. Each number was attempted a minimum of four times before it was taken off the callback list. A total of eight hundred interviews were completed among registered voters who indicated they cast a ballot in the 2008 general election, and interviews were conducted between November 6 and November 24, 2008, in both English and Spanish. The overall response rate is 17.4%, only a bit higher than our self-administered survey using the same response rate methodology (RR2 as defined by AAPOR 2011).

One limitation with our study is that we were unable to pull one random sample and then split it between the different survey designs, which would be the optimal test for mode differences. This was not possible because of the independent nature of the company we hired to do the phone survey, their process for matching records, and most importantly the timing of the two surveys. However, given that the sampling frames were identical, the comparison here presents one of the strongest research designs available for comparing the representativeness of both designs to the population and for determining differences across survey mode in response patterns due to social desirability.

4 Method

Voter registration data provide us with a rich set of measures to determine the characteristics of our population and sample respondents, including vote history and a variety of demographic information. This allows us to compare sample respondent characteristics with population characteristics to determine the quality of the collected data. We compare the population means to self- and interviewer-administered respondent means to determine representativeness. We are also able to compare sample respondents to a number of aggregate benchmarks related to the election, including several election outcomes. In all cases, we test for sample representativeness by comparing the population mean (both voter registration data and election outcomes) to the respondent means using a one-sample *t*-test.

We also compare socio-demographic and religious characteristics across the two samples using two sample *t*-tests to determine potential sampling differences across the two survey modes. Because we are limited to a small set of known attributes provided by the voter registration data, we expand our analysis of sample differences through a comparison of a broader set of measures across the two samples. Given what we know about the digital divide, we expect the respondent-driven mixed mode Internet/mail survey to have higher levels of socioeconomic resources, including income and education as well as a larger percentage of non-Hispanic whites.

Because we are interested in the quality of data between the surveys, we test for differences between them in terms of political attitudes and behavior, due to social desirability and satisficing. Using both two sample *t*-tests and multivariate regression, we test whether differences between the two samples exist in terms of variables that do and do not have a social desirability component. In the multivariate regression, we control for education, age, gender, ethnicity, and income.

However, one problem with these methods is that any significant differences between samples could be due to differences in sample composition from differences in nonresponse bias. Any comparison of social desirability bias across the modes is confounded with differences in sample composition due to different sample biases because of the different nature of the samples. Therefore, to solve this methodological problem, we extend this research and utilize matching methods to purge the surveys of any sampling effects due to nonresponse. Propensity scores are calculated for each observation by estimating the conditional probability that a respondent is in the interviewer- or

are SAQs and present largely the same format to the respondent, we are not surprised by the lack of a mode difference here.

[11] A couple of minor question wording differences were necessary because of the survey mode and we had two additional questions on the mixed mode survey.

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt049 Published online by Cambridge University Press

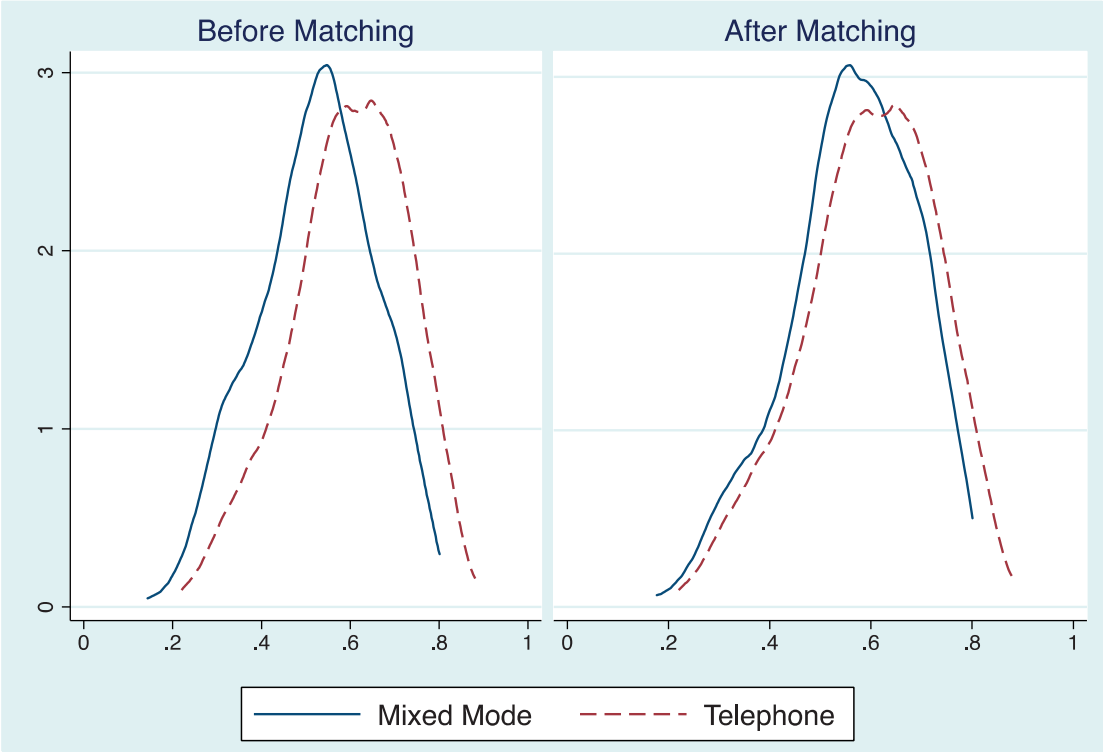


Fig. 1 Distribution of propensity scores for mixed mode and telephone respondents before and after matching.

respondent-driven survey,[12] which are then used by matching algorithms to pare down the samples so that the observed covariates are balanced across the two groups. Any remaining differences between survey means for our variables of interest after matching cannot be caused by differences in demographic characteristics between the samples and hence must be due to survey mode effects.

We used genetic matching in R's "MatchIt" routine to eliminate differences in sample composition between the self- and interviewer-administered samples (Ho et al. 2007a, 2007b; Sekhon 2011). Demographic characteristics and other predispositional factors that were different between the two survey modes are used as our covariates in the estimation of the propensity scores. These include: gender, age, education, ethnicity (white), voting mode (voted early, voted absentee), partisanship (included as dummy variables for Democrat and Republican), and born-again Christian. The propensity scores were then matched with the genetic matching, "...a multivariate matching method that uses an evolutionary search algorithm developed by Mebane and Sekhon (1998; Sekhon and Mebane 1998) to maximize the balance of observed covariates across matched treated and control units (Diamond and Sekhon 2012, 2)." The genetic match was unable to match 201 SAQs, but kept all of the responses from the IAQs.

We evaluate the balance of the propensity scores between mixed mode and telephone respondents before and after matching in Fig. 1. The figure shows that there is fairly good balance even before matching, which is not surprising given that both samples are probability-based designs and should be relatively close to one another. Nevertheless, the propensity score density plots are much more closely aligned after matching. Furthermore, Table S3 in the supplementary material shows that the covariate means between the "treated" respondents (IAQ) and the "control" respondents (SAQ) are much closer after matching. This suggests that the match is correctly balanced and that any differences observed in the matched sample are due to mode rather than sampling effects.

[12] $P(c_i = 1|x_i)$, where c_i is the "treatment" variable, taking a telephone survey, conditional on a vector x_i of the variables that may affect differences between survey modes.

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt049 Published online by Cambridge University Press

5 Results

5.1 *Evaluating Representativeness: Comparing Population Means to Survey Means*

Our general expectation is that both survey modes should represent well the population of interest, despite their relatively low response rates. This is especially important to show in the SAQ because it relies heavily on Internet responses. But this does not mean that we do not anticipate differences between the two samples. For example, due to the digital divide and the prominence of Internet responders in the SAQ, we expect the interviewer-administered phone respondents to be older and less educated than the self-administered survey.

The voter registration file provides demographic and voter history information about sample members, including their gender, age, party affiliation, and registration date. We use these data to first compare the self- and interviewer-administered respondent means to the population. We next use aggregate information about vote choice for President, Senate, and US Congress to examine each survey's ability to reflect the actual election outcomes.

Table 1 provides evidence about the quality of our voter samples. Specifically, column 4 shows the difference between the population mean (column 1) and the SAQ respondent mean (column 2) and the significance level from the one-sample *t*-tests. Column 5 presents the difference between the population mean (column 1) and the IAQ respondent mean (column 3). A positive sign in the fourth or fifth column indicates that the population mean is above the respondent mean, and a negative sign indicates that the population mean is below the respondent mean. Finally, column six reports the difference between the self- and interviewer-administered mean on each measure and whether they are significantly different from one another using a two-sample *t*-test.

We begin with gender. There is no significant difference between the percentage of males and females that voted in the state and those that voted in each survey type.[13] In terms of age, we do not see that the digital divide is greater for the SAQ. Both surveys underrepresent young voters and overrepresent older voters, but the SAQ underrepresents the young and overrepresents the old more than the IAQ.[14] However, the overall picture between the two surveys suggests a similar nonresponse bias. The number of years each respondent has been in the voter file also shows a slight overrepresentation of voters who have been in the voter file longer, but unlike with age here the SAQ does a slightly better job overall.

Both samples underrepresent respondents registered with no major party. Furthermore, the IAQ is spot on with Democrats, but overrepresents Republican respondents.[15] Conversely, the SAQs overrepresents Democrats, but there is no significant difference for Republicans.

We next examine region by looking at the proportion of voters in the three congressional districts in New Mexico. We find that the SAQs overrepresented the urban part of the state (CD1), where Internet access is likely high and where the university that sponsored the surveys is located, and underrepresented CD3 and CD2. The IAQs show a very similar pattern, but display substantively smaller and insignificant differences.

The last analysis focuses on differences between actual election outcomes and survey election outcomes in all the 2008 federal state contests. We find that we accurately represented all of the federal contests in the state in both survey modes, except that the SAQ significantly overrepresented the percent voting Democrat in the First Congressional District.

5.2 *Evaluating Representativeness across Survey Mode*

Table 1 also shows, in column 6, that there are few differences between both survey modes in the demographic characteristics examined from the voter file. Although most differences are insignificant, we find an approximately four-year difference between the modes on average for age.

[13]This finding is consistent across all three of our research designs, 2004, 2006, and 2008, using this method (Atkeson and Tafoya 2008; Atkeson and Adams 2010; Atkeson et al. 2011).

[14]A close look at the mixed mode survey suggests this is due to a higher rate of mail-requested surveys by older respondents (Atkeson et al. 2011).

[15]Unfortunately, we had to use party self-identification instead of party registration for the phone survey.

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt049 Published online by Cambridge University Press

Table 1 Mixed mode and telephone survey representativeness evaluations

	(1) Pop.	*(2) SAQ*	*(3) IAQ*	*(4) Pop.-SAQ*	*(5) Pop.-IAQ*	*(6) SAQ-IAQ*
% Male	45.8	45.0	47.0	0.8	−1.2	−2.0
Average age	48.0	55.7	52.3	−7.7***	−4.3***	3.4***
Age categories (years)						
18–30	21.0	9.8	15.5	11.2***	5.5***	−5.7***
31–45	24.1	15.4	19.5	8.7***	4.6***	−4.1*
46–50	9.8	8.7	8.5	1.1	1.3	0.2
51–65	26.8	36.0	32.1	−9.2***	−5.3***	3.9
66 and older	18.3	30.1	24.4	−11.8***	−6.1***	5.7**
Average registration (years)	10.0	15.2	16.9	−5.2***	−6.9***	−1.7
Registration categories						
0–9	55.9	51.7	49.0	4.2*	6.9***	2.7
10–19	29.7	27.4	32.8	2.3	−3.1	−5.4*
20 or more years	14.4	20.9	18.2	−6.5***	−3.8**	2.7
% Registered Dem	50.1	55.1	50.4	−5.0*	−0.3	4.7
% Registered GOP	31.6	33.0	37.4	−1.4	−5.8***	−4.4
% Registered no major party	18.2	11.9	12.3	6.3***	5.9***	−0.4
Region:						
% CD 1	35.4	44.4	38.1	−9.0***	−2.7	6.3*
% CD 2	30.4	25.5	29.9	4.9**	0.5	−4.4
% CD 3	34.2	30.2	32.0	4.1*	2.2	−1.8
% Voting democratic						
Presidential	56.9	59.0	57.3	−2.1	−0.4	1.7
Senate	61.3	60.4	61.9	0.9	−0.6	−1.5
CD 1	55.7	65.3	60.7	−9.6**	−5.0	4.6
CD 2	56.0	54.6	54.4	1.4	1.6	0.2
CD 3	56.7	62.4	57.7	−5.7	−1.0	4.7

Note. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; Pop. = 2008 NM Voter Population, SAQ = 2008 NM Mixed Mode Survey Respondent Mean, and IAQ = 2008 NM Phone Survey Respondent Mean.

Further examination reveals an almost six-year difference for the 18–30 age category, a four-year difference for the 31–45 age category, and almost a six-year difference for the 66 and older age category. This is surprising, given what we know about the digital divide, but from a close examination of the SAQ, we know that mail respondents were much older, skewing the average age upward (Atkeson and Adams 2010). However, the differences are rather small substantively, with an average age in the self-administered survey of 56 and an average age in the interviewer-administered phone survey of 52.

However, we can expand the number of comparison variables by examining additional differences between the two modes on other demographic indicators where we do not know the population mean. These include education, income, religiosity (measured as whether a respondent indicates they are born again), self-identified ethnicity, and marital status, which are shown in Table 2. Here we see that there were some additional differences in nonresponse between survey modes. First, we note that there is no difference in income, percent widowed, percent divorced, and married, but there is a large difference in the proportion of self-identified whites and Hispanics. A closer look at the SAQs shows that twice as many Hispanics chose the mail option over the Internet, suggesting a place where the digital divide may still be important and therefore presents an argument for providing different response options for different types of respondents.[16] We also see that the self-administered respondents overall are slightly more educated, which is

[16]There are also drawbacks, particularly on age.

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt049 Published online by Cambridge University Press

Table 2 Differences in other demographic characteristics survey modes

	IAQ/Phone	*SAQ/Mixed mode*	*Difference*
Income	3.5	3.4	0.1
Education	4.0	4.3	−0.3**
% Born again Christian	32.2	24.4	7.8**
% White	67.8	74.1	−6.3*
% Hispanic	26.3	17.0	9.3***
% Married	67.3	65.9	1.4
% Divorced	11.8	10.2	1.6
% Never married	12.2	16.4	−4.2*
% Widowed	7.7	6.9	0.8

Note. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, two-tail test.

consistent with what we know about the digital divide.[17] Finally, the IAQ also picked up fewer people who identified as born again, which may be an important characteristic for understanding relationships between variables.

5.3 *Mode Effects*

We examine mode effects with a set of questions that should display social desirability effects and with some that should not for comparison. We identified questions that presented an ego-driven impression of the respondent as these are the most likely to display social desirability effects. These include: personal voter confidence, state voter confidence, county voter confidence, overall voter experience, trust in government, and internal efficacy. Questions that should not display social desirability include behavioral questions that focused on how much voters watched or read the news, discussed politics, the amount of time they waited in line to vote, their vote choices for statewide President and Senate, and whether they regularly carry photo identification, convinced others how to vote, and gave money to parties and candidates. In our first analysis of social desirability effects, we examined each of these variables using a *t*-test and then examined them in a multivariate context. We controlled only for demographic variables in our multivariable model, including age, gender, Hispanic, other minority, education, and income.

Table 3 reports the difference in means between the self- and interviewer-administered surveys and the *p*-values for the variables in the analysis. Positive values indicate that the IAQ mean was greater than the SAQ mean. We see the expected variable clusters that we hypothesized. We find that respondents' personal behavior in the election was similar between survey modes. How much voters watched or read the news, discussed politics, gave money to parties and candidates, convince others how to vote, or carry identification did not significantly vary between survey modes. In addition, they waited in line about the same amount of time to vote and their vote choices for President and Senate were about the same.

However, voters' attitudes toward their experience with the election and with government display more socially desirable responses in the IAQ. Interviewer-administered respondents have more confidence that their votes are counted correctly, more trust in government, and are more internally efficacious, which is consistent with our hypothesis.

Due to social desirability effects and interviewer motivation, we posit that satisficing is less likely to occur in interviewer-driven surveys. To test the satisficing hypothesis, we examined a measure of nondifferentiation using the Mulligan et al. (2001) method as described by Chang and Krosnick (2009). For this measure, we compute the mean of the root of the absolute differences in the responses to all seven items from a set of questions about the perceived ideology of politicians.[18]

[17]This is especially true if we were to parse out the mail responders, who were on average less educated than those who responded via the Internet.

[18]These include: George W. Bush, John McCain, Sarah Palin, Barack Obama, Joe Biden, Steve Pearce, and Tom Udall.

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt049 Published online by Cambridge University Press

Table 3 Difference of means between interviewer- and self-administered modes on behavioral variables and ego-driven variables

	Difference of means (IAQ-SAQ)	P-*values*
Behavioral variables		
Watch TV news	0.096	0.465
Read news	0.000	0.999
Discuss politics	0.087	0.507
Time waited to vote	0.980	0.485
Presidential vote	−0.222	0.420
Senate vote	0.010	0.729
Frequency of carrying identification	−0.001	0.983
Convinced others how to vote	0.021	0.434
Gave money to candidate	−0.021	0.404
Socially desirable variables		
Vote confidence	0.112	0.004
State vote confidence	0.229	0.001
County vote confidence	0.237	0.001
Vote experience	0.083	0.011
Trust in Government	0.223	0.001
Internal efficacy	0.153	0.012
Nondifferentiation (Ideology Scores)	−0.020	0.031

We then transformed the variable, such that the nondifferentiation score ranged from 0, no nondifferentiation, to 1, complete nondifferentiation, to increase interpretability. We find that SAQ respondents were significantly more likely to nondifferentiate.

5.4 *Mode-Effects Controlling for Sampling Effects*

To control for the possibility that sampling effects are confounding our findings, we reran the regressions on the matched data set. Table 4 shows the mode coefficient from the multiple regression models for the entire sample and for the matched sample where sampling differences in the demographic and predispositional factors were taken into account. The advantage of the matched sample is that differences cannot be attributed to potential sampling differences observed in Tables 1 and 2 and therefore are most likely due to mode differences. We see that even though the coefficients are marginally different between the entire sample and the matched sample, the direction, significance levels, and hence conclusions are the same. Specifically, we still find that survey mode is not significantly related to responses regarding personal behavior, but that it is significantly correlated with variables that theoretically should display social desirability.

Voter confidence, trust, efficacy, and nondifferentiation measures are still significant in the matched sample, confirming our hypotheses with a strong empirical method. This provides evidence that respondents in the interviewer-administered survey appear to feel social pressure to indicate that they are more confident in the voting process and in themselves than those who take the survey independently of an interviewer. They also appear to be more cognitively engaged and as a consequence less likely to satisfice.

6 Discussion and Conclusion

We implemented an SAQ and an IAQ using the same sampling frame in New Mexico. Like Visser et al. (1996), who studied the success of a pre-election mail survey that accurately predicted the election outcomes, our results show that both survey modes represented several characteristics of the sample and the election outcomes, although significant differences were also seen on a variety of

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt049 Published online by Cambridge University Press

Table 4 Comparison of models for whole sample and for matched sample, controlling for age, gender, Hispanic, other minority, education

	Treatment coefficient unmatched data set	*Treatment coefficient matched data set*
Behavioral variables		
Watch TV news	0.207	0.232
Read news	0.218	0.049
Discuss politics	0.119	0.053
Time waited to vote	1.146	1.298
Presidential vote	−0.118	−0.119
Senate vote	−0.001	0.136
Frequency of carrying Government issued identification	0.017	0.070
Convinced others how to vote	0.073	0.009
Gave money to candidate	0.055	−0.021
Socially desirable variables		
Vote confidence	0.524**	0.382***
State vote confidence	0.580***	0.478***
County vote confidence	0.671***	0.532***
Vote Experience	0.400***	0.436**
Trust in Government	0.640***	0.713***
Internal efficacy	0.506***	0.600***
Non-differentiation (Ideology Scores)	−0.029**	−0.037**
Ns	904–1334	680–980

Note. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

variables. We find that both surveys substantively represent well the actual election outcomes, despite the relatively low response rates and potential coverage issues. There is some underrepresentation of the young, Independent, and rural voters, and overrepresentation of older and urban voters, but substantively these differences are not large. These results provide evidence that low response rates are not necessarily indicative of a poor research design and nonresponse error, which is consistent with recent research on this issue (Curtin, Presser, and Singer 2000; Keeter et al. 2000; Merkle and Edelman 2002; Groves, Presser, and Dipko 2004; Pew Research Center for the People and the Press 2004; Groves and Peytcheva 2008; Atkeson and Adams 2010; Atkeson et al. 2011). Importantly, the self-administered survey provides an alternative to expensive interviewer-administered designs while maintaining a probability-based sampling method. The self-administered survey cost approximately two-thirds less than the interviewer-administered phone survey, it could be done in-house, and the researchers remained in full control of all aspects of survey administration, providing us important feedback on all aspects of our research design. Online polls, another Internet alternative, do not offer the advantage of a probability-based sampling method (see AAPOR 2010). Self-administered surveys offer us the opportunity to combine contact and response modes in a way that builds upon our theoretical understanding of how random sampling from a subset of the population allows us to infer the attitudes and behavior of the larger population.

We find strong evidence for mode effects due to social desirability and the presence of an interviewer. We confirm this result by using matching methods to remove differences in sample composition that could also be responsible for mode differences. Our results are consistent with previous studies. As expected, variables that measure specific behaviors surrounding the campaign do not show a social desirability effect. But, variables that are more ego-driven were significantly more positive in the IAQ due to the presence of the interviewer. Because of the significant role played by social desirability, it is essential that scholars control for mode effects in surveys that combine SAQ and IAQs. Simply combining surveys with identical wordings by different answer modes is problematic and may create unreliable measures and influence relationships between variables.

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt049 Published online by Cambridge University Press

Data collection methodology matters to the quality of data and is one component that creates errors in survey research. Because mixed mode surveys are becoming quite popular, we need to be attentive to differences as we combine modes and control for mode effects within our models. The 2012 ANES study, for example, combines FTF with Internet panel responders. Importantly, these are distinctly different modes with distinctly different types of respondents (see e.g., Adams, Atkeson, and Karp 2013). Scholars need to be sensitive to these differences and their effects when they use survey studies that combine survey modes. The quality of our data is critical to our ability to test theories and hypotheses using observational data. The nature of the data is rapidly changing and we must be attentive to data quality and how different aspects of the data collection process contribute to the total survey error and how we can look and correct for it.

Funding

Funding for this research was provided by the Pew Charitable Trusts' Center on the States and the JEHT Foundation's Make Voting Work Initiative.

References

Adams, Alex, Lonna Atkeson, and Jeffrey Karp. 2013. Measurement error in online survey panels: Panel conditioning and data quality. Presented at the American Panel Survey (TAPS) Workshop, Washington University, St. Louis, November 8–9.

American Association for Public Opinion Research. 2010. *AAPOR report on online panels*. http://www.aapor.org/AAPOR_Releases_Report_on_Online_Survey_Panels/2263.htm (accessed January 10, 2014).

The American Association for Public Opinion Research. 2011. *Standard definitions: Final dispositions of case codes and outcome rates for surveys, 7th edition*. AAPOR. http://aapor.org/Content/NavigationMenu/AboutAAPOR/StandardsampEthics/StandardDefinitions/StandardDefinitions2011.pdf (accessed January 10, 2014).

Alexander, C. H. Jr., and S. Wetrogan. 2000. Integrating the American Community Survey and the Intercensal Demographic Estimates Program. Proceedings of the American Statistical Association at www.amstat.org/sections/srms/proceedings/papers/2000_046.pdf (accessed January 10, 2014).

Aquilino, W. 1994. Interview mode effects in surveys of drug and alcohol use. *Public Opinion Quarterly* 58:210–40.

Atkeson, Lonna Rae, Alex N. Adams, and R. Michael Alvarez. 2013. Replication data for: Nonresponse and mode effects in self- and interviewer-administered surveys. http://dx.doi.org/10.7910/DVN/23773 UNF:5:QH1dvPfRoJ+lbQkO32oeww== IQSS Dataverse Network [Distributor] V1 [Version].

Atkeson, Lonna Rae, and Alex N. Adams. 2010. Mixed mode (Internet and Mail) probability samples and survey representativeness: The case of New Mexico 2008. San Francisco, CA: Presented at the Western Political Science Association, April 1–4.

Atkeson, Lonna Rae, Alex N. Adams, Lisa A. Bryant, Luciana Zilberman, and Kyle L. Saunders. 2011. Considering mixed mode surveys for questions in political behavior: Using the Internet and mail to get quality data at reasonable costs. *Political Behavior* 33(1):161–78.

Atkeson, Lonna Rae, and Lori Tafoya. 2008. Surveying political activists: An examination of the effectiveness of a mixed-mode (Internet and mail) survey design. *Journal of Elections Public Opinion and Parties* 18(4):367–86.

Atkeson, Lonna Rae, R. Michael Alvarez, Alex N. Adams, and Lisa Bryant. 2011. The 2010 New Mexico election administration report. Typescript, University of New Mexico. http://polisci.unm.edu/common/documents/c-sved/papers/nm-2010-general-election.pdf (accessed January 10, 2014).

Atkeson, Lonna Rae, Lisa A. Bryant, and Alex N. Adams. 2013. The 2012 Bernalillo County election administration report. Typescript, University of New Mexico. http://polisci.unm.edu/common/documents/c-sved/papers/BernCoFinalReport.pdf (accessed January 10, 2014).

Atkeson, Lonna Rae, R. Michael Alvarez, and Thad E. Hall. 2010. Assessing electoral performance in New Mexico using an eco-system approach: New Mexico 2008. Typescript, University of New Mexico. http://polisci.unm.edu/common/documents/c-sved/papers/nm08pew.pdf (accessed January 10, 2014).

Balden, W. 2004. Multi-mode data collection: Benefits and downsides. Paper presented at the 2004 Conference of the Great Lakes. Cancun, Mexico: Chapter of the Marketing Research Association.

Belli, Robert F., Michael W. Traugott, and Matthew N. Beckmann. 2001. What leads to voting overreports? Contrasts of overreporters and admitted nonvoters in the American National Election Studies. *Journal of Official Statistics* 17(4):479–98.

Blair, Graeme, and Kosuke Imai. 2012. Statistical analysis of list experiments. *Political Analysis* 20:47–77.

Blumberg, Stephen J., and Julian V. Luke. 2009. Wireless substitution: Early release of estimates from the National Health Interview Survey. July–December 2008. http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless200905.htm.

———. 2013. Wireless substitution: Early release of estimates from the National Health Interview Survey. July–December 2012. http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201306.pdf.

Burden, Barry C. 2000. Voter turnout and the National Election Studies. *Political Analysis* 8(4):389–98.

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt049 Published online by Cambridge University Press

———. 2003. Internal and external effects on the accuracy of NES turnout: Reply. *Political Analysis* 11(2):193–95.
Campbell, Donald T. 1950. The indirect assessment of social attitudes. *Psychological Bulletin* 47):15–38.
Cannell, C. F., P. V. Miller, and L. Oksenberg. 1981. Research on interviewing techniques. In *Sociological Methodology*, ed. S. Leinhardt, 389–437. San Francisco, CA: Jossey-Bass.
Carley-Baxter, Lisa R., Andy Peytchev, and Michele C. Black. 2010. Comparison of cell phone and landline surveys: A design perspective. *Field Methods* 22(1):3–15.
Chang, LinChiat, and Jon Krosnick. 2009. National surveys via RDD telephone interviewing versus the Internet: Comparing sample representativeness and response quality. *Public Opinion Quarterly* 73(4):641–78.
———. 2010. Comparing oral interviewing with self-administered computerized questionnaires: An experiment. *Public Opinion Quarterly* 74(1):154–67.
Christian Leah M., and Don A. Dillman. 2004. The influence of graphical and symbolic language manipulations on responses to self-administered questions. *Public Opinion Quarterly* 68(1):57–80.
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.
Davis, James Allan, Tom Smith, and Peter V. Marsden. 2009. *General social surveys, 1972–2008: Cumulative codebook*. Chicago: National Opinion Research Center.
Day, N. A., D. R. Dunt, and S. Day. 1995. Maximizing response to surveys in health program evaluation at minimum cost using multiple methods. *Evaluation Review* 19(4):436–50.
de Leeuw, Edith. 2005. To mix or not to mix: Data collection modes in surveys. *Journal of Official Statistics* 21:233–55.
de Leeuw, E., J. Hox, E. Korendijk, G. Lensvelt-Mulders, and M. Callegaro. 2004. The influence of advance letters on response in telephone surveys: A meta-analysis. Maastricht: Paper presented at the 15 International Workshop on Household Survey Nonresponse.
de Leeuw, Edith D., and W. de Heer. 2002. Trends in household survey nonresponse: A longitudinal and international comparison. In *Survey nonresponse*, eds. Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little. New York: Wiley.
de Leeuw, Edith, and J. Van der Zouwen. 1988. Data quality in telephone and face-to-face surveys: A comparative meta-analysis. In *Telephone survey methodology*, eds. R. Groves, P. P. Bimer, L. Lyberg, I. T. Massey, W. L. Nicholls, and J. Waksberg. New York: John Wiley & Sons.
DeBell, Mathew, Catherine Wilson, Gary Segura, Simon Jackman, and Vincent Hutchings. 2011. *Methodology report and user's guide for the ANES 2010–2012 Evaluations of Government and Society Study*. Palo Alto, CA, and Ann Arbor: Stanford University and University of Michigan.
Diamond, Alexis, and Jasjeet S. Sekhon. 2012. Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics*. http://sekhon.berkeley.edu/papers/GenMatch.pdf.
Dillman, Don A. 2000. *Mail and Internet surveys: The tailored design method*. 2nd ed. New York: Wiley.
Dillman, Don A., R. Sangster, John Tarnai, and T. Rockwood. 1996. In *Current issues in survey research: New directions for evaluation, #70*. Chapter 4, eds. Mark T. Braverman, and Jana Kay Slater, 45–62. San Francisco, CA: Jossey-Bass.
Dillman, Don A., and Leah Malani Christian. 2005. Survey mode as a source of instability in responses across surveys. *Field Methods* 17(1):30–52.
Dillman, Don A., Jolene Smyth, and Leah Melani Christian. 2009. *Internet, mail, and mixed-mode surveys: The tailored design method*. New York: Wiley.
Fowler, Floyd Jackson., Jr, Anthony M. Roman, and Zhu Xiao Di. 1998. Mode effects in a survey of Medicare prostate surgery patients. *Public Opinion Quarterly* 62(1):29–46.
Fricker, S., M. Galesic, R. Tourangeau, and T. Yan. 2005. An experimental comparison of web and telephone surveys. *Public Opinion Quarterly* 69(3):370–92.
Gingerich, D. W. 2010. Understanding off-the-books politics: Conducting inference on the determinants of sensitive behavior with randomized response surveys. *Political Analysis* 18:349–80.
Groves, Robert M. 1989. *Survey errors and survey costs*. New York: John Wiley and Sons.
Groves, R. M., and R. L. Kahn. 1979. *Surveys by telephone: A national comparison with personal interviews*. New York: Academic Press.
Groves, R. M., and E. Peytcheva. 2008. The impact of nonresponse rates on nonresponse bias. *Public Opinion Quarterly* 72:1–23.
Groves, Robert M., Stanley Presser, and Sarah Dipko. 2004. The role of topic interest in survey participation decisions. *Public Opinion Quarterly* 68(1):2–31.
Groves, Robert, Eleanor Singer, and Amy Corning. 2000. Leverage-salience theory of survey participation. *Public Opinion Quarterly* 64:299–308.
Ho, Daniel, Kosuke Imai, Gary King, and Elizabeth Stuart. 2007a. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis* 15(3):199–236.
———. 2007b. Matchit: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*. http://gking.harvard.edu/matchit/.
Holbrook, A. L., M. C. Green, and J. A. Krosnick. 2003. Telephone versus face-to-face interviewing of national probability samples with long questionnaires. *Public Opinion Quarterly* 67:79–125.
Holbrook, Allyson L., and Jon A. Krosnick. 2010. Social desirability bias in voter turnout reports: Tests using the item count technique. *Public Opinion Quarterly* 74(1):37–67.

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt049 Published online by Cambridge University Press

Hox, J. J., and E. D. de Leeuw. 1994. A comparison of nonresponse in mail, telephone, and face-to-face surveys: Applying multilevel models to meta-analysis. *Quality and Quantity* 28:329–44.
Iannacchione, Vincent. 2011. The changing role of address-based sampling in survey research. *Public Opinion Quarterly* 75(3):556–75.
Keeter, Scott, Carolyn Miller, Andrew Kohut, Robert M. Groves, and Stanley Presser. 2000. Consequences of reducing non-response in a national telephone survey. *Public Opinion Quarterly* 64:125–48.
Kreuter, F., S. Presser, and R. Tourangeau. 2008. Social desirability bias in CATI, IVR, and web surveys: The effects of mode and question sensitivity. *Public Opinion Quarterly* 72:847–65.
Krosnick, J. A. 1991. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology* 5:213–36.
———. 1999. Maximizing questionnaire quality. In *Measures of political attitudes*, eds. J. P. Robinson, P. R. Shaver, and L. S. Wrightsman. New York: Academic Press.
London, K., and L. Williams. 1990. A comparison of abortion underreporting in an in-person interview and self-administered questionnaire. Toronto: Paper presented at the Annual Meeting of the Population Association of America.
Lupia, Arthur, Jon Krosnick, Pat Luevano, Matthew DeBell, and Darrell Donakowski. 2009. Users Guide to the ANES 2008 Time Series Study. Ann Arbor, MI, and Palo Alto, CA: University of Michigan and Stanford University. http://www.electionstudies.org/studypages/2008prepost/2008prepost_UsersGuide.pdf.
Martinez, Michael D. 2003. Comment on voter turnout and the National Election Studies. *Political Analysis* 11(2):187–92.
McDonald, Michael P. 2003. On the overreport bias of the National Election Study turnout rate. *Political Analysis* 11(2):180–86.
———. 2007. The true electorate: A cross validation of voter registration files and election survey demographics. *Public Opinion Quarterly* 71(4):588–602.
Mebane, Walter R. Jr, and Jasjeet S. Sekhon. 1998. GENetic Optimization Using Derivatives (GENOUD). Software package. http://sekhon/berkeley.edu/rgenoud/.
Mensch, Barbara S., and Denise B. Kandel. 1988. Underreporting of substance use in a National Longitudinal Youth Cohort. *Public Opinion Quarterly* 52(Spring):100–24.
Merkle, Daniel M., and Murray Edelman. 2002. Non-response in exit polls: A comprehensive analysis. In *Survey non-response*, eds. Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, 243–58. New York: John Wiley & Sons Inc.
Mulligan, Kenneth, Jon A. Krosnick, Wendy Smith, Melanie Green, and George Bizer. 2001. Nondifferentiation on attitude rating scales: A test of survey satisficing theory. Unpublished manuscript.
O'Muircheartaigh, Colm, Ned English, Michael Latterner, Stephanie Eckman, and Katie Dekker. 2009. Modeling the need for traditional vs. commercially available address listings for in-person surveys: Results from a National Validation of Addresses JSM. Proceedings of the American Statistical Association, Section on Survey Research Methods 6193–202.
Peterson, Robert A., and Roger A. Kerin. 1981. The quality of self-report data: Review and synthesis. In *Review of marketing*, eds. Ben M. Enis and Kenneth J. Roering, 5–20. Chicago: American Marketing Association.
Pew Research Center for the People and the Press. 2004. Polls face growing resistance, but still representative: Survey experiment shows. http://people-press.org/reportsdisplay.php3?ReportID=211.
Presser, Stanley, Michael W. Traugott, and Santa Traugott. 1990. Vote "over" reporting in surveys: The records or the respondents. *Technical Report No. 39*. Ann Arbor, MI: National Election Studies.
Rivers, Douglas. N.d. Sample matching: Representative sampling from Internet polling. *Polimetrix*.
Rosen, R., and T. Gomes. 2004. Converting CES reporters from TDE to web data collection. Toronto, Canada: Paper presented at the Joint Statistical Meetings.
Schuman, Howard, and Stanley Presser. 1981. *Questions and answers in attitude surveys: Experiments on question form, wording and context*. New York: Academic Press.
Schuman, Howard. 1992. Context effects: State of the past/state of the art. In *Context effects in social and psychological research*, eds. N. Schwarz and S. Sudman, 5–20. New York: Springer.
Schwarz, N. 1996. *Cognition and communication judgmental biases, research methods, and the logic of conversation*. Mahwah, NJ: Lawrence Erlbaum.
Schwarz, N., and E. Strack. 1985. Cognitive and affective processes in judgments of subjective well-being: A preliminary model. In *Economic psychology*, eds. H. Brandstatter and E. Kirehler, 439–47. Linz: Austria: R. Trauner.
Sekhon, Jasjeet Singh, and Walter R. Mebane Jr. 1998. Genetic optimization using derivatives: Theory and application to nonlinear models. *Political Analysis* 7:189–203.
Sekhon, Jasjeet S. 2011. Multivariate and propensity score matching software with automated balance optimization: The matching package for R. *Journal of Statistical Software* 42(7):1–52.
Shettle, Carolyn, and Geraldine Mooney. 1999. Monetary incentives in government surveys. *Journal of Official Statistics* 15:231–50.
Smyth, Jolene D., Don Dillman, Leah Malani Christian, and Michael J. Stern. 2006. Effects of using visual design principles to group response options in web surveys. *International Journal of Internet Science* 1:6–16.
Stoop, I. A. L. 2005. *The hunt for the last respondent*. The Hague: Social and Cultural Planning Office (SCP).
Seymour, Sudman, Norman M. Bradburn, and Norbert Schwarz. 1996. *Thinking about answers*. San Francisco, CA: Josey-Bass.

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt049 Published online by Cambridge University Press

Tourangeau, Roger, Mick Couper, and Frederick Conrad. 2004. Spacing, position and order: Interpretive heuristics for visual features of survey questions. *Public Opinion Quarterly* 68(3):368–93.

Tourangeau, R., and K. A. Rasinski. 1988. Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin* 103:299–314.

Tourangeau, Roger, and Ting Yan. 2007. Sensitive questions in surveys. *Psychological Bulletin* 133:859–83.

Traugott, Santa. 1989. Validating self-reported vote: 1964–1988. *ANES Technical Report Series, No. nes010152.* Ann Arbor, MI: National Election Studies.

U.S Census Bureau. *American community survey 2006–2008 three-year estimates. Data profile: New Mexico: Bernalillo County: Hispanics (All).*

Visser, Penny S., Jon A. Krosnick, Jesse Marquette, and Michael Curtin. 1996. Mail surveys for election forecasting? An evaluation of the Columbus Dispatch Poll. *Public Opinion Quarterly* 60(2):181–227.

Zickhur, K., and A. Smith. 2012. *Digital differences*. Washington, DC: Pew Internet and American Life Project. http://pewinternet.org/Reports/2012/Digital-differences.aspx (accessed August 7, 2012).

Downloaded from http://pan.oxfordjournals.org/ at California Institute of Technology on August 10, 2016

https://doi.org/10.1093/pan/mpt049 Published online by Cambridge University Press