

Title: Reducing Over-Reporting of Voter Turnout: An Experiment Using a 'Source Monitoring' Framework

Authors: Belli, Robert, Traugott, Santa and Steven J. Rosenstone

Date: November, 1994

Dataset: Special Fall '94 telephone survey

Keywords: over-reporting of vote, source monitoring, church attendance

At its June 1994 meeting, the NES Board of Overseers authorized a small experiment to test out new wording of the voter turnout question intended to reduce the tendency of respondents to report having voted when they had not. The experiment, conducted in the weeks following the 1994 general elections, was a collaborative effort between NES and the Survey Research Center's Survey Methods Program.

This report provides the results of this experiment. We begin by laying out the general problem of over-reporting and briefly reviewing the earlier, unsuccessful attempts (by NES and others) to solve the problem. We then introduce the source monitoring framework that underlies our approach to the problem. Next we describe the study design and data collection. We then report the results.

The Problem of Over-Report of Voter Turnout and Attempts to Solve It

A well known problem in survey research is that people often report to survey interviewers having voted, when in fact, they had not. Two general strategies have been taken to deal with over-reports of turnout. Voter validation efforts have been directed at generating a more accurate measure of whether the respondent actually voted by checking, ex post facto, official records to identify and correct the response of those respondents who have misreported. Voter validation studies have been carried out in conjunction with the SRC/CPS National Election Studies in 1964, 1976, 1978, 1980, 1984, 1986, 1988, and 1990.¹ As part of the 1990 voter validation, NES revalidated the validation conducted in 1988 as a means of checking the reliability of voter validation itself. What we have learned from voter validation over the years, and in particular from the revalidation exercise, is that voter validation is very expensive, it introduces errors in its own right, and it does not appreciably increase the reliability of the data (Traugott 1989; Presser, Traugott and Traugott 1990; Traugott, Traugott, and Presser 1992).

A second approach has been to tinker with the way the voter turnout question is asked. Some efforts have experimented with alternative question wordings in an attempt to reduce the social desirability of reporting that one has voted (e.g. Presser 1990). Other recent approaches have considered over reporting as a telescoping and memory problem (Abselson, Loftus, and Greenwald 1992). But, these approaches have not been successful. These failures to reduce over-reporting suggest that the psychological processes responsible for over-reporting have not been adequately identified.

The Source Monitoring Approach

The present experiment treated over-reporting of turnout within the *source monitoring* framework of Johnson, Hashtroudi, and Lindsay (1993) and others (e.g., Foley and Johnson 1985; Johnson, Foley, Suengas, and Raye 1988; Johnson and Raye 1981). Source monitoring is concerned with the ability to determine the source of one's memories. According to this framework, the source of one's memories is not directly specified in the memories themselves, but rather, source information is inferred from various aspects of the content which is contained in the memories. That is, to determine the source of one's memories, say between thought and action, involves an attribution process based on the characteristics of the remembered information. At a basic level, most external events (that occur in reality), such as actions, are encoded with rich information on the perceptions (e.g., visual, kinesthetic, and acoustic detail) that are associated with experiencing a real world event. On the other hand, most events that occur internally, such as thought or imagination, are encoded with information on the cognitive, thinking and reflective processes that the individual uses in generating the images. Thus, during later remembering, events that originated externally are likely to have richer perceptual information than events internally generated, whereas internally originated events are likely to have richer information about cognitive operations than events externally experienced (Johnson et al. 1988; Schooler, Gerhard and Loftus 1986). People, on the basis of these characteristics, are often able to correctly distinguish real from internally generated events (Johnson et al. 1993).

On the other hand, source confusions will occur when the characteristics of the remembered information are not reliable indications of source. Such lacking of reliability will occur when there is considerable overlap in the quantity or quality of the characteristics of information remembered from real and internally generated events. With voting, both actual voting and only thinking about voting share the intention to vote which may involve substantive cognitive operations associated with who to for and for what reasons. In other words, there may be confusion among respondents between having only thought about voting with having actually voted. Analogously, in psychological experiments, subjects show source confusions when the memory characteristics of real and imagined events are similar (Johnson, et al. 1993).

An examination of data characteristics suggests that the inaccuracies found in remembering voting behavior may be a particularly good example of the presence of source confusions. One characteristic involves an asymmetry in source confusion; experimental work (Anderson 1988; Belli, Lindsay, Gales and McCarthy 1994; Foley, Durso, Wilder, and Friedman 1991) has revealed that there is a greater likelihood to confuse internal experience as external experience than vice-versa. As is well known, a similar asymmetry exists in reports of voting behavior. People overwhelmingly tend to over report having voted in a particular election when in fact they did not; there is only a very small degree of under reporting by some individuals who actually did vote and report not voting (Abelson et al. 1992). In addition, the psychological literature has found that thinking about intentions can exacerbate source confusion (Foley, et al. 1991). Interestingly, people who claim on pre-election interviews an intention to vote are more likely to over-report than others (Silver, Anderson, and Abramson 1986) and high over-reporting occurs more rapidly after a Presidential than other elections, suggesting that the elections with more intended thought about voting lead to more confusions (Abelson et al. 1992). Finally, people with a history of voting more often

have a greater tendency to over-report in comparison to those who typically vote less often (Abelson et al. 1992); having a history of voting apparently exacerbates the likelihood of confusing a more remote activity as a more recent one, which in fact had only been intended.

This line of thought does not rule out a role for social desirability. Given that people may be uncertain on whether they actually performed an action (such as voting) or merely thought about doing so, the social desirability of an action may bias respondents to over-report in conditions of uncertainty. However, attacking the social desirability component alone has not been successful; what was needed, we thought, was an approach that would address the role of source confusion more directly.

By recognizing the potential role of source confusion, we thought that a well designed question may reduce over-reporting. Psychological experimentation, for example, has found that a source monitoring test which asks subjects to scrutinize the sources of their memories for events that were externally or internally generated reduces source confusion in comparison to simpler yes/no tests on whether the events were external in origin (Lindsay and Johnson 1989; Zaragoza and Lane, in press). At present, the NES question on voting behavior asks for a "yes/no" response; we devised an alternative, experimental "source monitoring test" that we hypothesized would be effective in reducing over-reporting. Using this approach, respondents were made aware of the possibility that sometimes people confuse their thinking about voting and who to vote for with actually having voted. Respondents were also provided with "only thought about voting" as a response option. We used an additional strategy to help respondents scrutinize their memories, and thereby to recover the perceptual information (e.g., visual, kinesthetic and acoustic detail) that would be associated with actually having voted and that likely would not be as pervasive in respondents who did not vote. (Full question text appears in Appendix A.) By emphasizing those memory characteristics that would differentiate voting from non-voting, we thought, respondents might be able to more accurately remember whether they had actually voted or only thought of doing so.

Study Design

We undertook a controlled experiment where subjects were randomly assigned to either a treatment group which received the new improved turnout question, or a control group, which received the standard, NES voter turnout question. To test the replicability of the results on the turnout wording experiment, within the same study we conducted a second question wording experiment that focused on reported church attendance. A random half of the subjects were asked the standard, Gallup question on church attendance; the other half were asked a new improved question which, like the experimental question on voter turnout, both walked respondents through the experience of church attendance and expanded the response options (beyond yes/no) to include responses that permitted the respondent to indicate that she "thought about going this weekend, but didn't" and that she usually goes, "but didn't this weekend." (The full text of the church attendance items appear in Appendix A.)

Subjects were randomly assigned to one of four groups: Those in Group 1 received the standard NES version of the vote question and the standard Gallup version of the church attendance question. Those in Group 2 were asked the standard NES turnout and the experimental church attendance. Group 3 subjects were administered the experimental turnout question and the standard church attendance question. Group 4 subjects were asked both experimental questions. (See Appendix A for a summary of the number of respondents assigned to each condition.)

A total of 700 persons (350 from Ann Arbor, Michigan and 350 from Ypsilanti, Michigan) participated in the study which consisted of a 9-minute telephone interview conducted by the Survey Research Center's Telephone Facility. Respondents were randomly selected from the list of persons registered to vote in these two communities. Because the voter registration lists in these two communities did not include the individual's telephone number, a random subsample of these two lists was passed through a commercial service that matched names and addresses from lists with names and addresses in recent (computerized) telephone directories.² Thus, individuals eligible to take part in this study were a random sample citizens registered to vote in these two communities who had a listed telephone number.

The content of the questionnaire was designed to mimic the questions on the 1994 National Election Study that lead up to the voter turnout question, which appears about 10 minutes into the 70 minute long, face-to-face interview. The turnout wording experiment appeared about 6 minutes into our survey; the church attendance question wording experiment appeared a few minutes later. In every respect, the interview was designed to mimic conditions that would prevail in a post-election survey of the American electorate.

We also interviewed over a 10 week period, from November 9 through December 20 to emulate as much as possible the time during which interviews were being conducted in the 1994 National Election Study. We released sample on a regular basis, so that any effects of time on misreporting would not be confounded with length of time it took to get an interview. The sample was divided into random "replicates" so that those whose names came early in the alphabet were not introduced early into the sample.

In order to minimize costs, we deviated from usual standards of survey practice in several ways. First, we instructed interviewers not to probe, and not to write down answers to open-end questions. To further economize, we limited to two the number of calls to any selected respondent unless a firm appointment had been set.³ This produced an overall response rate for the study of only 29 percent. Were we conducting a sample survey, the biased sample frame and low response rate would be serious deficiencies in design. Because our interest was not in a representative sample, but in the differences between the control and treatment groups, we considered these deficiencies to be less serious. However, if there is an interaction between being the willingness of subjects to take part in our study and the efficacy of the experimental interventions, then biases might well result. We think that this is unlikely.

By selecting subjects from lists of registered voters in Ann Arbor and Ypsilanti, we not only could control costs, but could validate, by checking the public records after the election, whether the subjects in our study actually did or did not vote in the 1994 general election. This validation of the vote achieves the "golden standard" not usually attainable in vote validation of respondents to national surveys. The NES staff only had to navigate through two offices, both of which maintained machine readable registration lists. We knew that every subject in the study was registered to vote and we knew her correct name,

address, and town where she was registered. In short, the problems that reduce the reliability of voter validation efforts conducted on respondents in the typical national sample, were not present in this study (Traugott 1989; Presser, Traugott, and Traugott 1990; Traugott, Traugott, and Presser 1992).

Results of the Experiment to Reduce the Over-Report of Voter Turnout

Respondents who were asked the experimental version of turnout question were not less likely to report having voted than respondents asked the standard NES question:

Question Format	Percent who Reported Voting
Standard NES Question	87.8
Experimental Question	87.1

χ^2 (1 d.f.) = .065 (prob = .798)

When judged against the baseline of validated vote, those respondents who were administered the experimental version of the turnout question were 3.6 percent less likely to report having not voted (when they in fact did not vote) than respondents who were asked the standard NES question. The paucity of true non-voters in the sample (118), however, means that one should have little confidence that this effect did not occur by chance.

Question Format	Percent of Actual Non-Voters Who Reported Having Voted
Standard NES Question	29.8
Experimental Question	26.2

χ^2 (1 d.f.) = .189 (prob = .664)

In the aggregate, the experimental version of the turnout question does not appreciably improve the overall accuracy of reported turnout:

Question Format	Percentage of Respondents Correctly Reporting Whether they Voted
Standard NES Question	94.5
Experimental Question	95.2

χ^2 (1 d.f.) = .217 (prob = .642)

In summary, there is little, if any evidence that the experimental question on voter turnout reduces the propensity of respondents to over-report having voted.

We tested whether (despite random assignment) differences between the treatment and control groups might have contaminated these results. To do so we estimated two logistic equations: one for reported turnout and one for whether turnout was correctly reported. In each equation we estimated the effect of the experimental question controlling for age, education, race, gender, internal efficacy, and external efficacy. After controlling for these differences among respondents, in neither equation did the experimental question produce a detectable improvement in the measurement of voter turnout. 4

As a final exercise, we considered a handful of plausible interaction effects to test whether the experimental question format might reduce over-reporting of turnout, but for only some kinds of respondents. Our logistic analysis found that the experimental question works no better for Democrats, than for Republicans; it works no better for those who cared very much about how the election to the House came out than those who did not care at all; it worked no better among the well-educated than those with few years of formal schooling. The evidence hints that the experimental question may do a slightly better job of reducing mix-reports of turnout among those with the least amount of political information, though the logistic coefficients are substantively small and fail to meet the customary hurdles of statistical significance. 5 There is also some hint that close to the election, the standard NES item works better than the experimental question. The farther away from the election one gets, the better the experimental question works relative to the standard NES item. It should be stressed that none of these effects pass the standard hurdles of statistical significance and the substantive size of the effects are small. 6

Results of the Experiment to Reduce the Over-Report of Church Attendance

The results from the question wording experiment on church attendance are no more persuasive than the findings for the experimental question on voter turnout. Respondents who were asked the experimental version of church attendance question were not less likely to report having attended religious services than were respondents who were asked the standard Gallup question:

Percent who

Question Format	Reported Attending Religious Services
Standard Gallup Question	49.3
Experimental Question	50.0

χ^2 (1 d.f.) = .027 (prob = .864)

Controlling for educational, age, gender, and racial differences among respondents does not reveal any effect of the experimental question that might have been masked by these demographic differences among respondents. The experimental question does not work any better for men than for women, for blacks than for whites, for the young, than the elderly, or for the well-educated than those with few years of formal schooling.

Summary and Recommendations

1. We can find no evidence to suggest that our experimental question designed to help respondents detect source confusion reduces over-reports of voting. Nor does our experimental question reduce reports of attendance at religious services. We think these findings are fairly robust. We doubt that further statistical tinkering is likely to uncover an effect that we have not yet detected.
2. There is one nagging concern that we may want to think more about. Maybe the process of sample selection (both what we did to select respondents and what potential respondents did to select themselves into our study) created a pool of subjects for whom differences in question wording are not going to matter. Those not registered to vote, who do not have phones, or who recently moved were excluded from the study by design. The validated turnout of participants in the study was 15 points higher than the validated turnout of those in the sample who did not take part (and had valid phone numbers), suggesting a hefty selection process.
3. The most obvious and straightforward recommendation is to do nothing (for the moment). We have no evidence that should be interpreted as suggesting that NES should scrap its standard turnout question in favor of the experimental item.
4. Although we failed in our effort to devise a better voter turnout question, there is still additional work to be done with these data. Remember, for the first time we have an extremely reliable measure of validated vote, one that can distinguish mix-reporting from poor record keeping and meager interviewer detective skills. This means that we are better equipped to understand the correlates of misreporting in ways not possible with the earlier NES validation studies. For example, preliminary analysis of these data shows that African Americans, the poor, those with the fewest years of formal education, and those with low external efficacy are most likely to mix-report having voted. These findings are exactly the opposite of those that have been reached by those who have analyzed the various NES voter validation studies (e.g. Silver, Anderson, and Abramson 1986). More analysis needs to be done here -- work that might provide some hints about what NES's next move should be.
5. The search should continue for the mechanisms that produce over-reports in voter turnout. Earlier rounds of experiments ruled out social desirability; this experiment rules out source confusion. Although we do not know for sure what the next move should be, we do think that R&D work on the problem of over-reporting of turnout should remain a high priority for NES. We need to think about other mechanisms that might be producing over-reports in turnout. We need to consider undertaking further R&D work in 1996.
6. The Board was prudent in its recommendation to undertake this experiment on a small scale. Small scale R&D efforts should continue. It's in the nature of research and development work that some ideas will pan out and others will not.

APPENDIX A

NES TURNOUT REPORTING EXPERIMENT 1994: Treatment and Number of Cases by Form

Form	N	Turnout Question	Church Attendance Question
Form 1	182	Standard NES	Standard Gallup
Form 2	161	Standard NES	Experimental version
Form 3	170	Experimental version	Standard Gallup
Form 4	187	Experimental version	Experimental version

Turnout Questions

Standard NES Question: In talking to people about elections, we often find that a lot of people were not able to vote because they weren't registered, they were sick, or they just didn't have time. How about you -- did you vote in the elections this November? (N=343)

Experimental Question: In talking about elections, we sometimes find that people who thought about voting actually did not vote. Also, people who usually vote may have trouble saying for sure whether they voted in a particular election. In a moment, I'm going to ask you whether you voted on Tuesday, November 8th, which was [day(s)/week(s)] ago. Before you answer, think of a number of different things that will likely come to mind if you actually did vote this past election day; things like whether you walked, drove, or were driven by another person to your polling place [pause], what the weather was like on the way [pause], the time of day that was [pause], and people you went with, saw, or met while there [pause]. After thinking about it, you may realize that you did not vote in this particular election. [Pause]. Now that you've thought about it, which of these statements best describes you? [IWER:: READ STATEMENTS IN BOXES 1-4 to R] (N=357)

1. I did not vote in the November 8th election.
2. I thought about voting this time but didn't.
3. I usually vote but didn't this time.
4. I am sure I voted in the November 8th election.
7. (VOLUNTEERS) I VOTED BY ABSENTEE BALLOT.

Church Attendance Questions

Standard Gallup Question: Now I would like to ask you whether or not you happened to attend church or synagogue (this/this past weekend. Did you yourself attend church or synagogue (this/this past) weekend? (N=352)

Experimental Question: Lots of people tell us that they attend religious services quite regularly, but that this weekend, for some reason or another, they could not attend. In a moment, I am going to ask you whether you attended religious services this weekend. Before you answer, think of a number of different things that will likely come to mind if you actually did attend religious services; things like whether you walked, drove or were driven by another person to the service [pause]; what the weather was like on the way [pause]; if someone you expected to see was not there [pause]; and the particular hymns or prayers that took place. [pause]. After thinking about it, you may realize that you did not go this weekend, but you can remember another earlier weekend in which you did attend. [pause]. Now that you've thought about it, which of these statements best describes you? (IWER:: READ STATEMENTS ..1-4 to R] (N=348)

1. I did not attend religious services this weekend.
2. I thought about going this weekend but didn't.
3. I usually go, but didn't this weekend.
4. I am sure I attended religious services this weekend.
5. (VOLUNTEERED) I went on an earlier weekend.
6. (VOLUNTEERED) I never go to religious services

References

- Abelson, R. P., Loftus, E. F., & Greenwald, A. G. (1992). Attempts to improve the accuracy of self-reports of voting. In J. M. Tanur (Ed.), *Questions about questions* (pp. 138-153). New York: Russell Sage Foundation.
- Anderson, R. E. (1984). Did I do it or did I only imagine doing it? *Journal of Experimental Psychology: General*, 113, 594-613.
- Belli, R. F., Lindsay, D. S., Gales, M. S., & McCarthy, T. M. (1994). Memory impairment and source misattribution in postevent misinformation experiments with short retention intervals. *Memory & Cognition*, 22, 40-54.
- Foley, M. A., Durso, F. T., Wilder, A., & Friedman, R. (1991). Developmental comparisons of explicit versus implicit imagery and reality monitoring. *Journal of Experimental Child Psychology*, 51, 1-13.
- Foley, M. A., & Johnson, M. K. (1985). Confusions between memories for performed and imagined actions. *Child Development*, 56, 1145-1155.
- Johnson, M. K., Foley, M. A., Suengas, A. G., & Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General*, 117, 371-376.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114, 3-28.
- Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, 88, 67-85.
- Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory and Cognition*, 17, 349-358.
- Presser, S. (1990). Can context changes reduce vote over reporting? *Public Opinion Quarterly*, 54, 586-593.
- Schooler, J. W., Gerhard, D., & Loftus, E. F. (1986). Qualities of the unreal. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 171-181.
- Silver, B. D., Anderson, B. A. & Abramson, R. P. (1986). Who overreports voting? *American Political Science Review*, 80, 613-624.
- Zaragoza, M. S., & Lane, S. M. (in press). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Endnotes

1. The vote reports of respondents in the 1972 and 1974 Election Studies were validated in the 1976 exercise.
2. Such commercial services charge by the name. Because our voter lists contained thousands of names, it was obviously preferable to subsample before the matching procedure began. The telematch service did the random subsampling. However, they did not save the exact procedure they used to generate the subsample. At this point, therefore, we do not have a good subset of "unmatched" numbers -- i.e., of citizens with unlisted telephone numbers, or who have moved away or died.
3. The relationship of household to respondent is not perfectly one to one. Because a random algorithm was used to subsample from the voter lists, we had some households in which more than one registered voter fell into the sample. An attempt was made to interview each of the selected respondents in the household.

4. The parameter estimate in the equation for reported turnout was .02 (with a standard error of .27). The parameter estimate in the equation for a correct report of turnout was -.15 (with a standard error of .42).

5. Here we estimated two logistic equations: one for reported turnout and one for whether turnout was correctly reported. In each equation we estimated the effect of the experimental question controlling for political information and the interaction between political information and the experimental question. In the equation for reported turnout, the parameter for the interaction term was 1.21 (with a standard error of .95). The parameter estimate in the equation for a correct report of turnout was -1.45 (with a standard error of 1.30).

6. Once again we estimated two logistic equations: one for reported turnout and one for whether turnout was correctly reported. In each equation we estimated the effect of the experimental question controlling for the date of interview and the interaction between date of interview and the experimental question. In the equation for reported turnout, the parameter for the interaction term was -.03 (with a standard error of .02). The parameter estimate in the equation for a correct report of turnout was .04 (with a standard error of .03).