

Question Wording and the House Vote Choice: Some Experimental Evidence



Janet M. Box-Steffensmeier; Gary C. Jacobson; J. Tobin Grant

The Public Opinion Quarterly, Vol. 64, No. 3. (Autumn, 2000), pp. 257-270.

Stable URL:

<http://links.jstor.org/sici?sici=0033-362X%28200023%2964%3A3%3C257%3AQWATHV%3E2.0.CO%3B2-G>

The Public Opinion Quarterly is currently published by Oxford University Press.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/oup.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

QUESTION WORDING AND THE HOUSE VOTE CHOICE

SOME EXPERIMENTAL EVIDENCE

JANET M. BOX-STEFFENSMEIER
GARY C. JACOBSON
J. TOBIN GRANT

Abstract Since 1978, the vote reported for House incumbents in the American National Election Studies (NES) has been significantly higher than the actual incumbents' vote in the districts surveyed; in NES surveys before 1978, the reported vote was much closer to the actual vote. The prime suspect for the source of this bias is the new question format introduced in 1978 and used in all subsequent studies. We document the problem and review the results of several question-wording experiments that confirm the superior accuracy of a format that does not mention the candidates' names over the ballot format currently in use. We also find evidence that a modified version of the ballot format may reduce the pro-incumbent bias, so that improvement may be possible without a major interruption of the post-1978 NES times series.

The American National Election Study (NES) was overhauled in 1978 to enhance its coverage of congressional elections. Before the overhaul, the vote for House candidates reported in the NES surveys generally matched the district level vote quite satisfactorily. No systematic bias favoring one set of candidates—incumbents or, more generally, winners—appeared in survey responses. From 1978 onward, however, respondents have consistently overstated their support for House incumbents. Table 1 shows the magnitude and persistence of the problem. Prior to 1978, the reported vote and the actual district vote rarely differed by more than a percentage point or two; since

JANET M. BOX-STEFFENSMEIER is at Ohio State University, GARY C. JACOBSON is at the University of California, San Diego, and J. TOBIN GRANT is at Ohio State University.

Table I. Question Format and the Overreport of Vote for House Incumbents

Old Question Format			New Question Format		
Year	<i>N</i>	Overreport	Year	<i>N</i>	Overreport
Postelection studies:					
1956	967	1.5	1978	761	10.8
1958	592	.9	1980	713	5.0
1960	659	1.6	1982	528	8.7
1964	806	-.2	1984	967	6.8
1966	567	5.6	1986	699	7.8
1968	708	-1.1	1988	756	8.8
1970	541	1.1	1990	528	13.7
1972	951	6.4	1992	1,061	10.0
1974	582	2.1	1994	694	4.9
1976	848	1.5	1996	906	8.2
1956-76	7,221	1.9	1978-96	7,614	8.5
Senate election study:					
			1988, unweighted	1,433	8.8
			1988, weighted ^a		8.1
			1990, unweighted	1,330	10.6
			1990, weighted ^a		12.5
1992, unweighted	950	-1.1			
1992, weighted ^a		-1.8			

NOTE.—The overreport is calculated by taking the mean of the difference between the vote for the House incumbent reported in the survey and the actual district vote, weighted by the number of respondents in each district. See Jacobson and Rivers (1993).

^a Weighted by the number of congressional districts in the state.

1978, the reported vote for House incumbents in contested districts has been, on average, 8.5 percentage points higher than the actual district vote.¹

The systematic overreport of votes for incumbents since 1978 is ironic and disconcerting, because the whole purpose of redesigning the NES was to improve understanding of congressional elections, and the House vote is obviously a key variable in this endeavor. In this article, we examine how and why this problem has arisen and consider what can or should be done about it. In the first section, we examine three changes made as part of the 1978 overhaul that could have inadvertently led to the overstating of votes for incumbents: sampling, question order, and the wording and presentation of the vote question. The altered vote question turns out to be the prime suspect. We then report some split sample experiments from three 1996 surveys that

1. The NES data are from Warren E. Miller and the National Election Studies, *American National Election Studies Cumulative Data File, 1952-96* [computer file], 9th ICPSR version (Ann Arbor: University of Michigan, Center for Political Studies [producer], 1998; Ann Arbor, MI: Inter-University Consortium for Political and Social Research [distributor], 1998).

confirm these suspicions. Next we consider why the question format in use since 1978 produces a pro-incumbent bias. Finally, we conclude that although both formats induce errors, the older NES format probably produces less distortion. Fortunately, a slightly modified version of the newer question format shows promise of reducing the pro-incumbent bias and could be adopted without radically altering the NES time series initiated with the 1978 study.

The Source of Bias

Initially, it appeared that the pro-incumbent bias in the reported House vote might be the result of an unfortunate sample. Looking to explain the 10.8 point overstatement of incumbent support in the 1978 survey, Jacobson (1981) noticed that districts with strong incumbents and weak challengers had been oversampled, accounting for at least some of the discrepancy between the actual vote share won by incumbents and that reported in the survey. Optimistically, he attributed the problem to the (bad) luck of the draw and expected it to disappear on its own. The problem did not disappear. When Eubank and Gow (1983; Gow and Eubank 1984) found a clear pro-incumbent bias again in the 1980 and 1982 surveys, the sample of districts could no longer be blamed, for the sampling frame was changed after 1980.

A second possible source of the bias was the addition to the survey instrument in 1978 of a large battery of new questions about the incumbent's activity. Gow and Eubank noted that the increase in the reported vote for incumbents between the pre- and post-1978 surveys occurred primarily among voters who identified with the challenger's party and who were so poorly informed that they could not recall the name of either candidate. They also pointed out that the survey asked a number of questions dealing with the incumbent's activities prior to asking the vote question. They concluded that hearing and responding to these questions had prompted some of the uninformed respondents to "remember" voting for the incumbent when they had not (Eubank and Gow 1983). Eubank and Gow proposed to solve the problem by asking the vote question before questions about the incumbent's activities. The NES Board of Overseers agreed. In 1984, the vote question was asked before any of the candidate questions except name recall and recognition (derived from the thermometer scale). Since 1984, the only questions about candidates asked before the vote question are those that refer to both candidates equally (name recall, feeling thermometers, likes and dislikes, and the contact battery); all questions referring exclusively to the incumbent are asked after the vote question.

Unfortunately, changing the question order did not solve the problem either. Overreporting of the vote for House incumbents has continued at the same level, as the data in table 1 demonstrate. If anything, the pro-incumbent bias has been worse since the question ordering was revised after 1982, rising

from an average of 8.2 percentage points in 1978–82 to 8.8 percentage points subsequently. It was not significantly lower in 1984, when potential contamination from the other questions was minimized, than in the other post-1976 studies. In 1990, overreporting reached a disconcerting peak of 13.7 percentage points.

The third possible source of bias is the change in the wording and presentation of the House vote question. Prior to 1978, the vote question was:

How about the vote for Congressman—that is, for the House of Representatives in Washington? Did you vote for a candidate for Congress? [IF YES] Who did you vote for? Which party was that?

The votes of respondents who did not name one of the district's candidates correctly were determined by the "which party?" question.

From 1978 onward, respondents have been handed a ballot card listing the

BALLOT CARD

For the November 1994 General Election

State: Georgia

Congressional District: 01

Democratic
Party

Republican
Party

CANDIDATES FOR THE
U.S. HOUSE OF
REPRESENTATIVES:

Raymond Beckworth

Jack Kingston

CANDIDATES FOR THE
STATE GOVERNOR'S
OFFICE:

Zell Miller

Guy Millner

Figure 1

candidates and their parties. Figure 1 reproduces, as an example, the sample ballot card used in Georgia's 1st District in 1994. Respondents are then asked:

Here is a list of candidates for major races in this district. How about the election for House of Representatives in Washington. Did you vote for a candidate for the U.S. House of Representatives? [IF YES] Who did you vote for?

In telephone surveys attempting to replicate the ballot format, the candidates' names and parties are read to the respondent.

The logic of switching to the ballot card seemed compelling. The new question format was intended to reproduce more faithfully the situation in the voting booth, where the names and parties of the candidates are right there in front of voters as they make their choice. The innovation was expected to increase the accuracy of reported voting behavior. Instead, it was followed by a substantial overstatement of the vote for incumbents in every subsequent survey.

Additional evidence that question wording was the source of the problem was provided serendipitously by the NES's Senate Election Study (SES).² The SES consisted of surveys taken in 1988, 1990, and 1992 of statewide samples in every state with a Senate race. Interviews were by telephone. Respondents were asked, in addition to an extensive list of questions about the Senate candidates, how they voted in House contests. In 1988 and 1990, the vote question was intended to replicate the ballot form as closely as is possible over the telephone,³ and it appears to have had the same unfortunate effect. As the lower section of table 1 reveals, the vote for House incumbents reported in the 1988 and 1990 Senate Election Studies is just as overstated as that reported in the in-person postelection studies. For 1992, however, redistricting made it impossible to list the names of the House candidates. There was not enough time to match telephone exchanges with the new congressional districts in many states, so candidates' names could not be given in the vote question. Thus the survey asked the vote question in its pre-1978 form, which makes no mention of names. Congressional districts could be identified after the survey was completed, making it possible to add the appropriate district-level information on the candidates and election results. Writing before the 1992 data were available for analysis, Jacobson and Rivers (1993) predicted that the 1992 SES would show little overreporting of the

2. Warren E. Miller, Donald R. Kinder, Steven J. Rosenstone, and the National Election Studies, *American National Election Study: Pooled Senate Election Study, 1988, 1990, 1992* [computer file], 2d release (Ann Arbor: University of Michigan, Center for Political Studies [producer], 1993; Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1993).

3. The telephone version of the House vote question used in the Senate Election Study for 1988 and 1990 was phrased this way: "I'm going to read a list of candidates for the major races in your district. In the election for the House of Representatives, the ballot listed [rotate: read names and parties of all house of representatives candidates]. Did you vote for a candidate for the U.S. House of Representatives? [IF YES] Who did you vote for?"

vote for House incumbents; they were right (see the SES entry for 1992 in table 1).

Experimental Evidence

To supplement the “natural” experiments provided by temporal changes in NES and SES question-wording formats, we conducted three question-wording experiments during the 1996 elections, using the Ohio Union Study, National Black Election Study (NBES), and Texas Post Election Survey.⁴ Each survey used a split sample design in which respondents were randomly assigned into two groups. Half of each sample was asked the pre-1978 (no names) version of the vote question. The other half was asked the ballot format version. The results of these experiments are reported in table 2.

The experiments confirm that, just as the NES and SES survey results suggested, the post-1978 ballot format produces a higher level of reported votes for incumbents. The Ohio Union Survey shows a 6.1 percentage point difference, the Texas Survey a 5.2 point difference, and the NBES a 4.5 point difference. The average difference for the three surveys is 5.3 percentage points, quite close to the NES result. With the relatively small number of cases available for analysis, differences of this magnitude cannot reach statistical significance, but the consistency of the results is nonetheless impressive. Because the split sample design controls for all other potential influences, the differences in reported incumbent support can be attributed only to differences in the survey questions.

Sources of Bias in the Ballot Question Format

Why does the ballot form bias the reported vote for House incumbents? Wright (1993) concluded that the new question wording triggered a “bandwagon” effect, drawing the more uncommitted and uninformed voters toward support of the election’s winner. Wright argued that “the ballot format does not fully replicate the polling booth. The respondent has been exposed to post-election media and conversations. These signal the winner as well as provide new

4. The Principal Investigators (PIs) for the Ohio Union Study were Herb Asher and Randall Ripley, Ohio State University; the sample was predominantly union members; the response rate was 73 percent. The PI for the National Black Election Study was Katherine Tate, Ohio State University; the NBES sample was predominantly African Americans; the response rate was 65 percent. The PI for the 1996 Texas Post Election Survey was Robert Stein, Rice University; the Texas survey was of voters; the response rate was 64 percent. The National Science Foundation grant (number SBR-974014) was used to provide partial funding for the 1996 Texas Post Election Survey. The authors are grateful to all of them for their contributions to this study. The Ohio Union Study and the NBES do not provide national or statewide samples, so the appropriate comparison is between results of the alternative ballot formats, not between reported and actual results at the district level.

Table 2. Overreport of Vote for House Incumbents by Question Format

	Pre-1978 Format	Post-1978 Format	Difference	<i>p</i> -value ^a
National Election Study (1956–96)	1.9 (7,221)	8.5 (7,614)	6.6	.00
Ohio Union Study (1996)	-2.5 (178)	3.6 (174)	6.1	.19
Texas Survey Experiment (1996)	1.6 (144)	6.8 (127)	5.2	.33
National Black Election Study (1996)	-5.0 (296)	-.5 (266)	4.5	.27

NOTE.—The overreport is calculated by taking the mean of the difference between the vote for the House incumbent reported in the survey and the actual district vote, weighted by the number of respondents in each district; see Jacobson and Rivers (1993); the number of survey cases is in parentheses.

^a The *p*-value is for the significance of the difference between the estimates of overreport of the incumbent's vote generated from the two ballot question formats.

information about all the candidates. Respondents unable to reliably recall their earlier vote choice must reconstruct that evaluation, and this reconstruction is based on information currently in memory” (Wright 1993, p. 298). The ballot form thus promotes a “bandwagon” effect; winners get more (and more positive) attention, inducing some uninformed voters mistakenly to recall voting for them.

Wright's bandwagon explanation implies that uninformed voters should overreport voting for all winners, not merely incumbents (Wright 1993, p. 305). But the data on winners of open seats tell a different story. The change in the vote question had no effect in open seat races; in the NES studies, the vote for winners of open seats was overstated by 2.0 percentage points for the 1956–76 period and was exactly on target for the 1978–96 period. No evidence of a bandwagon effect is discernible for these contests.⁵

If not a bandwagon effect, what is it? Jacobson and Rivers (1993) argued that the pro-incumbent bias is produced by the ballot form itself without

5. In fairness to Wright (1993), there is evidence that in 1994, victorious Republican candidates did enjoy a postelection bandwagon (Box-Steffensmeier and Jacobson 1995). We also examined whether being interviewed longer after the election induced more pro-incumbent responses. In both the Texas survey (which was completed within a short time after the election) and the NBES (which interviewed into January 1997) the probability of giving a pro-incumbent response was not related to the date of the survey. Jacobson and Rivers's (1993) analysis of the NES surveys also found no evidence that the timing of the survey affected the degree of overreporting of votes for incumbents.

needing any help at all from postelection publicity. The ballot card simply changes the information available to respondents through the interview process itself. The old form of the vote question contains only one cue for respondents who cannot remember the name of the candidate they voted for: party. The ballot card gives them two cues: party and name. One name belongs to an incumbent who is recognized by a large majority of respondents, even those who cannot recall his or her name spontaneously. In elections since 1978, only 47 percent of voters have been able to recall the incumbent's name unaided, but 93 percent could recognize it on a list—including 87 percent of those who could not initially recall it. Far fewer recalled the challenger's name (18 percent), and only 43 percent of those who did not recall the name could nonetheless recognize it; the overall recognition rate for challengers is only 52 percent. Thus if some voters are prompted to "remember" their vote choice by a familiar name—a prompt that was not available before the ballot card—the incumbent's support will be exaggerated.

Jacobson and Rivers provided a variety of circumstantial evidence for this explanation. They showed, for example, that virtually all of the increase in overreporting could be attributed to respondents who could recall neither candidate's name. Among these voters (in elections from 1978 through 1990) 85 percent reported voting for the incumbent if they recognized only the incumbent's name, 69 percent if they recognized neither candidate's name, and 66 percent if they recognize both candidates' names. Two of our 1996 surveys included questions tapping candidate name recall and recognition, allowing us to test this explanation more directly. Table 3 reports the results.

Both experiments confirm that the question wording has no effect on the votes of respondents who recognize both candidates' names. Also as expected, voters who recognize only the incumbent's name are more likely to report voting for the incumbent under the post-1978 ballot format—by 4.6 percentage points in the Texas Survey, by 6.5 percentage points in the NBES—although with the small number of cases, the differences do not reach statistical significance. The only anomalies appear in the "recognize neither candidate" category. The number of cases from the NBES that fall into this category is so small that the results allow no systematic interpretation. The Texas Survey results are more curious. Voters who could not recognize either candidate's name were much more likely to report voting for the incumbent under the ballot format. If name recognition could not trigger this response, we are at loss to figure out what could have triggered it. We suspect that this survey may have inadvertently understated the level of recognition enjoyed by incumbents. In NES surveys, about 93 percent of voters recognize the incumbent's name; 91 percent of the 1996 NBES respondents recognized the incumbent; but only 75 percent of Texas Survey respondents are identified as recognizing their incumbent's name. Doubting that Texans are abnormally uninformed, we think it more likely that some of the Texas respondents in the "recognize neither candidate" category would have ended up in the "rec-

Table 3. Familiarity with Candidates and the Reported Vote for House Incumbents (Percentages)

	Pre-1978 Format	Post-1978 Format	Difference	<i>p</i> -value ^a
Texas survey (1996):				
Recognize neither candidate	56.9 (39)	77.8 (27)	20.9	.07
Recognize incumbent only	80.6 (67)	85.2 (61)	4.6	.49
Recognize both candidates	62.2 (37)	58.3 (36)	-2.9	.74
NBES (1996):				
Recognize neither candidate	80.0 (10)	60.0 (10)	-20.0	.36
Recognize incumbent only	80.0 (75)	86.5 (89)	6.5	.27
Recognize both candidates	57.9 (37)	57.8 (36)	-.1	.87

NOTE.—The number of survey cases is in parentheses.

^a The *p*-value is for the significance of the difference between the estimates of the incumbent's vote generated from the two ballot question formats.

ognize incumbent only" category had the survey been administered in identical fashion to the other surveys we examine.⁶ In any case, the split-sample experiments leave little doubt that the ballot form induces some relatively uninformed voters to mistakenly recall voting for the incumbent.

In sum, then, the ballot format evidently exaggerates the incumbent's support because people are far more likely to recognize (and thus have their uncertain memories prompted by) the incumbent's name than by the challenger's name. One testable implication of this view is that if one were to examine the choices of only those respondents who could name the candidate they voted for without being given the candidates' names, the incumbents' support should be even *more* exaggerated, because respondents are so much more likely to be able to recall an incumbent's than a challenger's name.

6. Unlike the NES and NBES studies, Texas respondents were not given a large battery of the thermometer scales that are used to measure name recognition unobtrusively; they were asked only to rate the two House candidates on the 100-point thermometer scale. Thus they got the question "cold," without the usual easy identifications (the president, vice president, and so forth) to stimulate their political memories.

Table 4. Reported Vote for Incumbents under the Pre-1978 Question Format (Percentages)

Year	Voting for Incumbent			Overreport for Incumbent		
	Who?	Which Party?	Combined	Who?	Which Party?	Combined
1970	73.6 (367)	54.9 (204)	65.9 (571)	9.8 (367)	-11.9 (204)	1.0 (571)
1972	80.4 (536)	58.1 (422)	69.9 (958)	16.2 (536)	-5.8 (422)	5.9 (958)
1974	72.6 (394)	49.3 (201)	64.5 (595)	10.3 (394)	-15.4 (201)	1.5 (595)
1976	74.0 (651)	51.3 (197)	67.3 (848)	7.8 (651)	-16.4 (197)	.8 (848)
1970-76	75.4 (1,948)	54.4 (1,024)	67.3 (2,972)	11.1 (1,948)	-10.9 (1,024)	2.6 (2,972)

SOURCE.—1970-76 American National Election Studies. The “Combined” column in this table does not match the equivalent figures in table 1 exactly because they were derived from different versions of the NES data sets. The data in this table are from data files for specific years; those in table 1 are from the cumulative data file.

NOTE.—“Who?” lists the distribution of votes determined by the first part of the question (“Who did you vote for?”); “Which party?” lists the distribution of votes determined by the second part of the question (“Which party was that?”) if not ascertained by the first question; the “Combined” column lists the vote as determined by both questions together; the number of cases is in parentheses.

This is exactly what we observe in the pre-1978 surveys. Table 4 shows how the votes of respondents were distributed depending on whether they were determined by the first (“Who did you vote for?”) or the second (“Which party was that?”) part of the vote question. Consistently, the format picks up a disproportionate share of incumbents’ voters when the “who?” question is asked, then restores the balance with a disproportionate share of the challengers’ voters when the “which party?” question is asked. When both responses are combined, the result usually matches the actual vote quite closely. The second part of the question, then, allows the party prompt to kick in, redressing the imbalance created by differential familiarity with the candidates. Both questions probably induce errors of unknown magnitude—the first through the incumbent’s greater salience (leading to an overstatement of the incumbent’s support), the second through the voter’s partisanship (leading to an overstatement of support for challengers of the respondent’s party)—but if so, the errors are of similar size, thus offsetting one another (Jacobson and Rivers 1993).

By this logic, the ballot form should have the most impact on the challenger’s partisans, because they are much more likely than the incumbent’s partisans to recognize the other party’s candidate but not their own. Table 5 confirms this expectation with data from both the NES surveys and the 1996

Table 5. Partisanship and the Reported Vote for House Incumbents (Percentages)

	Pre-1978 Format	Post-1978 Format	Difference	<i>p</i> -value ^a
NES (1970–96):				
Challenger's partisans	34.7 (1,658)	44.7 (2,787)	10.0	.00
Incumbent's partisans	90.9 (2,178)	93.4 (4,284)	2.4	.02
Texas survey (1996):				
Challenger's partisans	40.5 (37)	47.4 (38)	6.9	.56
Incumbent's partisans	94.4 (72)	92.9 (56)	-1.5	.73
NBES (1996):				
Challenger's partisans	4.4 (45)	30.8 (52)	26.4	.00
Incumbent's partisans	94.7 (94)	98.9 (87)	3.2	.18

NOTE.—Number of observations is in parentheses.

^a The *p*-value is for the significance of the difference between the estimates of the incumbent's vote generated from the two ballot question formats.

experiments.⁷ The question format has at most a very small effect on the reported vote of the incumbent's partisans, but the ballot format increases reported defections from the challenger's partisans by from 6.9 to 26.4 percentage points, depending on the data set. The ballot question format thus leads to an underestimate of party loyalty in House elections as well as an overestimate of support for incumbents. If, for example, the ballot format exaggerates defections by challengers' partisans by 10 percentage points, we can calculate that party loyalty in elections since 1978 was actually 3.4 percentage points higher than the NES surveys indicate (82.2 percent rather than 78.8 percent).

7. We confined analysis to the post-1964 NES surveys so that the comparison is not contaminated by the notorious increase in the incumbency advantage that occurred, by all accounts, in 1966. See Jacobson (1997, chap. 3).

What Is to Be Done?

The implications of these results are disconcerting. On the one hand, we have evidence that the pre-1978 format produces reported voting patterns that match the actual vote much more closely than does the new format (although it produces errors of its own that are only netted out). On the other hand, the NES now has an eleven-election time series employing the new format, and breaking the series by reinstating the old format would render cross-time comparisons thoroughly suspect. The 1994 elections forcefully remind us of the knowledge that could be lost were format-induced changes commingled with real behavioral changes.

The 1994 NES actually tested one potential solution, though the test has not, to our knowledge, been previously evaluated. If the overreport of votes for incumbents arises from the ballot form's emphasis of the name cue at the expense of the party cue, it might be possible to reduce the bias by giving the party label greater prominence on the ballot card. To test this possibility, the 1994 sample was randomly divided into two groups, each receiving either the traditional ballot card or a revised ballot card. An example of the traditional ballot card is shown in figure 1. The revised card differed by printing the candidates' party labels in bold, italicized letters in a different font directly below their names (see fig. 2). The results, reported in table 6, are quite promising, although statistically somewhat inconclusive because of the small sample.

The overreport of votes for incumbents in contested districts was 6.7 percentage points under the old ballot format, compared with only 3.3 percentage points under the revised format highlighting the party label. Substantively, the improvement is considerable, reducing the pro-incumbent bias by more than half. In addition, the overreport under the new ballot format is not significantly different from the actual result, while under the old ballot format, it is ($p = .01$). Although the difference between the accuracy of the results using the two formats, 3.5 percentage points, does not reach conventional levels of statistical significance ($p = .31$), even a .69 probability of getting a true reduction in bias by adopting the revised ballot card makes the change attractive.

Moreover, in comparisons to other election years, the revised format produces an overreport significantly smaller than that of all of the 1978–96 studies taken together ($p = .02$), while the 1994 overreport under the old ballot format does not ($p = .46$). Compared with each biennial survey individually, the revised format shows a significant reduction in bias compared to seven of the nine other post-1978 surveys, failing to improve only on the two with the smallest pro-incumbent bias, 1980 and 1984. Under the old ballot format, the 1994 result is significantly less biased only compared to 1990, the survey with the greatest bias in the series. These findings suggest that, at the very least, NES ought to repeat this experiment, for the revised ballot form offers

BALLOT CARD

General Election of November 1994

STATE: Georgia
CD: 01

CANDIDATES FOR U.S. HOUSE OF REPRESENTATIVES

Raymond Beckworth
Democrat

Jack Kingston
Republican

CANDIDATES FOR STATE GOVERNOR

Zell Miller
Democrat

Guy Millner
Republican

Figure 2

a way to improve accuracy without a major interruption of the time series. The effect of emphasizing party should also be tested for telephone surveys, for example, by putting party before name: "Did you vote for the *Democrat*, Joe Smith, the *Republican*, Jane Jones, or someone else?"

Conclusion

The results of our survey experiments and reanalysis of NES and SES data persuade us that any future congressional election study that is not part of a time series should avoid the new question format in its present form. Of the two, we consider the old question superior; although it also induces some erroneous responses, the errors are not systematically biased in favor of incumbents or winners, at least in House elections. However, the value of the post-1978 time series, along with NES's current practice of including in the study a panel component that carries a subsample of respondents from one election's survey to the next, argues strongly against reinstating the old question format regardless of its greater accuracy in measuring the vote. But the promising results of the 1994 experiment suggest that the problems created

Table 6. Results of the 1994 NES Ballot Experiment

	Incumbent's Share of Votes (%)			<i>p</i> -value ^a
	Actual Results	Survey Results	Difference	
Old ballot format (names emphasized)	62.3	69.0 (339)	6.7	.01
New ballot format (party emphasized)	63.4	66.7 (327)	3.3	.15

NOTE.—Includes districts contested by both major parties only; number of survey cases in parentheses.

^a The *p*-value is for the significance of the difference between the estimates of the incumbent's actual vote share and the vote share estimated by the alternative ballot formats.

by the ballot form might be reduced if party is emphasized on the ballot. A change of this dimension would be considerably less disruptive of the time series, though scholars using the data to examine trends in incumbent support or party loyalty or ticket splitting would have to be sensitive to the change. A replication of the 1994 experiment would determine if the 1994 result was a fluke and, if it was not, would provide a second data set for examining the consequences of adopting the revised ballot format for subsequent studies. We strongly recommend that NES undertake this replication.

References

- Box-Steffensmeier, Janet, and Gary C. Jacobson. 1995. "Question Wording and the House Vote: Some Experimental Evidence." Prepared for delivery at the annual meeting of the Southern Political Science Association, Tampa, FL, November 1–4.
- Eubank, Robert B., and David John Gow. 1983. "The Pro-Incumbent Bias in the 1978 and 1980 National Election Studies." *American Journal of Political Science* 27:122–39.
- Gow, David John, and Robert B. Eubank. 1984. "The Pro-Incumbency Bias in the 1982 National Election Study." *American Journal of Political Science* 28:224–30.
- Jacobson, Gary C. 1981. "Congressional Elections 1978: The Case of the Vanishing Challengers." In *Congressional Elections*, ed. Louis Sandy Maisel and Joseph Cooper. *Sage Electoral Studies Yearbook* 6:219–47.
- . 1997. *The Politics of Congressional Elections*. 4th ed. New York: Longman.
- Jacobson, Gary C., and Douglas Rivers. 1993. "Explaining the Overreport of Votes for Incumbents in National Election Studies." Paper delivered at the annual meeting of the Western Political Science Association, Pasadena, CA, March 18–20.
- Wright, Gerald C. 1993. "Errors in Measuring Vote Choice in the National Election Studies." *American Journal of Political Science* 37:291–316.