Issue Attitudes and Survey Continuity across Interview Mode in the 2000 NES

Brian J. Fogarty

Department of Political Science,
University of North Carolina at Chapel Hill,
Chapel Hill, NC 27599-3265
e-mail: fogarty@email.unc.edu

Nathan J. Kelly

Department of Political Science,
University at Buffalo, The State University of New York,
Buffalo, NY 14260-4120
e-mail: nkelly@buffalo.edu

H. Whitt Kilburn

Department of Political Science,
University of North Carolina at Chapel Hill,
Chapel Hill, NC 27599-3265
e-mail: wkilburn@email.unc.edu

Can researchers draw consistent inferences about the U.S. public's issue attitudes when studying survey results from both the in-person and telephone interview modes of the 2000 National Election Studies (NES) survey? We address this question through an analysis contrasting the distribution of issue attitudes across modes in the dual sample design of the 2000 NES. We find clear differences across mode even when applying a method devised by the NES to improve comparability by recoding issue attitude scales from the in-person mode. We present an alternative method of recoding these scales, which substantially improves comparability between modes. Through an analysis of the covariance structure of the issues and simple models of vote choice, we discuss the implications of the results for the study of issue attitudes in the 2000 NES.

1 Introduction

One of the central purposes of large academic surveys, most notably the General Social Survey and the National Election Studies (NES), is to provide valid and consistent measures of major social trends. In an alteration that is likely inconsistent with this purpose, the 2000 NES departed from the presidential election series by relying on telephone interviews for approximately half of the usual number of respondents. Previous research has identified several mechanisms that can make survey data gathered in person incomparable to data collected over the telephone. One source of this incomparability is that telephone and in-person surveys sample from different populations. Not all individuals are part of a telephone-owning household, and telephone owners are wealthier, more educated, older, more likely to be female, and more often white than are nonowners (Klecka and Tuchfarber 1978; Wolfe 1979; Mulry-Liggan 1983; Thornberry 1987; Gfroerer and Hughes 1991). There is also evidence that telephone respondents are more likely than in-person respondents to report socially desirable attitudes, offer no opinion, and acquiesce to survey questions (Leeuw and van der Zouwen 1988; Johnson et al. 1989; Krosnick 1991; Aquilino 1994; Holbrook et al. 2003). So there could be inherent incomparability in the mixed-mode design of the 2000 NES due to differences in sampling and the quality of the survey response.

Researchers interested in studying survey results from the 2000 NES must confront another related concern. Not only does the 2000 NES make use of two interview modes, but the question formats vary across mode for most issue attitude questions. For example, when interviewed in person, survey respondents are presented with the following question about jobs and a good standard of living, along with the card in Fig. 1:

> Some people feel the government in Washington should see to it that every person has a job and a good standard of living. Suppose these people are at one end of a scale, at point 1. Others think the government should just let each person get ahead on their own. Suppose these people are at the other end, at point 7. And, of course, some other people have opinions somewhere in between, at points 2,3,4,5, or 6. Where would you place yourself on this scale, or haven't you thought much about this?

In the telephone mode, however, the question is presented in a branching format with two follow-up questions:

> Some people feel the government in Washington should see to it that every person has a job and a good standard of living. Others think the government should just let each person get ahead on their own. Which is closer to the way you feel or haven't you thought much about this?

Next, depending on the response to the first question, survey respondents were asked:

> Do you feel strongly that the government should see to it that every person has a job and a good standard of living, or not so strongly?

Or they were asked:

> Do you feel strongly that the government should just let each person get ahead on their own, or not so strongly?

(See the appendix for the wording of other issue attitude questions).

What should be clear from this example is that respondents interviewed in person were given a standard seven-point response scale on a cue card. Those interviewed over the phone were presented with the question in branching format, which combines to only five response categories. In general, while either modal or question wording differences in the 2000 NES are likely to lead to serious comparability concerns, the wording and modal differences present an unusual challenge for cross-sectional analysis and survey comparability over time.
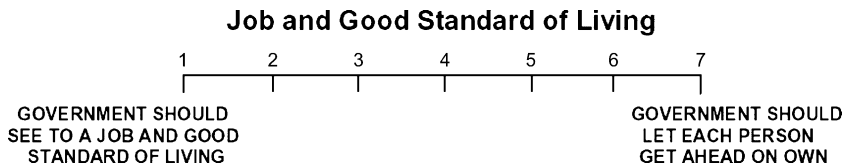
Job and Good Standard of Living

Fig. 1 National election studies (NES) show card for in-person respondents.

While we are interested in the broad implications of differences in survey mode for social research, the mixed mode and format in the 2000 NES prevents an analysis that isolates the effect of survey mode on issue attitudes. Our primary goal in this article, therefore, is to address the practical issue of the comparability of issue attitude data gathered using the two interview modes and question formats. Are the seven-point attitude scales used in the in-person survey comparable to the five-point scales in the telephone survey? Is there a reasonable way to recode the in-person scales to produce comparability? Are inferences from multivariate analyses using issue attitudes from the 2000 NES conditional on question format and survey mode? All of these questions are of practical importance to scholars who wish to conduct cross-sectional or time series studies of political attitudes with the 2000 NES while at the same time maintaining the largest possible sample size.

Our analysis focuses on the measurement of attitudes toward six issues at the heart of politics in the United States: government spending on services in general, attitudes toward government intervention in the economy to raise standards of living, public versus private sector provision of health care, spending on national defense, access to abortion, and government intervention to improve the socioeconomic position of African-Americans. We examine self-reported attitudes and where possible the perceptions of the parties and presidential candidates. While there is little hope based on previous studies that measures of issue attitudes can be readily compared across mode (Jordan et al. 1980; Calsyn et al. 1992; Wessel et al. 2000; Holbrook et al. 2003), the evidence we present shows that such comparisons are better using some approaches than others. We suggest an approach that enhances comparability compared to the approach incorporated into the publicly distributed dataset for the 2000 NES. Our goal in this paper is to provide specific guidance to users of the NES in how to manage the mixed-mode and mixed-format design of the survey.

2 Comparability across Survey Mode

One test of the comparability of issue attitude measures across mode is to examine the two attitude distributions. If the two survey modes are comparable, then no statistically discernable differences in attitude distribution should exist between the two groups. For five policy issues, part A of Table 1 reports the arithmetic mean and standard deviation—two statistics fundamental to statistical inference—across each mode, with the in-person issue scales transformed to a five-point scale using two different transformation schemes. The first column, labeled "Telephone," lists the mean and standard deviation across each issue for telephone respondents. These statistics serve as a basic point of comparison for the transformed in-person scales. The second column, labeled "NES Recoding," presents these statistics for the in-person attitude measures transformed from seven- to five-point scales by collapsing the three middle response categories into a single value. This, in fact, is the recoding scheme that is used in the publicly distributed 2000 NES data file when summary measures including all respondents are reported. Once the in-person data are

Table 1   Interview mode and the distribution of issue attitudes in 2000

| | A. Format Variation across Survey Mode | | |
Issue	Telephone	In-Person (NES Recoding)	In-Person (Alternative Recoding)
Services and spending	−.41 (1.30)	−.19* (.92)*	−.36 (1.30)
Health insurance[a]	−0.14 (1.77)	−.13 (1.14)*	−.16 (1.49)*
Jobs and standard of living	0.64 (1.64)	.30* (1.05)*	.57 (1.41)*
Defense spending	.60 (1.28)	.25* (.85)*	.57 (1.21)
Aid to blacks	.49 (1.49)	.44 (1.04)*	.64* (1.33)*
	B. No Format Variation across Survey Mode		
Issue	Telephone		In-Person
Abortion[b]	−.42 (1.06)		−.36 (1.13)

Note. Cell entries are mean respondent self-placement with standard deviation in parentheses, coded in the direction of conservatism. The NES recoding scheme creates a five-point scale by collapsing the three middle responses of the seven-point (1 to 7) scale. Our alternative recoding collapses the two most extreme values on each end of the seven-point scale to create five response categories. All issues are then centered at 0 and range from −2 to +2. Differences in means across survey mode (telephone versus in person) are tested using two-tailed mean comparison tests, assuming unequal variance. Differences in variance are tested using the variance ratio (F-distributed) test.
[a]Health insurance is the combined standard and experimental (reversed response scale) form.
[b]Abortion is measured on a four-point (−1.5 to +1.5) scale that is identical across survey modes.
*Mean (SD) is statistically different from the "Telephone" column, $p < .05$.

transformed to a five-point scale, all of the issues in part A of the table are recoded such that the middle category is zero, scores range from −2 to +2, and positive values indicate more conservative attitudes.

The initial contrast between these two columns provides little grounds for optimism regarding the comparability of the two survey modes. All of the five issues for which question format varies across survey mode show significant differences between telephone and in-person respondents in either mean or variance. The differences in means are quite large and the variance is consistently lower for the NES recoded seven-point scales. The significant reduction in variance can be accounted for by the way in which the NES recoding scheme collapses the three middle values of the seven-point scale—a part of the scale that produces a large portion of the variance—into one category.[1]

There is, however, a more theoretically sensible way to recode seven-point scales for comparison across format. Given that the general attitude direction is the first decision made by a respondent in the branching format, we believe it makes more sense to collapse the two most extreme response categories on each end of the scale. This recoding scheme preserves a finer measurement of attitude direction while placing less importance on extremity of attitude.

[1]In addition to comparability, the lack of distinction between respondents who favor one side over another raises a number of problems for assessing issue preferences, particularly from the standpoint of the directional theory of issue voting (Rabinowitz and MacDonald 1989). For directional theorists, the midpoint of a scale is extremely important. A key piece of information needed from individual voters in directional theory is the direction of their preference, knowing whether a respondent is on one side or the other of an issue. The NES recoding scheme thus masks variation that is of central importance to those testing hypotheses rooted in directional theory.

The third column of Table 1 shows the mean and standard deviation of in-person attitude measures scaled with our alternative recoding scheme. In comparison across the two modes, the alternative coding performs much better than the NES scheme. The mean differences across mode are almost negligible and the formal tests of difference show no statistically significant variation in means, with the exception of attitudes toward government aid to blacks. This issue, in fact, is the only one to display a significant difference in both mean and variance. While there are statistically discernable differences in variance for three of the five issues rescaled with our alternative recoding scheme, our approach approximates the results from the telephone mode better than the NES coding strategy.

Part B of Table 1 is particularly important. It displays the mean and standard deviations for attitudes toward abortion access. For this issue attitude, measured across mode in identical question and response format, there are no significant differences in mean or variance. It seems likely, based on the results in Table 1, that the distributions of issue attitudes measured over the phone as opposed to in person are generally comparable. The primary challenge in the 2000 NES is the divergence in question format across mode. We should also point out, however, that there are certain issue domains and types of attitudes (such as racial attitudes) that may be particularly sensitive to survey mode, and no coding strategy could or should disguise those differences. Nevertheless, when telephone and in-person issue attitudes are coded in a theoretically appropriate way, evidence suggests that they show reasonably high levels of comparability with regard to basic summary statistics.

We take a more detailed look at attitude distributions in Table 2. This table compares the reported attitude distributions of telephone respondents to in-person respondents using each of the two recoding strategies. The χ^2 test reported in the final column tests the null hypothesis that there is no difference between the modes. The results in this table generally corroborate those reported earlier; significant differences exist, but these differences are reduced by using our recoding strategy as opposed to that provided by the NES. This is illustrated by the consistently smaller values for the χ^2 statistic for our alternative recoding. In fact, our strategy eliminates statistically discernable differences across mode for attitudes toward defense spending.

What these results add to the comparison of means and standard deviations is a better sense of source of difference across question format in the 2000 NES. The main hurdle to comparability across format appears to be that respondents are much less likely to report neutral and moderate attitudes when queried with questions in the branching format compared to the visual scale of the in-person mode. Yet the NES recoding scheme collapses the three middle categories of the seven-point scale. As Table 2 clearly shows, the NES recoding produces higher estimates of neutral attitudes than in the telephone survey. Our recoding scheme also reports more moderate issue sentiment, but to a smaller degree than the NES strategy.

3 Comparing Party and Candidate Placement across Survey Mode

To this point we have discussed reports from citizens about their own attitudes. These attitudes, of course, are only one area of interest to political scientists. Theories of vote choice and party competition usually model citizen perceptions of candidate and party issue positions. So we turn to respondent perceptions of Gore, Bush, and their respective parties across interview mode in the 2000 NES. Table 3 reports the mean and standard deviation of citizen perceptions of candidate and party issue positions.

Consistent with the results for respondent self-placement, the NES coding scheme produces statistically significant differences across mode for perceptions of both

Table 2 Distribution of issue attitudes in 2000 across survey mode

Issue	Respondent Self-Placement						
	−2	−1	0	1	2	N	Chi-Square (df)
Services and spending							
Phone	22%	36%	14%	16%	11%	646	
In-person (NES recoding)	11%	13%	63%	7%	5%	856	357.09 (4)*
In-person (alternative recoding)	25%	22%	29%	12%	12%	856	70.70 (4)*
Health insurance							
Phone	41%	8%	7%	11%	33%	684	
In-person (NES recoding)	17%	11%	52%	10%	11%	935	424.52 (4)*
In-person (alternative recoding)	27%	17%	21%	14%	21%	935	120.78 (4)*
Jobs and standard of living							
Phone	22%	6%	8%	14%	50%	686	
In-person (NES recoding)	8%	5%	50%	21%	15%	909	445.46 (4)*
In-person (alternative recoding)	13%	10%	19%	21%	37%	909	89.00 (4)*
Defense spending							
Phone	9%	10%	25%	24%	32%	585	
In-person (NES recoding)	4%	4%	64%	17%	10%	836	237.44 (4)*
In-person (alternative recoding)	8%	9%	27%	26%	29%	836	6.48 (4)
Aid to blacks							
Phone	14%	17%	16%	13%	40%	637	
In-person (NES recoding)	6%	4%	52%	17%	21%	924	273.47 (4)*
In-person (alternative recoding)	10%	9%	27%	16%	38%	924	50.50 (4)*

Note. Chi-square is nonparametric, testing independence of telephone responses from the two in-person recoding schemes.
*$p < .05$

candidates and parties in nearly all issue means and standard deviations, with the exception of defense spending for Gore and the Democratic Party.[2] Looking at our alternative recoding scheme, the situation is less dire. We find statistically significant differences in the mean and standard deviation for roughly half of the issue comparisons. Our scheme generally produces smaller differences in means and standard deviations across question format than the NES strategy. Furthermore, in cases in which our coding scheme does not eliminate statistically discernable differences between telephone and in-person data, it produces smaller differences in mean and standard deviation than the NES scheme in every case. While we do not offer a singular solution to the mixed-mode problem, these results again suggest that the alternative recoding method enhances the comparability of modes in the 2000 NES.

4 Covariance Comparisons

Beyond comparisons of means and standard deviations, however, lies an interest in how the covariance structure of the issue attitudes compares across mode. To examine the

---

[2]In 2000, no candidate or party placements are available for attitudes toward health care, while the NES does not ask respondents for party positions on access to abortion.

Table 3 Issue placement of candidates and parties across telephone and in-person modes

	Gore			Bush		
Issue	Telephone	NES Recoding	Alternative Recoding	Telephone	NES Recoding	Alternative Recoding
Services and spending	−1.01 (1.11)	−0.52* (.81)*	−0.99 (1.11)	.47 (1.40)	.16* (.77)	.40 (1.27)*
Jobs and standard of living	−0.78 (1.58)	−0.17* (.86)*	−0.37* (1.30)*	1.08 (1.43)	.43* (.84)*	.82* (1.21)*
Defense spending	−0.02 (1.11)	.02 (.70)*	.06 (1.23)*	1.07 (1.09)	.44* (.80)*	.90* (1.13)*
Aid to blacks	−0.88 (1.30)	−0.27* (.89)*	−0.52* (1.28)	.67 (1.37)	.36* (.83)*	.72 (1.15)*

	Telephone Mode		In-Person Mode	Telephone Mode		In-Person Mode
Abortion	1.61 (.91)		1.80* (.99)*	2.86 (.88)		2.87 (.85)

	Democratic Party			Republican Party		
	Telephone	NES Recoding	Alternative Recoding	Telephone	NES Recoding	Alternative Recoding
Services and spending	−1.07 (1.10)	−0.56* (.83)*	−1.08 (1.06)	.65 (1.34)	.30* (.81)*	.57 (1.26)
Jobs and standard of living	−0.86 (1.52)	−0.23* (.89)*	−0.49* (1.33)*	.13 (1.33)	.53* (.85)*	.94* (1.16)*
Defense spending	−0.09 (1.15)	.00 (.73)*	−0.04 (1.28)*	.96 (1.15)	.42* (.77)*	.86 (1.15)
Aid to blacks	−0.94 (1.29)	−0.33* (.85)*	−0.65* (1.26)	.71 (1.37)	.46* (.83)*	.87* (1.14)*

Note. Cell entries are mean respondent placement of candidates and parties with standard deviation in parentheses, coded in the direction of conservatism. The NES recoding scheme creates a five-point scale by collapsing the three middle responses of the seven-point scale. Our alternative recoding collapses the two most extreme values on each end of the seven-point scale to create five response categories. All issues are then centered at 0 and range from −2 to +2. Differences in means are tested using two-tailed mean comparison tests, assuming unequal variance. Differences in variance are tested using the variance ratio (F-distributed) test. No candidate or party placements were included in the 2000 NES survey for attitudes toward health insurance.

*Mean (SD) is statistically different from the "Telephone" column, $p < .05$.

Table 4   Fit statistics for measurement models across mode

	A. Tests of Model Fit (χ^2)	
	χ^2 value (df)	p value
Unconstrained Models		
Telephone mode	.43, (4)	.98
In-person mode, NES recoding	17.71, (4)	.001
In-person mode, alternative recoding	16.92, (4)	.002
Constrained Models		
In-person mode, NES recoding	253.87, (4)	< .00
In-person mode, alternative recoding	49.84, (4)	< .00
	B. χ^2 Difference Tests (Constrained minus Constrained)	
In-person mode, NES recoding	236.15, (5)	< .00
In-person mode, alternative recoding	32.13, (5)	< .00

Note. In the constrained models, factor loadings are constrained to the estimates for these parameters in the telephone mode. See note 4 for a brief definition of the χ^2 distributed test of model fit.

structure, we estimated a close-fitting factor analytic model for five issues in the telephone mode.[3] Next, we estimated the same model parameters for both recoding strategies. To compare model fit between the telephone mode and recoding strategies, Table 3 displays conventional χ^2 test statistics.[4] Part A of Table 4 shows that while the model fits well in the telephone mode, each of the recoding strategies fail to achieve a good fit, although the alternative coding strategy fits just slightly better than the NES strategy.

Yet in order to assess the preferability of either recoding strategy, the parameter estimates of these models—most importantly the factor loadings—can be compared to the estimates in the telephone mode. The lower portion of Table 4 provides such an assessment: a test of the difference in χ^2 for the models estimated from each recoding strategy (labeled

---

[3]A maximum likelihood estimator, based on the variance-covariance matrix of five issues (excluding abortion, because it is measured identically across mode), was used to estimate loadings on two factors, an inter-factor correlation and uncorrelated disturbances. The parameter estimates, and standard errors in parentheses, are as follows.

For telephone mode (N = 384), loadings on Factor 1: jobs and standard of living 1.17 (.11) with disturbance 1.11 (.21), aid to blacks .80 (.09) with disturbance 1.54 (.14); loadings on Factor 2: services and spending .90 (.08) with disturbance 1.09 (.12), defense spending .60 (.08) with disturbance 1.32 (.11), and health insurance 1.11(.10) with disturbance 1.94 (.20); inter-factor correlation .75 (.07).

For in-person mode (N = 669) NES recoding, loadings on Factor 1: jobs and standard of living .84 (.07) with disturbance .34 (.11), aid to blacks 44. (.05) with disturbance .82 (.05); loadings on Factor 2: services and spending .47(.05) with disturbance .62 (.05), defense spending .20(.04) with disturbance .63(.04), and health insurance 57 (.06) with disturbance .90 (.07); inter-factor correlation .76.

For in-person mode (N = 669) alternative recoding, loadings on Factor 1: jobs and standard of living 1.01 (.08) with disturbance .85 (.13), aid to blacks .87 (.08) with disturbance 1.97 (.14); loadings on Factor 2: services and spending .80 (.06) with disturbance 1.09 (.09), defense spending .36 (.06) with disturbance 1.27 (.07), and health insurance .87 (.07) with disturbance 1.44(.12); inter-factor correlation .74.

[4]This χ^2 statistic is a test of fit for overidentified models. The statistic, $(N-1)F$, where N is the number of observations and F is the value of the fitting function evaluated at its final (minimized) parameter estimates, is asymptotically distributed as χ^2 with $1/2(v)(v+1) - t$ degrees of freedom, where t is the number of free parameters and v is the number of observed variables. The test evaluates the null hypothesis $H_o: = \Sigma = \Sigma(\Theta)$, where Σ is the population covariance matrix of observed variables and $\Sigma(\Theta)$ is the value of the matrix as a function of the model parameters Θ. Beyond overall fit, the χ^2 statistic can be differenced between nested models as a likelihood ratio test (Bollen 1989).

"unconstrained") compared to the χ^2 for each of these models in which the factor loadings are constrained to be equal to the loadings in the telephone mode. A significant χ^2 difference would indicate that either recoding strategy fails to recover the factor loadings observed in the telephone mode. As displayed in Part B of Table 4, both difference tests are highly significant.

These results support the conclusion that there is a statistically significant difference in the covariance structure of issue attitudes collected via the telephone versus in-person modes. Nevertheless, the covariance structure of the data collected in different modes would have to match almost perfectly in order produce a χ^2 difference small enough to be statistically insignificant, and in relatively large samples, very small differences in χ^2 produce statistically significant test statistics. It is important, then, to look not only at the statistical significance of χ^2 differences, but also at their absolute magnitude. Larger χ^2 differences indicate greater discrepancies in the covariance structure across survey mode, and the results in Table 4 show that the NES recoding scheme produces a χ^2 difference of over 236, while our alternative produces a difference of just over 32. This is clear evidence that, while our coding scheme does not produce an identical covariance structure across question format, it does perform far better than the NES strategy.

5   Mode and Coding Implications for Vote Choice and Candidate Evaluation

Another way to examine the impact of mode is to examine the issues in a regression analysis. We are particularly interested in two questions. First, do differences in covariance structure across mode affect the conclusions one draws about the influence of issue attitudes on vote choice and feeling thermometers? Second, does the NES or our alternative recoding strategy produce more comparable results across question format? We answer these questions by combining data gathered via telephone with in-person data and include dummy interactions to compare the impact of issue attitudes across survey mode.

Table 5 includes all five issue attitudes as predictors of vote choice and a combined candidate feeling thermometer ranging from -100 (pro-Gore) to 100 (pro-Bush). In addition to the five issue attitudes, the right-hand side of each model includes a dummy variable for survey mode (1 equals in-person respondents) and an interaction between this mode dummy and each issue. Every column in the table therefore reports two coefficients for each issue as well as a single coefficient for the mode dummy.

The main effect for each issue is the estimated effect that it has in the data gathered over the telephone and in person (with either recoding method), while the dummy interaction coefficient associated with each issue attitude is the amount by which the main effect changes with only the recoded in-person data. As an example, for services and spending on vote choice, neither the NES in-person coding method (.113 coefficient with a .102 standard error) nor the alternative in-person coding method (.052 coefficient with a .087 standard error) is expected to significantly change the main effect of services and spending on vote choice for in-person respondents compared with the combined samples.

We clearly see in Table 5 that in all ten interaction comparisons—five for the vote choice model and five for the candidate feeling thermometer model—not one is statistically significant. This suggests that either coding method for the in-person sample provides similar results in a regression context. Although this is the case, we know from the mean and variance analysis and measurement model analysis that our recoding method produces a better behaving covariance structure when we compare it to the telephone sample.

6   Conclusion: Survey Analysis across Mixed-Mode Designs

In the two major NES studies conducted since 2000, the 2002 study relied solely on telephone interviews and, in the resumption of the presidential election survey in 2004,

Table 5 Covariance regression models of vote choice and candidate thermometers

Issues	Probit Vote Choice Models		OLS Candidate Thermometer Models	
	Telephone & NES In-Person	*Telephone & Alt In-Person*	*Telephone & NES In-Person*	*Telephone & Alt In-Person*
Services and spending combined	.297*** (.068)	.296*** (.068)	9.31*** (1.65)	9.31*** (1.60)
Services and spending dummy interaction	.113 (.102)	.052 (.087)	1.59 (2.41)	−.152 (2.02)
Health insurance combined	.128** (.050)	.128** (.050)	1.77 (1.26)	1.77 (1.22)
Health insurance dummy interaction	.067 (.077)	.003 (.068)	3.31 (1.92)	2.05 (1.63)
Jobs and standard of living combined	.228*** (.062)	.228*** (.062)	3.31* (1.47)	3.02* (1.42)
Jobs and standard of living interaction	−.068 (.094)	−.079 (.082)	2.22 (2.25)	.533 (1.88)
Defense spending combined	.183** (.072)	.183** (.072)	8.31*** (1.63)	8.31*** (1.58)
Defense spending interaction	.045 (.111)	−.014 (.092)	−1.09 (2.49)	−2.79 (2.03)
Aid to blacks combined	.129* (.062)	.129* (.063)	6.63*** (1.49)	6.63*** (1.44)
Aid to blacks dummy interaction	.096 (.093)	.048 (.082)	−1.72 (2.21)	−2.25 (1.90)
Mode dummy	−.723 (.569)	−.078 (.492)	−14.78 (11.23)	4.41 (9.37)
Constant	−1.44*** (.504)	−3.21*** (.380)	−34.48*** (11.61)	−90.35*** (5.24)

Note. The first section reports probit coefficients with standard errors in parentheses of vote choice for the 2000 NES. The second section reports OLS coefficients with standard errors in parentheses of the combined feeling thermometers for Al Gore and George W. Bush for the 2000 NES. This combined feeling thermometer ranges from −100 (strong Gore) to +100 (strong Bush). The predictor variables for each of the five issues include a combined phone and in-person (with the two codings) and a dummy interaction with phone respondents dummied out.

*$p < .05$, **$p < .01$, ***$p < .001$

interviews are being conducted solely in person (although sample size is smaller than in the past). Clearly, for compatibility across years, it is best to maintain continuity of survey mode and question format with consistent sample sizes. Hopefully, funding and survey design decisions will provide for in-person interviews in all future NES presidential election-year surveys. Regardless of what happens in the future, however, researchers cannot change the data that were collected in 2000. Those interested in studying issue attitudes using the 2000 NES data must confront the problems associated with a mixed mode and format design.

When issue attitudes are queried in multiple modes, with forced (branching) choice and seven-point scales, the results are more comparable when the issue scales are recoded with greater sensitivity to attitude direction than intensity. While recoding the issue scales does not obviate all differences across mode, it does provide a sounder basis for comparing results across mode, at least based on the data collected in the 2000 NES. This improvement, while especially pronounced in the estimation of means and variances across mode, is also consistent in the analyses of covariance structure. A further implication of this analysis is that relying on our alternative coding strategy will not fundamentally change the conclusions researchers may draw from regression analyses of issues within the 2000 NES. In general, when researchers are faced with the need to combine survey data from alternate question formats, careful attention to the coding decision is important. The alternative coding strategy we propose is decidedly better than the one provided by the NES to combine question formats across mode.

Appendix: Question Wording across Mode

Attitudes toward abortion access are tapped in identical format across mode. Of the issue attitudes we examine, the following are different across mode:

Jobs and a Good Standard of Living

In person: Some people feel the government in Washington should see to it that every person has a job and a good standard of living. Suppose these people are at one end of a scale, at point 1. Others think the government should just let each person get ahead on their own. Suppose these people are at the other end, at point 7. And, of course, some other people have opinions somewhere in between, at points 2,3,4,5, or 6. Where would you place yourself on this scale, or haven't you thought much about this?

Where would you place [Al Gore/George W. Bush/The Democratic Party/The Republican Party] on this issue?

Telephone: Some people feel the government in Washington should see to it that every person has a job and a good standard of living. Others think the government should just let each person get ahead on their own. Which is closer to the way you feel, or haven't you thought much about this? [Depending on response:] Do you feel strongly that the government should [see to it that every person has a good job and a good standard of living/just let each person get ahead on their own], or not so strongly?

What about [Al Gore/George W. Bush/The Democratic Party/The Republican Party]? Do you think [he/they] feel[s] that government should see to it that every person has a job and a good standard of living, or that government should just let each person get ahead on their own? [Depending on response:] Do you think [he/they] feel[s] strongly or not so strongly that government should [see to jobs and a good standard of living/just let each person get ahead on their own?

Services and Spending

In person: Some people think the government should provide fewer services even in areas such as health and education in order to reduce spending. Suppose these people are at one end of a scale, at point 1. Other people feel it is important for the government to provide many more services even if it means an increase in spending. Suppose these people are at the other end, at point 7. And, of course, some other people have opinions somewhere in between, at points 2,3,4,5, or 6. Where would you place yourself on this scale, or haven't you thought much about this?

Where would you place [Al Gore/George W. Bush/The Democratic Party/The Republican Party] on this issue?

Telephone: Some people think the government should provide fewer services even in areas such as health and education in order to reduce spending. Other people feel it is important for the government to provide many more services even if it means an increase in spending. Which is closer to the way you feel, or haven't you thought much about this? [Depending on response:] Should the government [reduce/increase] services and spending a great deal or only some?

What about [Al Gore/George W. Bush/The Democratic Party/The Republican Party]? Do you think [he/they] would like the government to provide fewer services in order to reduce spending or provide more services even if it means an increase in spending? [Depending on response:] Do you think [he/they] would like to [reduce/increase] services and spending a great deal or only some?

Defense Spending

In person: Some people believe that we should spend much less money for defense. Suppose these people are at one end of a scale, at point 1. Others feel that defense spending should be greatly increased. Suppose these people are at the other end, at point 7. And, of course, some other people have opinions somewhere in between, at points 2,3,4,5, or 6. Where would you place yourself on this scale, or haven't you thought much about this?

Where would you place [Al Gore/George W. Bush/The Democratic Party/The Republican Party] on this issue?

Telephone: Some people believe that we should spend much less money for defense. Others feel that defense spending should be greatly increased. Do you have an opinion on this issue, or haven't you thought much about this? [If R has opinion:] Do you feel government should decrease defense spending, increase defense spending, or is the government spending on defense about the right amount now? [Depending on response:] Should the government [reduce/increase] defense spending a lot or a little?

What about [Al Gore/George W. Bush/The Democratic Party/The Republican Party]? Do you think [he/they] believe[s] we should decrease defense spending, increase defense spending, or that government spending on defense is about the right amount now? [Depending on response:] A lot or a little?

Aid to Blacks

In person: Some people feel that the government in Washington should make every effort to improve the social and economic position of blacks. Suppose these people are at one

end of a scale, at point 1. Others feel that the government should not make any special effort to help blacks because they should help themselves. Suppose these people are at the other end, at point 7. And, of course, some other people have opinions somewhere in between, at points 2,3,4,5, or 6. Where would you place yourself on this scale, or haven't you thought much about this?

Where would you place [Al Gore/George W. Bush/The Democratic Party/The Republican Party] on this issue?

Telephone: Some people feel that the government in Washington should make every effort to improve the social and economic position of blacks. Others feel that the government should not make any special effort to help blacks because they should help themselves. Which is closer to the way you feel, or haven't you thought much about this? [Depending on response:] Should [the government help blacks/blacks have to help themselves] to a great extent or only to some extent?

What about [Al Gore/George W. Bush/The Democratic Party/The Republican Party]? Do you think [he/they] feel[s] that government should make every effort to improve the social and economic position of blacks, or that government should not make any special effort to help blacks because they should help themselves? [Depending on response:] Do you think [they feel the government should help blacks/blacks should have to help themselves] to a great extent or only to some extent?

Health Insurance

In person: There is much concern about the rapid rise in medical and hospital costs. Some people feel there should be a government insurance plan that would cover all medical and hospital expenses for everyone. Suppose these people are at one end of a scale, at point 1. Others feel that all medical expenses should be paid by individuals through private insurance plans like Blue Cross or other company-paid plans. Suppose these people are at the other end, at point 7. And, of course, some other people have opinions somewhere in between, at points 2,3,4,5, or 6. Where would you place yourself on this scale, or haven't you thought much about this?

Telephone: There is much concern about the rapid rise in medical and hospital costs. Some people feel there should be a government insurance plan that would cover all medical and hospital expenses for everyone. Others feel that all medical expenses should be paid by individuals through private insurance plans like Blue Cross or other company-paid plans. Which is closer to the way you feel or haven't you thought much about this? [Depending on response:] Do you feel strongly or not strongly that [there should be a government insurance plan/individuals should pay through private insurance plans]?

References

Aquilino, William S. 1994. "Interview Mode Effects in Surveys of Drug and Alcohol Use: A Field Experiment." *Public Opinion Quarterly* 58:210–240.

Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables.* New York: Wiley.

Burns, Nancy, Donald R. Kinder, Steven J. Rosenstone, Virginia Sapiro, National Election Studies. 2001. National Election Studies, 2000: Pre-/Post-Election Study [dataset]. Ann Arbor, MI: University of Michigan, Center for Political Studies.

Calsyn, Robert J., Laurie A. Roades, and Dylan S. Calsyn. 1992. "Acquiescence in Needs Assessment Studies of the Elderly." *Gerontologist* 32:246–252.

Gfroerer, Joseph C., and Arthur L. Hughes. 1991. "The Feasibility of Collecting Drug Abuse Data by Telephone." *Public Health Reports* 106:384–393.

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

Johnson, Timothy P., Jr. James G. Hougland, and Richard R. Clayton. 1989. "Obtaining Reports of Sensitive Behavior: A Comparison of Substance Use Reports from Telephone and Face-to-Face Interviews." *Social Science Quarterly* 70:174–183.

Jordan, Lawrence A., Alfred C. Marcus, and Leo G. Reeder. 1980. "Response Styles in Telephone and Household Interviewing: A Field Experiment." *Public Opinion Quarterly* 44:210–222.

Klecka, William R., and Alfred J. Tuchfarber. 1978. "Random Digit Dialing: A Comparison to Personal Surveys." *Public Opinion Quarterly* 42:105–114.

Krosnick, Jon A. 1991. "Response Strategies for Dealing with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5:213–236.

Leeuw, Edith D. de, and Johannes van der Zouwen. 1988. "Data Quality in Telephone and Face to Face Surveys: A Comparative Meta-Analysis." In *Telephone Survey Methodolgy*, eds. R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls, and J. Waksberg. New York: Wiley. 283–299.

Mulry-Liggan, Mary H. 1983. "A Comparison of a Random Digit Dialing Survey and the Current Population Survey." In *Proceedings of the American Statistical Association, Section on Survey Research Methods.* Washington, DC: American Statistical Association. 214–219.

Rabinowitz, George, and Stuart Elaine MacDonald. 1989. "A Directional Theory of Voting." *American Political Science Review* 83:93–121.

Thornberry, Owen T. 1987. *An Experimental Comparison of Telephone and Personal Health Interview Surveys.* Hyattsville, MD: U.S. Department of Health and Human Services, Public Health Service, National Center for Health Statistics.

Wessel, Christina, Wendy Rahn, and Thomas Rudolph. 2000. *An Analysis of the 1998 NES Mixed-Mode Design 2000* [cited 2004]. (Available from http://www.umich.edu/~nes.)

Wolfe, Lee M. 1979. "Characteristics of Persons with and without Home Telephones." *Journal of Marketing Research* 16:421–425.