

ALTERNATIVE APPROACHES TO SURVEY DATA COLLECTION
FOR THE NATIONAL ELECTION STUDIES:

A Report on
The 1982 NES Method Comparison Project

by

J. Merrill Shanks
Maria Sanchez, and
Betsy Morton

Submitted to Warren E. Miller, Principal Investigator

and

The NES Board of Overseers

INITIAL VERSION

March 15, 1983

OUTLINE

INTRODUCTION: THE NES RATIONALE FOR CATI AND METHOD COMPARISON

Potential Advantages of Telephone Surveys for Electoral Research

Method Comparison and CATI Evaluation Issues

- Differences in sample or measurement quality
- Instrument complexity and new measurement formats
- Comparability with NES time series
- Sample size, interview length, and data collection costs

DESIGN AND GOALS FOR THE 1982 METHOD COMPARISON PROJECT

Immediate Analytic Objectives

- Univariate differences between methods
- Differences in coverage, completion, and measurement
- Differences in relationships between variables
- Identifying congressional districts
- "House effects": Differences between CATI operations

Testing for differences: A Note on Statistical Procedure

A SUMMARY OF INITIAL RESULTS

Sample outcomes and response rates for the three projects

Modest (but significant) differences in social characteristics

Consistent differences in opinionation or involvement

Modest (but inconsistent) differences in political attitudes

Measurement differences and relationships between variables

Alternative ways to identify congressional districts

GENERAL CONCLUSIONS AND RECOMMENDATIONS

General similarity or comparability between the two methods

The importance of maintaining personal and telephone methods

The need for continuing analysis and experimentation on cost-quality tradeoffs for alternative data collection procedures

Organizational issues: Staffing implications of growing technical complexity; and alternatives for large sample designs

INTRODUCTION

Since the discussions which led eventually to the initial proposal for the National Election Studies, no single subject has taken more time within the NES community than the appropriateness of alternative survey designs based on telephone methods -- and the potential advantages and/or disadvantages associated with the newer systems for computer-assisted telephone interviewing (or CATI). This general interest first came sharply into focus in 1978-79, during the Board's deliberations concerning the most efficient way to study the primary election period before the 1980 presidential nominating conventions. The Board's 1980 Committee ultimately chose to recommend against using telephone-based methods for any of the major design elements for that election, because of uncertainty in the areas to be discussed below. But many of the same issues also arose in the Board's discussions of alternative approaches to research on congressional elections, based on analytic limitations imposed by the sample design used in NES' otherwise successful study of the 1978 election.

Issues concerning the design and/or data quality tradeoffs associated with telephone vs. conventional survey methods -- and the new (but still uncertain) potential of computer-assisted systems for data collection -- have therefore been central to Board deliberations throughout NES' short history as a national research facility. In 1982, these issues culminated in a Board decision to sponsor a direct comparison between the two methods in conjunction with the regular 1982 study. Although the initial

results of that study were not available until mid-February of 1983, they have already influenced many of the Board's decisions concerning NES data collection plans for 1984. Analyses of the complex set of method comparisons made possible by the 1982 project will no doubt continue for some time, both within the Board and staff and in the larger NES research community, and the results of those analyses will structure Board deliberations concerning alternative designs throughout the next grant period.

The purpose of this report is to review the background and or rationale for the 1982 Method Comparison Project, and to summarize those initial results from the Project which bear directly on the feasibility and cost-effectiveness of the new survey techniques. We also hope to encourage other researchers to exploit the (unique) methodological data produced by the project, for this report represents only the first step in what should become a continuing program of research and development on telephone- and CATI-based techniques. As we will argue, much remains to be done before NES can take full advantage of the new survey techniques involved.

Potential Advantages. In its 1983 proposal to the National Science Foundation, the NES Board has identified those substantive areas where NES surveys are likely to depart from prior studies, based on recommendations from active scholars who use NES materials on a continuing basis. In specifying the research objectives which the Board has given a high priority for the mid-1980's, the proposal also emphasizes the method- or design-

related implications of several recommendations, for they would require an at least partial shift to telephone-based methods. In general, recommendations of this sort -- that NES utilize telephone surveys and/or computer-assisted systems -- cite one or more of the following potential advantages of the new techniques:

1. With telephone-based techniques, it may be possible to achieve a major reduction in cost per interview or a major increase in sample size, through some combination of lower interviewing costs on the telephone and greater production efficiency through the use of computer-assisted telephone interviewing (CATI);
2. CATI systems make it possible to use more complicated survey instruments or measurement procedures, based on computer-assisted branching, cycling, and automatic modification of question content;
3. CATI may minimize the amount of error involved in survey data through "online" verification and/or edit checking;
4. Both centralized telephone interviewing and CATI may make it easier to control (or schedule) the administration of NES instruments to independent national samples in relatively short periods of time, in order to provide a "time sensitive" design for assessing the impact of campaign-related and other political events; and
5. Both telephone methods and CATI should facilitate the administration of NES instruments to specific subnational constituencies (such as deliberately selected states and/or congressional districts), for which the traditional NES national sampling frame and interviewing staff is impractical.

As discussed more fully in the 1983 NES proposal, the above data collection objectives have played a central role in discussions about a wide variety of future NES research objectives. Prior to 1982, however, the Board and staff were reluctant to commit significant data collection resources to the new techniques involved because of general uncertainty about the consequences of such a decision, and because of specific concerns about the tradeoffs

between cost, sample size, and data quality involved.

Possible Problems or Methodological Issues. In particular, Board discussions about telephone interviewing have for some time reflected the following concerns or reservations:

1. Telephone-based methods might lead to a loss in sample quality, based on a combination of non-coverage (of citizens living in households without telephones) and the potential for lower response rates in telephone surveys.
2. Most observers have shared a belief that the average length of telephone interviews cannot be extended to the 60 to 70 minutes required to cover the traditional NES content.
3. Many researchers have also been concerned about the absence of show cards and other visual displays in telephone interviews. NES instruments have made extensive use of such aids, so that telephone methods might lead to losses in comparability or data quality vis a vis the traditional (face-to-face) interviews. It was also recognized that in principle CATI systems might compensate for the loss of visual aids by reformulating questions into complex sequences which take advantage of automatic "branching", but no evidence existed concerning the effectiveness of such compensating capabilities for NES-type items.
4. Telephone surveys based on national samples produced by random digit dialing may not be feasible for research objectives which require that each interview be assigned to a particular congressional district before substantive questions begin.
5. The use of telephone-based surveys for studies which need probability sampling, high response rates, and long interviews is relatively new. As a result, information presented to the Board about the relative costs of telephone vs. traditional methods has been inconsistent at best. It has thus been impossible to assess the hard cost-related tradeoffs involved in any design changes for future studies.
6. This uncertainty about costs has been compounded with respect to computer-assisted telephone interviewing, for CATI systems themselves -- and detailed NES procedures for using CATI -- were still in an exploratory stage. Thus, even the cost-related results of demonstration CATI projects were difficult for NES to assess, because

the operational procedures involved had not reached sufficient maturity to permit stable cost estimates for large scale NES utilization.

During the early 1980's, however, experience with telephone interviewing and CATI systems within the academic, governmental and private sectors advanced to the point where the new techniques demanded serious consideration by NES. In particular, the Berkeley Program in Computer-Assisted Survey Methods had demonstrated in 1981 that the length of telephone interviews on political topics could be extended to 50-55 minutes, thereby overcoming what had been assumed to be a fundamental barrier to NES use of telephone surveys (See Shanks, et al, 1981). Nevertheless, it was still the case that the concerns about tradeoffs in cost and data quality for NES could not be effectively addressed by reviewing other CATI demonstration or method comparison projects.

DESIGN AND GOALS FOR THE 1982 METHOD COMPARISON PROJECT

For these reasons, the Board decided to commit all of its R&D resources for 1982 to a systematic comparison of the two methods. As originally proposed, the same NES interview schedule would be administered to two independent national random adult samples, one of which would use traditional household sampling and face-to-face personal interviewing, and the other would use an RDD sample of telephone households and computer-assisted telephone interviewing. As the project developed, it was further proposed that the CATI component for the study be equally (and randomly) divided between the two approaches to CATI system

design and production data collection being developed at Michigan and Berkeley. Given probable budgetary constraints, the specific goals for these three independent data collection projects were set at 1500 traditional (or face-to-face) interviews, 500 interviews produced by ISR's CATI facility, and 500 interviews produced by CSM's CATI facility.

With both approval and supplementary support from the Foundation, this specific design was carried out during the period following the 1982 congressional elections. The next section of this report reviews the objectives and procedures used in our first analysis of the resulting data, and the following section highlights the analytic results which have already been presented to the NES Board. A concluding section then provides a brief summary of the lessons so far obtained from the Method Comparison and CATI demonstration aspects of the 1982 study. This discussion includes our own recommendations concerning the methodological and/or technical issues which need further attention before NES can regard the 1982 design as "established" or routine, and a short discussion of the organizational and staffing issues associated with large scale NES utilization of the new (computer-assisted) techniques.

Immediate Analytic Objectives. Given the above set of general goals, answers to precoded questions from the 1982 survey were processed during the few weeks between the completion of all interviewing activity and the Board's February meeting. As of this writing, coding and editing for open-ended questions and

"other specify" responses are still underway, but method comparisons based on those materials will be incorporated in a subsequent version of this report.

In the following pages, we have summarized our early conclusions from the 1982 Method Comparison Project in six broad categories, dealing with:

- 1) univariate differences between the traditional and CATI samples, with respect to:
 - social and demographic characteristics;
 - the existence of political interest and opinion; and
 - the direction of political opinions (when present);
- 2) where possible, univariate differences between methods are discussed in terms of the contributions made by method differences in sample coverage, response rates, and measurement techniques;
- 3) differences between the traditional and CATI samples in relationships between variables;
- 4) alternative approaches to congressional district classification;
- 5) "House" effects, based on differences between the Michigan and Berkeley portions of the CATI sample; and
- 6) the impact of telephone methods and CATI on costs and study organization.

Testing for Differences: A Note on Statistical Procedure.

Previous studies provide some guidance concerning the probable direction and magnitude of any differences which might arise between telephone-based and traditional methods -- and the extent to which they may be attributed to differences in coverage, response rates, or measurement (Groves and Kahn, 1979, Freeman, et al, 1982). The NES Board and Staff, however, are primarily concerned with the combined effect of all three of these sources

(or types) of "method effects". For this reason, and because the three types of effect cannot always be separated, our initial analyses have been structured in terms of an overall null hypothesis. Thus, we have repeatedly tested the proposition that our traditional and telephone methods produced two independent random samples from the same population, or that each difference between the two 1982 samples can be attributed solely to chance.

As our first, and deliberately exploratory, analytic step, we therefore examined the relationships between type of data collection (personal vs. CATI) and substantive NES variables in each of the three areas mentioned above -- i.e. in social or demographic characteristics, in the degree to which respondents expressed political opinions or interest, and in the direction of political attitudes. For each such relationship (between method and substance), we then calculated the apparent probability that a relationship (or a departure from complete independence) that large could have arisen by chance. Since this part of the analysis has necessarily examined a very large number of potential differences between the two methods, our search for differences was carried out on only a random half of the interviews from both samples. In those cases where this exploratory procedure yielded evidence of a difference which is "significant" at the .05 level, the same calculations are then replicated on the second random half of our sample. For those variables where a significant difference persists, we then discuss the degree to which those differences may be due to the three types or sources of method differences -- i.e. to coverage (of non-telephone

households), differential completion rates, or differences by mode in the measurement of identical concepts. In some sections, we have re-examined the same relationships after restricting the analysis to those respondents who voted in 1982. Many NES analyses are necessarily limited to the active (instead of the potential) electorate, so it is essential to see if overall method differences persist when the sample is restricted in that fashion.

The above split-sample procedure was also used to test for differences between the two CATI-based survey organizations that collected the data. But "method" differences were much more numerous (and significant) than "house" differences in the 1982 project.

A SUMMARY OF INITIAL RESULTS

All three data collection projects achieved their objectives in terms of completed interviews, and the final response rates for the three projects were consistent with prior expectations. Thus, 1418 complete or partial interviews were conducted in respondents' houses, for a completion rate of 72%. The Berkeley CATI study yielded 501 complete or partial interviews, the Michigan CATI operation produced 500 complete or partial interviews, and both CATI samples represent a completion rate of 62%. Based on the information now available, all three projects adhered to their design specifications for households and respondent selection, so that the three data sets can be interpreted as indepen-

dent samples from the same population -- i.e. of adult citizens living in households during the 1982 post-election period. Any differences between the three samples that cannot be attributed to chance should therefore be attributed to one or more intentional differences in method or procedure, rather than defects in the execution of our basic design.

Method-Related Differences in Social Characteristics. Out of 11 comparisons, five fixed personal characteristics or demographic variables yielded statistically significant differences between the traditional and CATI samples. These differences -- in respondent's education, income, race, religion, and region -- were of course initially "detected" in an exploratory analysis based on a random half of both samples, but Table A (on the following page) highlights those differences for the total 1982 sample, along with selected results for variables with no method-related differences.

The relationship between data collection method and education is of course familiar from previous studies (Groves and Kahn, 1979, Freeman, et al, 1982). As in other studies, some of this difference can be attributable to the small proportion (approximately 5%) of the population excluded from telephone surveys because they live in households without telephones. Thus, when this "coverage" effect is removed by restricting the traditional (or face-to-face) sample to those respondents who live in telephone households, the magnitude of the educational difference between the two samples is reduced, but it is still clearly sig-

Table A

METHOD-RELATED DIFFERENCES IN SOCIAL CHARACTERISTICS

<u>Significant Differences</u>		<u>Total Personal Interview Sample</u>	<u>Total CATI Sample</u>
Education --	Percent with no High School	23.2	16.6
	Percent With College Degree	18.7	24.4
	Number of Cases (Base N)	(1401)	(948)
Race --	Percent Black	10.6	5.6
	Percent Non-Black	89.4	94.4
	Number of Cases (Base N)	(1402)	(945)
Household -- Income	Percent With Income Less Than \$5,000	10.7	4.4
	Percent With Income Greater Than \$35,000	18.8	24.3
	Number of Cases (Base N)	(1246)	(861)
Religion --	Percent Protestant	62.6	51.6
	Percent With No Preference	13.5	20.9
	Number of Cases (Base N)	(1398)	(954)
Region of -- Residence	Percent in Northeast	19.8	28.1
	Percent in South	36.1	33.9
	Number of Cases (Base N)	(1407)	(954)
<u>Non-Significant Differences</u>			
Union -- Membership	Percent With Union Member in Household	21.4	20.0
Marital -- Status	Percent Married, Living With Spouse	60.0	60.4
Age --	Percent Less Than 26 Years Old	14.1	16.1

nificant. In any production utilization of telephone-based methods, weight variables would presumably be used to correct for differential response rates between areas, and to correct for the disproportionate representation of better educated citizens in telephone samples. But the magnitude of education-related differences should be recognized wherever those weights are used.

The noticeable differences between traditional and CATI samples with respect to race and income present more difficult problems of interpretation, for these effects may be at least partially caused by differences in the conceptualization and measurement of the basic variables involved -- as well as sample-related differences due to non-telephone households and differential response rates. Thus, race was (necessarily) measured through a direct question asked of all respondents on the telephone, while the 1982 NES personal interviews relied on direct interviewer observation. And income was measured on the telephone through a complex series of questions which took advantage of CATI's capability for automatic branching, since the "card" which is used to show income categories to respondents in personal interviews could not be used.

Whatever the cause of these differences, however, it seems important to note that household sampling and face-to-face interviews produced a national sample of the potential electorate in 1982 with nearly twice as many blacks as a comparably sized sample produced by random digit dialing sampling and telephone interviewing. And the 1982 telephone sample under-represented low

income families in general -- and included only half as many families with income less than \$5,000 per year.

As in the case of education, these differences can -- and should -- be partially corrected through weight variables which compensate for known under-representation of specific demographic groups. In developing such a weighting procedure, however, our current evidence is less than decisive, for the national data which might be used as a standard for correction are themselves based on different data collection procedures -- i.e. on interviewer observation or self-administered questionnaires. As discussed below, we recommend that NES continue to support research and development in a variety of telephone-related areas. The whole question of demographic representation and weighting provides a good case in point, for it raises a variety of issues concerning the correct standard for re-weighting both the telephone and the personal interview sample.

Although the representative quality of telephone vs. traditional samples must be judged on the basis of total sample comparisons, it is worth noting that both the education and income differences are much weaker (and in fact fail to obtain statistical significance on a half sample basis) when the analysis is restricted to those who actually voted in 1982. Thus, as in our prior analysis of non-telephone household distinctiveness based on data from the 1980 NES study, the impact of telephone methods on sample composition would appear to be much less for the active (instead of the potential) electorate.

As shown in Table A above, the 1982 CATI sample departed from the 1982 personal interview sample (and previous NES studies) in the proportion of respondents who live in the Northeast and Southern regions of the country. As of this writing, the explanation of this regional difference is not clear, but it appears to stem from the original telephone sample, rather than from differential response rates by region. Regional differences will be removed when the data are weighted, but the weight variable to be used has yet to be developed, and will depend on an extended analysis that has only recently begun.

The apparent method-related difference in religious preference was entirely unexpected, and may be a by-product of the regional and educational differences discussed above. As in many other sections of this report, however, further analysis of this difference must await the completion of a weight variable which compensates for known differences in other social characteristics.

In summary, it seems likely that some demographic differences will persist between samples produced by traditional vs. telephone-based methods. But the effects involved are modest in nature, they are to a considerable extent concentrated among non-voters, and they can be largely removed through the utilization of appropriate weights. To reinforce this general conclusion, we have begun to compare the distributions on variables from the traditional NES sample in 1982 with comparable figures from the 1980 census.

Table B

METHOD-RELATED DIFFERENCES IN EDUCATIONAL ATTAINMENT

<u>Years of School Completed</u>	<u>1980 Census</u>	<u>1982 Personal Interviews</u>	<u>1982 CATI Interviews</u>
0-11	32%	23%	17%
4 Yrs High School	36%	34%	36%
1-3 Yrs College	17%	23%	25%
4+ Yrs College	15%	20%	22%
	-----	-----	-----
	100%	100%	100%

As shown in Table B, the difference in education between the CATI and traditional NES samples represents only an extension of an already noticeable discrepancy between traditional NES samples and Census Bureau estimates for the national adult population. In general, even political surveys carried out by academic organizations that use probability sampling and seek the highest possible response rates under-represent citizens with low levels of education and income. In making inferences about the potential electorate, these differences have posed a potential threat to correct inferences about univariate distributions -- if not about relationships between variables -- for some time, and the use of telephone methods would appear to accentuate those difficulties. Were it not for considerations of cost, timing, or geographical location, traditional survey methods (based on household sampling and face-to-face interviewing) would apparently produce a slightly more accurate portrait of the potential electorate. But these

method-related differences are generally modest, while the cost differences between the two techniques are not.

Differences in Political Interest and Opinionation. The clearest and most consistent differences between CATI and traditional methods arise when we focus on the extent to which respondents offer opinions on issues, and on the level of political interest or involvement. As before, the magnitude of those differences is not great, but 21 out of 54 comparisons made in this area between the CATI and personal half samples produced differences which were significant at the .05 level. Most of those differences were again significant in our second half sample, but (perhaps understandably) only seven remained significant when the analysis was restricted to the active electorate. Table C (on the following page) presents five of these comparisons, based on the full 1982 sample for both techniques. We are not yet sure whether the differences in Table C are solely due to the greater ease with which politically uninterested potential respondents can refuse to participate in a telephone interview. It may also be the case that telephone respondents are less willing to admit they "don't know" in response to attitudinal questions, or that probing is more effective on the telephone. But we are tempted to conclude that non-response is more sensitive to low levels of political interest on the telephone than in face-to-face (personal) interviews.

Differences in the Direction of Political Attitudes. For those respondents in both samples who did express an opinion or

Table C

DIFFERENCES IN POLITICAL INTEREST AND OPINIONATION

	<u>Total Personal Interview Sample</u>	<u>Total CATI Sample</u>
Percent with No Opinion on Gov't Spending and Services	20.4	14.4
Percent with No Opinion on the Long Term Impact on the Economy of Reagan Policies	18.3	10.3
Percent with No Opinion on the Equal Rights Amendment	7.0	2.5
Percent Who Follow Gov't and Public Affairs "Hardly at All"	14.8	8.6
Percent Who Care "Very Much" About the Outcome of Congressional Elections	20.6	24.5
Number of Cases (Average Base N)	(1400)	(950)

attitude on NES issue variables, we have also tested the hypothesis that the distribution or balance of opinion is the same in both populations -- and then checked to see if significant differences persist when the analysis is restricted to the active electorate. Out of 64 (half sample) comparisons of that sort, only 19 are statistically significant at the .05 level, and 10 of these are no longer significant when the analysis is restricted to the active electorate. Table D presents six of these significant differences, along with insignificant differences in ideology and overall approval of Ronald Reagan's performance -- in order to illustrate the modest and inconsistent nature of method-related effects in this area.

To help us in interpreting these results, we have begun to use summary variables within each category of NES questions, in order to remove some of the ideosyncratic (and/or random) differences between items. In initial analyses, variables which summarize perceptions (traits) and feelings about Ronald Reagan have not shown a significant relationship to data collection mode, but we are not yet prepared to conclude that "no differences" exist. And we are not yet certain how to summarize sample differences in policy preferences. There are significant differences in individual 7-pt scales, but the direction of those differences is inconsistent. There is no significant difference in overall ideological self-identification, and we have as yet no insights into the role played by sampling and measurement differences between the two techniques.

Table D

DIFFERENCES IN THE DIRECTION OF POLITICAL ATTITUDES

(Base N's are in Parentheses)

	<u>Total Personal Interview Sample</u>	<u>Total CATI Sample</u>
* Average Score on Ideology (ISR) (1=Liberal, 7=Conservative)	4.3 (888)	4.3 (331)
Average Score on Aid to Minorities (ISR) (1=Favor Aid, 7=Oppose)	4.4 (1186)	4.1 (399)
Average Score on Defense Spending (Berkeley) (1=Cut, 7=Increase)	3.9 (1115)	4.3 (410)
Average Score on Gov't Services (ISR) (1=Reduce, 7=Increase)	3.8 (1109)	4.2 (403)
Average Score on Gov't Jobs Responsibility (Berkeley) (1=Favor, 7=Oppose)	4.3 (1202)	4.5 (407)
* Percent Who Approve of Reagan's Performance as President	51.4 (1310)	54.3 (902)
Percent Who Approve of the Way Reagan is Handling National Defense	55.7 (1198)	61.5 (832)
Party Identification:		
Percent Who are Strong or Weak Democratic Identifiers	45.2	37.4
Percent Who are Strong or Weak Republican Identifiers	24.5 (1373)	26.0 (940)

* Not Significant at the .05 Level

Because of the branching experiment, these results are based on either the ISR or the Berkeley half-samples who were asked the questions in the conventional 7-point (non-branching) format.

Based on our results so far, we have concluded that telephone methods yielded consistently (and significantly) different distributions in political attitudes only for partisan identification and some party-related evaluations. Both the ISR and Berkeley telephone samples included fewer democrats, and this difference remains when the analysis is restricted to actual voters. As in all other comparisons in this report, we do not yet know whether this sample difference in partisanship will disappear when we re-weight the telephone sample to correct for known differences in demographic characteristics. When appropriate weight variables have been constructed, all of the above analyses will be repeated, so that we can assess the degree to which the two methods produce different results -- through mechanisms other than the telephone's under-representation of less educated and politically interested citizens.

Measurement-Related Differences. As discussed above in connection with sample differences in race and income, results of telephone surveys (and CATI studies in particular) may differ substantially from those of traditional surveys because of measurement differences associated with the loss of visual aids. In the case of questions about income and detailed contacts with congressional candidates, this change in instrumentation is unavoidable, and the revised questions for CATI represent only a natural adaptation of the same basic objectives and categories.

In the area of political issues, however, the 1982 project included an additional experiment on the consequences of dif-

ferent approaches to the conceptualization and measurement of both the respondent's position and his perception of the position held by several candidates and parties. By again taking advantage of the random allocation of cases to Michigan or Berkeley, respondents were randomly assigned to one of the two following measurement "treatments" -- with respect to ideological position, attitudes towards government services, defense spending, aid to minorities, equality for women, and governmental responsibility for jobs and standard of living:

1. In the "traditional" treatment, the "card" used in personal interviews to explain the seven point scale was simply evoked for telephone interviews as part of the question text, and respondents were asked telephone questions which closely followed the traditional NES wording;
2. In the CATI or "branching" approach, respondents were first asked a simple trichotomous question concerning their own opinion, and their answer to that question determined the (different) follow-on question used to classify each respondent into one of seven final categories. (The same two stage process was then used to capture respondents' perceptions of candidates and party "positions" on the same policy issues.) In constructing these questions, care was taken to maintain the definition of the end points on the traditional 7 point scale, but the branching format led inevitably to some changes in the definition of final response categories.

Based on the 1979 NES pilot survey, and subsequent discussions concerning the potential advantages of CATI-type branching, some researchers have anticipated improvement in the accuracy or validity of data on issue positions produced through this multi-stage process. That possibility will be addressed below in our discussion of "Differences in Relations Between Variables", but it also suggests the importance of comparing the overall marginal distri-

butions between the two approaches to measurement. We are therefore using our (random) half samples to test the hypothesis that the branching format produces a systematically different impression of the overall balance of public opinion in each attitude area. Comparisons of this sort will be discussed in greater detail in subsequent versions of this Report, but the following results are already clear:

1. There are no consistent differences between the two methods in their tendency to over (or under) represent one side of the bipolar dimensions represented by the four NES questions used for this comparison;
2. The branching format does reduce the number of respondents who remain at an indeterminate or "middle" position, although it is not yet clear if such a change represents an improvement;
3. For whatever reasons, the branching approach maximizes the variation between respondents' in their apparent position on the 7 underlying dimensions, a fact which should be kept mind when interpreting differences between methods in relations between variables;
4. We believe that branching results in a noticeable increase in the amount of missing data, possibly because the initial trichotomous question provides more encouragement for respondents to say they "haven't thought much about this" -- but this result has not yet been fully tested; and
5. The branching method takes noticeably longer to administer when placements are needed for a substantial number of candidates.

This last point deserves special emphasis, as the Board's 1984 Committee begins its deliberations concerning the multiple design elements for that year. Continuity with previous election studies may well require that the traditional 7-pt scale be continued for the Fall (time series) component of the 1984 study. But branching methods may be attractive for the pre-convention por-

tions of the study unless the instrument calls for "placements" for a large number of candidates. As of this writing, method comparisons in the issue and branching areas have only begun, but future NES Committees may well have to choose between better measures of respondent positions and issue placements for more candidates.

Relationships Between Variables. In our initial -- and exploratory -- comparisons of the CATI and traditional samples, we have also examined the strength of relationships for each method between basic political predispositions and a substantial number of attitudinal or opinion variables. In our results so far, no significant differences have been found between the telephone and traditional samples in bivariate relationships between party identification (and ideology) and several dozen other political variables. There is an inconsistent (and insignificant) tendency for relationships between party identification and specific partisan evaluations to be slightly stronger in the personal interview sample. But this faint tendency is if anything reversed in relationships between ideological self designation and issue position, and we suspect that these preliminary differences are to some extent confounded by the differences between the two samples in levels of education and by the differences between branching vs. 7-pt measures -- as discussed above.

Confirmation (or rejection) of these last possibilities will require further analysis, and these same comparisons may have to be replicated in a subsequent election before we can dismiss sam-

pling error as an explanation, for none of our differences in relationships represent more than a 15% change in the measures of association involved. In summary, our initial results suggest that NES analysts who are primarily interested in relationships between variables need not worry about the impact of telephone-based methods on their results.

Congressional District Classification. Since 1978, NES Board and Committee meetings devoted to research on congressional election have repeatedly dwelled on the potential advantages of telephone surveys for congressional research -- primarily because of the larger sample size which could be produced. But all parties involved were also concerned about the potential barrier to such an application posed by the initial ambiguity concerning the current congressional district for households selected through random digit dialing (RDD). Since the instrumentation developed for the 1978 NES study requires that detailed information about the candidates be used to automatically modify all questions about congressional elections, some method had to be found and tested to produce such information before each interview.

For this reason, one of the major objectives of the 1982 Method Comparison study was an assessment of the relative efficiency and accuracy of alternative techniques for assigning randomly selected household telephone numbers to congressional districts. On the basis of extensive staff and Board discussions, it was decided that only a "streamlined" method would be used in 1982, in which congressional districts were simply attributed to

each randomly selected household telephone number (where any possibility of ambiguity existed) based on information provided by the respondent -- so that the maximum number of interviews could be conducted with the resources available. In principle, the Board might recommend (and may do so in the future) that interviews involving some ambiguity concerning district identification be interrupted, so that geographic information may be used to identify the district involved before resuming the interview as a callback. This option was not exercised in 1982, both because of the extra costs associated with such a venture, and because detailed geographic maps were not available from the 1981 and 1982 redistricting in time to support such a technique. Instead, the 1982 interviews included questions about the geographic location of the respondent's household, and each case was ultimately associated with a "true" congressional district, based on address information and detailed maps -- but that classification took place after all the interviews were completed. Based on that effort, an analysis is now underway concerning the detailed correspondence between the apparent district (to which each respondent was attributed during the interview) and the actual district which a more careful (but expensive) procedure could (in principle) have established before substantive questioning began.

Before discussing these results, it may be helpful to review what we believe is the major important conceptual and measurement issue before the Board concerning survey measurement in the congressional area. Based on the data from prior election studies, it has been evident for some time that a substantial pro-

portion of NES respondents are unable to evaluate congressional candidates, and appear to have very little information about any of the alternatives involved. Nevertheless, previous election studies have maintained the practice of asking all potential voters for their evaluations of the candidates and for evidence of contact with both the incumbent and/or challengers. Thus, a substantial amount of interview time has been used to obtain repeated negative answers like "no contact" or no specific "likes" or "dislikes". Many analysts (and the NES staff) have recommended for some time that screening questions be introduced, so that interviewers would automatically skip all questions about congressional candidates for individuals who fail some minimal standard for candidate visibility. The question of potential comparability between telephone surveys and the traditional NES procedure rests fundamentally on the way in which the Board resolves this question.

In 1982, two different methods were used to identify congressional districts and to select the subsample of respondents for whom detailed congressional questions were asked, and these two methods are directly linked to the two alternative answers to the above general question. In the random half of the telephone (CATI) sample administered by Michigan's Survey Research Center, respondents were "attributed" to a congressional district when the names recalled implied one and only one of the districts possible for that telephone number. At the other extreme, respondents who failed to recognize any of the major party candidates in any of the congressional districts to which their

telephone cluster might belong were not asked any of the congressional questions. When the respondent failed the recall test, but gave signs of candidate recognition for one of the congressional districts -- but no such recognition in the other possible districts -- that respondent was automatically "attributed" to the district in which recognition took place and the detailed congressional questions were administered in terms of the candidates in that district. In the approach used at Michigan, only those respondents who had failed the recall test but recognized candidates in more than one district were then explicitly asked to choose between the two or three districts that might be involved. Out of 468 interviews for which the essential information is currently available, this first (Michigan) method produced a national sample in which 77 respondents (or 16%) were asked no congressional questions because they failed to recognize any of the potential candidates involved. And an additional 7 respondents (or 1%) were not asked the congressional questions because they recognized candidates in different districts, and were unable to resolve the conflict in response to a direct question.

In contrast, respondents interviewed by Berkeley CSM's staff were asked the congressional questions even if they failed to recognize any of the congressional candidates involved if they lived in single district clusters. In multi-district clusters, except in those cases where the respondent was able to select a district through unaided recall, all respondents with telephone numbers for which there was some prior ambiguity about the

correct congressional district were asked to make an explicit choice between the multiple districts involved. Thus, "recognition" information was not used to attribute each respondent to a congressional district. As a result of this procedure, out of a total of 443 interviews for which the appropriate information is currently available, no respondents were "skipped" for the congressional questions because of a failure in recognition, but 66 respondents (or 14%) were not asked the congressional questions because they were unable to select their congressional district in a direct choice between the possible alternatives.

The two methods were also different in the accuracy of the congressional districts assigned to respondents on the basis of their answers to questions. As the following table shows, the method used for Berkeley's half sample permitted a larger proportion of the sample to be asked the congressional questions, and that a higher proportion of the attributions proved ultimately to be accurate. But nearly all of the "gain" in the Berkeley sample (both in the number of respondents asked congressional questions and the accuracy of congressional district attribution) was concentrated in those respondents that were unable to recognize any of the candidates involved. Therefore, if the Board subsequently decides to systematically skip all congressional questions for respondents who are ill informed about the candidates in their district, the method of "attribution through recognition" used in the ISR half of the 1982 study would be superior.

VERIFICATION AND ASSIGNMENT OF CONGRESSIONAL DISTRICTS

Alternative CATI-Based Procedures

	True CD Equals <u>Assigned CD</u>	True CD Does Not Equal <u>Assigned CD</u>	Never <u>Assigned</u>	DC+MD
ISR Method (N=484)	69% (335)	13% (65)	17% (84)	-- (16)
CSM Method (N=461)	76% (352)	9% (43)	14% (66)	-- (18)

In general, it seems clear that the losses or errors in both 1982 methods are heavily concentrated among those respondents who have very little to say when asked the detailed congressional questions. The random assignment of respondents to two different methods of congressional district classification in 1982 does permit us to assess the kind of information lost (or gained) when one method of district attribution is chosen instead of the other. For example, we believe that the Berkeley respondents who were not asked congressional questions because they could not explicitly choose their district had (on the average) very little to say about the candidates involved, so that we expect that the two samples will be nearly identical in the overall "amount" of information collected in response to the open-ended questions about the congressional candidates. And the Berkeley results should make it clear that those respondents who were asked congressional questions (even though they failed to recognize any of the candidates involved) had very little to say in response to those questions -- or that the method used at ISR lost very little information by skipping all congressional questions for such respondents.

Differences Between Survey Organizations and CATI Systems.

Although the questions used in the two CATI-based segments of the 1982 Method Comparison Project were identical, there are differences between the two CATI systems and survey organizations involved. Although the senior staff from both organizations in the 1982 project did not expect any significant differences between results in the Berkeley and Michigan samples, one of the study's major objectives was to test the validity of that expectation. For this reason, all of the comparisons discussed above between traditional and telephone methods have been calculated for the ISR and Berkeley halves of the CATI sample. As before, because so many potential differences were being "tested", only a random half of the two CATI half samples was used for these large scale (and exploratory) comparisons.

In summary, out of 125 potential relationships between specific variables and the "house" or system which collected the data, only 19 differences appear to be significant at the .05 level. These differences, which are not large in substantive terms, involve variables as diverse as length of residence (at both the current address and city), turnout, and the frequency of "don't know" responses to attitude questions. A cursory examination of the content of these variables also suggests a slight difference in socio-economic status and attitudes towards the Reagan administration, but these differences fail to pass a test of significance which takes into account the very large number of tests involved.

Based on our results so far, it seems likely that the two CATI systems and the basic procedures used by the two survey organizations are sufficiently similar to produce national samples which are fairly difficult to distinguish. To some extent, this similarity in results may be due to the standardization in interviewer behavior imposed by the CATI systems and the detailed instrumentation used for the 1982 project. But the two organizations also share a great deal in their approach to interviewer recruitment, training, and supervision. Whatever the cause, however, the new survey technology and methodology can apparently be used to produce similar results in separate organizational settings, despite modest differences between the survey organizations and CATI systems involved.

In a subsequent version of this report, we will present the detailed results of these comparisons between CATI organizations. That presentation will also review the specific areas in which survey practices were known to differ between organizations. As background for that discussion, however, it may be helpful to summarize the major differences between the two organizations and CATI systems involved. The Michigan Survey Research Center (SRC) has maintained a telephone survey facility and professional staff for several years, but the CATI system used by SRC only recently became available for production utilization. In contrast, the Berkeley Program in Computer-Assisted Survey Methods (CSM) is a research and development group devoted to the creation of new systems and techniques, and does not maintain a production survey capability. The Berkeley system is now in use in several other

survey organizations around the country, but the CSM unit maintains staff expertise at only the senior level. As a result, Berkeley's interviewers, supervisors, and coders were recruited specially for the 1982 project, while Michigan's effort relied on a continuing professional staff for these services. These differences should be kept in mind in reviewing both the general similarity and modest differences between the "houses" involved in 1982. And these same organizational differences should be helpful in drawing conclusions from 1982 concerning the probable costs and the staffing requirements for future NES CATI utilization.

Costs and Staffing for Traditional vs. Telephone Methods.

Projections for future studies concerning the comparative costs for traditional- and telephone-based methods are still problematic in several respects. But the budgets and cost information from the 1982 project, coupled with revised cost estimates for the future, do support some conclusions about the tradeoffs in cost or sample size associated with the two alternative techniques. Thus, after appropriate adjustments have been made for both 1982 cost overruns and future cost reductions based on the lessons and instrumentation from 1982, we believe that the 1982 study will confirm that telephone-based methods can produce more than two (and less than three) times as many interviews (lasting an average of 50 minutes) as can be collected using traditional household sampling and face-to-face interviewing. This ratio of cost and/or sample size could become larger (surpassing three to

one) if comparatively short interviews were conducted with those citizens who did not vote in the most recent election. But future planning activities would appear to be on safe ground if they operated within the above range of possibilities.

Even these approximate answers are of course based on a number of operating assumptions and ambiguities, each of which deserve emphasis. In particular, the above estimates are based on the following assumptions:

- (a) Both techniques are used to collect fresh (i.e. no re-interview) random adult cross-sections of the nation's eligible electorate, with a sample size between 1500 and 2500. (Smaller studies will drive up the cost per case somewhat differentially between the two methods, but the 1982 experience makes it difficult to assess the comparative startup costs associated with relatively small projects, due to ambiguities in the division of labor between the NES study staff and the survey organizations involved).
- (b) The above estimates assume that the basic instrument design (including both the analysis-ready codebook and instructions for coding open-ended questions as well as the interviewer's instrument) are developed by the study staff or on a separate budget, so that the above estimates reflect only "production" costs associated with sampling, field, coding, and direct data entry.
- (c) None of the figures which have gone into the above comparisons incorporate any institutional overhead or indirect costs.
- (d) No attempt has been made to explicitly adjust for the loss in data associated with the shorter length of telephone interviews (e.g. from 70 to 50 minutes).

This last point deserves special emphasis. Although it might be said that telephone interviews "produce" 5/7ths as much information as the single traditional interview, we have not used that fraction to compare the costs per interview or interview

equivalent. In many contexts, we suspect that the maximum sample size which can be produced is the more important figure for planning purposes, but it should be noted that the telephone survey will continue to require some sacrifice in the number of questions which can be asked. In this connection, it also seems appropriate to point out that no attempt was made in 1982 to administer a telephone interview which exceeds an average length of 50-55 minutes. Just as the results of 1981 and 1982 CATI surveys have given support to the idea that telephone studies can approach one hour in length, subsequent pilot studies should explore the possibility of even longer interviews, on at least a subsample basis, so that future NES analysts may gain the maximum benefit from each completed interview.

The Impact of CATI. As discussed at the beginning of this report, the 1982 Method Comparison Project was structured in a way which would permit the Election Studies to assess the potential utility of CATI systems as well as telephone-based sampling and interviewing. Thus, the final 1982 data tape will contain cases that were produced by three different systems for the entry, management, and cleaning of survey information. The face-to-face interviews were of course first collected in the traditional paper and pencil form, and then converted to numeric form and cleaned using the Institute's system for direct data entry. In contrast, the data records from Michigan's telephone sample were created in machine readable form by interviewers using ISR's CATI system, and then converted to use the same procedures used for cleaning and coding the face-to-face interviews. At Berke-

The NES Board has proposed that future Election Studies include a variety of separate (or independently executed) data collection operations -- each of which would require substantial advance planning for full CATI utilization. The full implications of such a change for the permanent staff of the National Election Studies are just now becoming visible, and we urge the staff to proceed carefully, in order to maximize the integrity of each data collection involved while making an orderly transition to the new technology. In general, we recommend that future studies use CATI-type systems only if (or when) they meet the following guidelines:

1. CATI should be used only when the system and instrumentation involved have reached a level of technical stability that reduces overall costs or provides a demonstrable enhancement in the overall quality of the resulting data; or
2. When the complexity of instrumentation is such that computer-assisted techniques are essential in order to satisfy the research objectives involved; and
3. The extra investment in advance planning and instrument development required for CATI studies should not be undertaken without the prior recruitment and training of additional senior staff with the appropriate substantive and technical expertise.

GENERAL CONCLUSIONS AND RECOMMENDATIONS

The results of the 1982 Method Comparison Study suggest that telephone-based methods are feasible and cost-effective for NES. Given our current estimates of comparative costs for future studies, it is clear that the newer survey methods are generally appropriate for those NES research objectives which could only be

NES decisions with respect to CATI utilization may therefore vary considerably in future studies -- as a function of cost and design requirements. For the design elements proposed for 1984, the NES Board has decided to defer decisions concerning CATI until later in 1983, but all parties concerned are agreed that decisions in this area should be based on the degree to which measurement objectives require CATI-type capability, on the degree to which CATI offers a clear advantage in the overall cost of data collection, and on the impact of CATI on senior technical staff.

The general implications of the new survey technology for survey organization and staffing have been evident for some time. But the NES 1982 project has provided a powerful independent demonstration of the need for additional development and testing of instrumentation for computer-assisted survey data collection -- for the cleaning and coding phases, as well as interviewing itself. From the 1982 experience, it is clear that the amount and complexity of advance planning required calls for a substantial increase in the permanent (senior) technical staff assigned to direct the studies. The two CATI systems used in the 1982 Method Comparison Project were designed in such a way as to save time and costs in the later stages of survey data preparation. But the 1982 project has shown that this expectation will only be realized if sufficient staff expertise and time are available for instrument development at the "front end" of the process, and if study budgets reflect the complexities associated with the coding of open ended NES questions and the reformatting of data and documentation for ICPSR distribution.

tradeoffs associated with newer survey technology continue to change at a very rapid pace. In particular, NES staff and Board members will be engaged in continuing analyses of the 1982 study, with special emphasis on its internal experiments with respect to measurement, congressional district identification, and CATI effectiveness. And any use of a dual frame design for the traditional (or time series) component of subsequent studies will make it possible for NES to replicate basic results from the 1982 study -- and thereby monitor changes in the cost and data quality tradeoffs associated with telephone-based methods. We therefore urge the Board to allocate resources for a continuing evaluation of method differences in sample or measurement quality, and of the cost- and staff-related tradeoffs associated with computer-based techniques.

In summary, the 1982 Method Comparison Project has made it possible for NES to take a major step forward in utilizing new survey techniques to maximize the efficiency of data collection, while maintaining the maximum possible comparability with the NES time series. If appropriate attention is given to the complications posed by using two parallel systems for data collection, future studies should be able to substantially enlarge the set of research objectives that can be accommodated.

accomplished with telephone surveys, and that telephone-based methods can also serve as a valuable complement to household sampling and face-to-face interviewing for the NES time series.

This last point deserves special emphasis, for the Board is considering a dual-frame design for the traditional (Fall) surveys which extend the NES time series. In such a plan, both traditional and telephone-based methods would be employed simultaneously to collect two independent (but complementary) national cross-sections. The resulting telephone interviews would be collected at lower cost and would be used to enlarge the overall sample size for those analyses which require more cases or better coverage of specialized subpopulations. And each sample would serve to correct for known weaknesses in the other, based on the differences between methods in both coverage and response rates. As a result, the combined national sample would provide a better description of the potential electorate than has heretofore been possible. At the same time, however, time series analyses could be restricted to the personal interview-only component, in order to avoid confusing true differences with method effects in the interpretation of aggregate change over time.

Continuing Methodological and Technical Issues. These general conclusions are already reflected in the Board's long term proposals for data collection although the dual frame design is not currently scheduled for use in 1984. But it is increasingly clear that NES has only begun what must become a continuing program of method-related research, for the cost and data quality

REFERENCES

Freeman, Howard E., K. Jill Kiecolt, William L. Nicholls, and J. Merrill Shanks, "Telephone Sampling Bias in Surveying Disability, Public Opinion Quarterly, Fall 1982, 392-407.

Freeman, Howard E., and J. Merrill Shanks, ed, "The Emergence of Computer-Assisted Survey Research", a forthcoming special issue of Sociological Methods and Research.

Groves, Robert M. and Robert L. Kahn, Surveys by Telephone, Academic Press, 1979.

Shanks, J. Merrill, Wm. Michael Denney, J. Stephen Hendricks, J. Merrill Shanks, Wm. Michael Denney, J. Stephen Hendricks, and Richard A. Brody, "Citizen Reasoning About Public Issues and Policy Tradeoffs: A Progress Report on Computer-Assisted Political Surveys", submitted to the Russell Sage Foundation, October 1981.