
How the ANES Used Online Commons Proposals and Pilot Study Reports to Develop Its 2008 Questionnaires

Jon A. Krosnick and Arthur Lupia

THE ANES SEEKS to produce questionnaires that allow scholars to explain vote choice, turnout, and related topics. When we led the ANES from 2005 to 2009, we chose to construct ANES questionnaires by consulting as many scholars as possible. In chapter 2, we outlined how our goal of broad consultation led us to develop the ANES Online Commons.

In response to our invitation to participate in the Online Commons (OC), hundreds of scholars proposed new ideas. In this chapter, we describe how these ideas informed our decisions about which questions to include in the questionnaires for the ANES 2008–2009 Panel Study and the 2008 edition of the ANES Time Series Study.

Two sets of scholarly products were particularly influential in our decision-making processes. The first set consists of the proposals that were submitted during the initial use of the OC. These proposals provided ideas for the 2006 ANES Pilot Study, and a number of them are described in greater detail in the other chapters of this book. The second set included Pilot Study Reports written by OC participants and other scholars in 2007 and 2008. The ANES used these reports to evaluate the potential value of pilot study questions for the questionnaires that we were developing for the 2008 and 2009 ANES surveys. A number of these Pilot Study Reports provided the initial versions of data and analyses that appear in the other chapters of this book.

Hundreds of scholars wrote OC proposals. Dozens wrote Pilot Study Reports. Many of these contributions provided important evidence about the performance of specific question wordings in an election survey environment. Other scholars contributed new and rigorous arguments to the questionnaire-development processes. The collective impact of these efforts was to transform the questionnaires that the ANES fielded in 2008 and 2009. Indeed, as a result of these efforts, ANES surveys included hundreds of new or improved questions.

Given that the ANES is a public good, it is remarkable that so many talented scholars from a wide range of disciplines took the time to contribute to the ANES questionnaire-development process. The magnitude of their contributions is especially noteworthy because scholars who wrote OC proposals and Pilot Study Reports had no privileged access to the data collected with the questionnaires that they helped develop—ANES data become available to all scholars at exactly the same moment. The social sciences are fortunate that so many scholars are willing to contribute to broad collaborations such as the ANES.

The number of questions that scholars proposed for each of the studies that the ANES ran in 2006–2009 was far greater than the number of questions that could be included in the surveys. In the case of the 2006 Pilot Study, for example, more than 1,100 questions were proposed for a survey whose time budget allowed only about 120 questions to be asked. OC proposals sent for ANES studies in 2008 and 2009 had comparable supply-demand imbalances. As a result, we had to make difficult choices about which items each survey would include.

When making these choices, our main priority was to fill ANES surveys with questions that would provide high analytic value to as many scholars as possible. Since the ANES mission makes explanations of vote choice and turnout a top priority, we followed suit. We were also interested in supporting a broader, theoretically rigorous, and multidisciplinary range of election-related scholarship, if possible. The procedures that we developed to evaluate questionnaire proposals were designed to achieve these goals. An equally important priority was to be fair to all participants and to maximize the transparency of the questionnaire-development processes. For this reason, and well in advance of the deadline dates for decisions, we developed and publicized the evaluative criteria that we would employ when making difficult choices.

In what follows, we describe the manner by which we converted the insights of many generous people into strong and innovative questionnaires. We break the description into two stages to reflect the nature of our decision-making processes.

The first stage began in 2006, two years before the general election of 2008. At that time, the OC was used to solicit ideas for the 2006 ANES Pilot Study. The purpose of that study was to create data with which scholars could evaluate how a large set of new questions would perform in an election survey. In section 2 of this chapter, we describe how we evaluated those proposals and chose which questions to ask in the 2006 Pilot Study.

The second stage began about six weeks after we ran the pilot study, when we released the data to the public. This period started at the end of 2006 and continued into the early months of 2008. During this time, we received Pilot Study

Reports. In section 3 of this chapter, we describe how we evaluated those reports and how these evaluations affected choices of which questions would appear on one or more of the 2008 studies.

The criteria that we used at each stage were similar—which is not surprising, since the main objective in both cases was to build better ANES questionnaires in 2008. But there were also important differences in the two sets of criteria. For the 2006 Pilot Study, we encouraged scholars to offer riskier proposals. For the 2008 studies, where our margins for error were narrowed and our deadlines for producing final questionnaires were quickly approaching, our appetite for such risk decreased.

Using the procedures that we describe below, we learned many important facts and lessons from a wide range of people. This advice changed our thinking about the questionnaires and improved their quality immensely. We present this information in this chapter in the hope that it will increase understanding of how questionnaire design decisions were made during the months and years leading up to the 2008 elections. We hope that this level of transparency will also stimulate further conversation about procedures and criteria by which future ANES decision makers and other survey producers can increase the analytic value of the data that they collect and distribute.

EVALUATION OF ONLINE COMMONS PROPOSALS FOR THE PILOT STUDY

Principal investigators (PIs) of national election studies, or other shared data collection instruments, face many challenges. No challenge is more daunting than that presented by the clash of resource constraints against the diverse expectations amongst potential users of ANES data. Many scholars can imagine how an enterprise such as the ANES could increase or deepen its coverage of a topic about which they care deeply. The challenge is that there are literally hundreds, even thousands, of such nascent demands. Many of these ideas, if implemented into the ANES, would surely provide value to the scientific community. But resource constraints are unforgiving. For any given dollar amount, and any particular sampling design, finite resources imply a hard limit on interview length and sample size. These limits, in turn, constrain how many topics the ANES can cover. As ANES PIs, our goal was to derive the greatest aggregate scholarly benefit available from the resources that we could acquire.

Of course, everyone who runs a survey has resource constraints. Scholars who develop surveys for the sole purpose of advancing their own research must also choose among questions that they would like to ask. The key difference between scholars who develop their own surveys and scholars who develop

unusable data. For example, a few proposals featured questions for which it was unlikely that more than a handful of respondents would say yes. While such frequencies can be valuable to know, they severely limit the usefulness of such variables to most ANES users. Of course, for people with temporal hypotheses, having no one say yes at a certain time can be quite informative—and we are willing to consider proposals in which such outcomes are possible. But we would encounter problems if such null results were expected most or every time the question is run (that is, if a “variable” does not vary, then it is limited in the power it can provide to statistical analyses).

We also wanted questions that would both strengthen interdisciplinary ties and contribute to ongoing debates:

Bridge-Building: Can the ideas proposed build intellectual bridges from one or more research traditions to others? There is no single approach to explaining turnout or vote choice. Where possible, however, we would like to challenge researchers within certain paradigms to explore the consequences of interacting with other scientific communities.

We received several proposals that brought ideas from outside traditional ways of looking at elections into the context in exciting and potentially path-breaking ways. Subject to such proposals meeting the other state criteria, we would like to see more of this.

Controversy-Relevant: Are the ideas proposed relevant to ongoing controversies among researchers, such that our adding particular questions can advance the debate?

Most of the proposals that we received attempted to satisfy these criteria. As a result, we were able to base many of our decisions on sound scientific foundations. Even in cases where we could not include the questions that a particular author proposed, their advice often helped us improve the questionnaires in other ways.

After the deadline for submitting proposals had passed, we distributed the proposals and all comments about them received on the OC to the ANES Board of Overseers. Each proposal was initially assigned to a subset of the board. This procedure ensured that each proposal would be read in its entirety by at least three board members and by both PIs. The entire board then discussed all proposals at a multiday meeting.

The process of going from the many questions proposed to the relatively few that we could include was made easier by the board's rigorous and frank assessments of each proposal. Also making things easier was the fact that some scholars proposed questions that did not need to be tested in a pilot study. For exam-

ple, some proposals sought to run certain questions from past ANES surveys. Others proposed including questions that had already appeared on other major surveys. Given that the mandate of the 2006 Pilot Study was to evaluate new questions, we could not accommodate these requests. Finally, some proposals described new questions that lacked an apparent connection to determinants of turnout and vote choice. They were not included either.

Following the meeting of the ANES board, the PIs and the ANES staff worked with OC proposers and the board to derive a final set of questions from the proposals. We spent a great deal of time during the months that followed working with authors of OC proposals that best met the criteria stated above on ways to make their question wording more effective. We leveraged emerging insights from literatures on the psychology of survey responses and optimal questionnaire design to transform many of the questions that were proposed into more effective instruments for measuring important concepts.

Because we had more good ideas than we could implement, efficiency in measurement became a focal issue. For example, several proposals asked that we invest twenty, thirty, or sixty questions in measuring a single concept. Given the fixed lengths of our survey instruments, the opportunity costs inherent in making such investments are substantial. Many concepts that are valuable to ANES users can be measured in one, two, or three questions. Hence, devoting sixty questions to a single measure would come at the expense of twenty to sixty typical ANES concepts. When there was sufficient enthusiasm among reviewers and the board for such content, we worked with proposal authors to explore whether their constructs could be measured more efficiently. When such conversations were successful, we included the resulting questions. When they were not, we had to exclude them.

Another frequent challenge for us came from proposals whose relevance to a single hypothesis was more obvious than its value to the broad ANES user community. ANES studies are not partitioned into independent playgrounds for discrete groups of scholars. Instead, a main objective of our questionnaire design procedures is to facilitate a wide range of hypothesis tests. When a questionnaire includes a hundred questions (like the 2006 ANES Pilot Study), several hundred questions (like the 2008 ANES Time Series), or more than a thousand questions (like the 2008–2009 Panel Study), millions of hypotheses can be evaluated if data from individual questions can be meaningfully mixed and integrated with others. Therefore, data from individual questions that can be used to explore a wide range of hypotheses provide more value to the ANES user community than data from questions that have a single use for a single scholar or set of scholars. Indeed, many valuable lessons have been learned

from inquiries that seek to go beyond the single-factor explanations of electoral phenomena that are so common in the mass media. The practice of asking questions that many scholars can use has allowed thoughtful researchers to rigorously evaluate many meaningful and theoretically relevant propositions.

Following the criteria and practices stated above, we used OC proposals and the advice and consultation we received from many other scholars to form the questionnaire for the 2006 Pilot Study. The questionnaire contained twenty-nine modules that covered a range of interesting topics. Table 20.1 lists the topics covered.

The resulting data proved to be very helpful in our efforts to design effective questionnaires for the 2008 election. Simply having pilot study data, however, was insufficient for determining which of the new questions tested in that study merited inclusion in one of our election year surveys. We also needed rigorous evaluations of the new data. This is why, during the months that followed the release of this pilot study data, we invited scholars to write Pilot Study Reports.

EVALUATION OF PILOT STUDY REPORTS

The OC produced proposals for questions to be asked in the 2008–2009 ANES Panel Study and the 2008 Two-Wave Time Series Study. We evaluated these proposals using the criteria listed above, plus additional criteria that were appropriate for each specific study. For our panel study, for example, we sought proposals that effectively leveraged its temporal design.

With our panel and time series studies, and given the number and range of people who use ANES data, the stakes were too high to allow the inclusion of untested questions. So we required evidence that questions performed effectively (that is, were capable of making new and significant contributions to explanations of turnout and vote choice) before they could be considered for inclusion in a 2008 or 2009 questionnaire. Some proposals that the ANES received in 2007 and 2008 (that is, after the 2006 Pilot Study data was released) provided no evidence about how those questions would be useful to researchers in an election study. In cases where the proposals had the potential to generate such value, we followed up with proposal authors in search of such evidence, or we attempted to find it ourselves. In cases where the searches were unsuccessful, we could not include the questions.

Most of the proposals that we received in 2007 and 2008 did provide such evidence. Some presented evidence from other major surveys. Some presented evidence from surveys that the researchers had run themselves. But many proposals drew on 2006 Pilot Study Reports.

TABLE 20.1
2006 Pilot Study Modules with a Brief Description of Content

Module #	Module title	Descriptive question or phrase
1	Character Judgments	Can people change who they are?
2	Defensive Confidence	Can you defend your opinions?
3	Need for Closure	A psychological measure of a desire for quick answers
4	Belief in a Just World	Do people get what they deserve?
5	Self-Monitoring	Questions about comfort with self-expression
6	Trust	Can you trust other people?
7	Values	A set of questions creates a values scale
8	Borrowing	Can you borrow money if you need to?
9	Sociotropic Voting	Questions about gas prices and unemployment
10	Religion	Questions about religious practices
11	Christianity	Questions about specific aspects of Christian theology
12	Optimism/Pessimism	Feelings about the future
13	Social Networks	With whom do you discuss politics?
14	Attention to Politics	How often do you pay attention to what is going on in government and politics?
15	Ambivalence	Do you have positive and negative feelings about government?
16	Efficacy	Questions about the abilities and perceptions of federal officeholders
17	Trust in Government	To what extent do you trust the federal government?
18	Media	Updated and improved measures of media access and usage
19	Political Party	As of today, do you think of yourself as a Republican, a Democrat, an independent, or what?
20	Identification	Questions about abortion opinions under various scenarios
21	Abortion	An experimental question about tolerance towards Muslims
22	Tolerance	Questions about whether the policy treat various groups fairly
23	Justice	Questions about women's and men's abilities to be president
24	Gender	An experiment on how to ask questions on these topics and Estate Taxes
25	Progressive Taxation	How different are Republicans and Democrats?
26	Partisan Differences	An experiment on how to ask questions on these topics
27	Voting and Turnout	Questions solicit general approval and topic-specific approval
28	Presidential Approval	Better or worse?
29	National Economy	Feelings about and likelihood of catastrophe
29	Death	

This book includes a subset of the Pilot Study Reports that were submitted. Many of the reports included analyses that showed how the 2006 Pilot Study's new questions provided improved measurement of important concepts. Hence, the Pilot Study Reports were highly influential. Indeed, they were a frequent and focal reference for us as we developed questionnaires for the 2008 Time Series Study and the 2008–2009 Panel Study. We are grateful to the authors of these reports for their effort and insight in providing rigorous assessments that helped us make a series of difficult and complex decisions about the 2008 questionnaires. This work is valuable not just for its scholarly advances, but also for its direct contributions to improved questionnaire quality.

In the remainder of this section, we document successes and failures of our attempts to use Pilot Study Reports to inform our questionnaire-development decisions. First, we describe some of the challenges we faced in obtaining sufficiently helpful Pilot Study Reports. Then, we describe how the Pilot Study Reports in the chapters of this book did—or did not—influence our 2008 and 2009 questionnaires.

CHALLENGES OF PILOT STUDY REPORTING

During the years leading up to our initial use of the OC, we spent considerable time and effort attempting to understand how it would work. We worried a lot about being fair to all proposers. We wanted to motivate honest interchange and encourage broad participation in the OC. We sought a balance between accountability for actions taken in the OC and privacy protection for people who wanted to be evaluated by their ideas and not their identities.

According to many indicators, the initial trials of the OC were successful. Many people participated. They represented various disciplines. Many of them made rigorous arguments. There were no privacy complaints.

Yet problems arose that we had not fully anticipated. These problems made the inaugural version of the OC less efficient with respect to our goal of producing high-quality questionnaires. In several cases, we received no Pilot Study Reports from question proposers who agreed to write such reports. This was a surprise to us. We had hoped that we would get a report from everyone whose proposals were accepted.

When people who promised to write Pilot Study Reports failed to do so, we sought to fill the gaps by providing new research opportunities to other scholars. During the summer of 2007, we invited scholars—including graduate students—to conduct their own analyses of the 2006 Pilot Study data in areas where we needed pilot reports. We offered as an incentive the opportunity to

include their work in this book. We received a number of interesting drafts of such papers, and we are grateful to the authors who provided them. In the end, one of these reports, “Perceptions of Similarity and Disagreement in Partisan Groups,” by Eric Whitaker and John Fulwider, is included in this volume, alongside side the chapters written by the original authors of OC proposals.

In some cases, we did not receive Pilot Study Reports because proposal authors wanted questions implemented in a way that was contrary to our goal of serving a broad user community. For example, several authors proposed that we include a set of questions that they had used in other work. A problem that we encountered more than once was that scholars proposed questions that they had used in other domains (such as surveys of experts or in laboratory experiments with college undergraduate participants) rather than in general-public samples. No doubt, many questions would be similarly understood by experts, undergraduates, and representative samples of the vote-eligible American public, but there are many other questions that do not transfer so easily across these populations. So we required evidence of effectiveness with general-population samples.

Because the vote-eligible population is considerably more diverse than many of the convenience samples or specialized populations that social scientists are accustomed to dealing with, our pilot tests of questions sometimes entailed changes from the question wordings proposed initially on the OC. In surveys of experts, for example, one can use advanced vocabulary and abstract concepts, knowing that all respondents will understand. However, many of those same words would be interpreted differently by members of the general public. Similarly, words and phrases used in studies of undergraduates are not always universally understandable by more general populations. We worked with proposers of pilot study questions on optimal wording. In almost all cases, we and the proposers saw the virtue of wording questions in a broadly accessible manner and in accordance with leading-edge theories of, and best practices in, question wording. But in a few cases, we and the authors agreed to disagree, and they declined to participate in writing a report on the changed question wording tested in the pilot study.

In other cases, we received Pilot Study Reports that were informative but not directly applicable to our criteria. For example, we received some reports that attempted to make the case for the inclusion of items based on analyses showing how the questions related to items that had little or nothing to do with turnout or vote choice. In many of these cases, the analyses were sound, but provided little help in our effort to determine whether the questions were consistent with the ANES's core mission of helping scholars explain vote choice and turnout.

In cases in which we did not receive Pilot Study Reports that evaluated the likely value of including items on an ANES questionnaire from the original question proposers or other scholars, we attempted to conduct these analyses ourselves. This practice was suboptimal—the best case for these questions could have been made by scholars who were more familiar with the theoretical underpinnings of the affected content than we were. But moving forward in this way was better than having no analysis at all.

In almost all such cases, we commissioned graduate students at our respective universities to take the first crack at analysis. We chose students who had rigorous training in statistical methods, and we asked these students to document everything they did. Our standard request of these students was to use the new content in multivariate analyses of vote choice and turnout. To the extent that the proposal or the kind of question implied a specific manner of evaluating data, we followed that lead. For example, some new questions measured constructs that are expected to moderate the relations of other variables with vote choice or turnout, so we tested moderation.

Beyond that, we asked the students to do little more. In particular, we asked the students not to go on fishing expeditions to try to find some specification, some set of control variables, or some nonobvious manner of manipulating the data to discover an effect of the questions at hand. We asked for simple multivariate analyses evaluating the questions about which we wanted to learn, and only those control variables or interactions directly implied by relevant theory. In a sense, this way of proceeding set the bar for evaluative success quite low. If a measure did not have a significant effect on vote choice or turnout in such bare-bones analytic environments, then it would likely not have such an effect when more control variables were added. Surely, this manner of proceeding led us to evaluate some questions less favorably than we would have if we had had access to broader input about the variables in question.

Some of the Pilot Study Reports that we received focused on the contribution of the items to analyses of vote choice and turnout but did not always lead us to accept the questions for inclusion in later studies. We sometimes learned that new versions of questions were not better than old versions. We sometimes learned that the added analytic value of new questions was smaller than we anticipated. We do not consider such lessons instances of failure. Rather, we are grateful that sound investigation led to valuable insights about optimal measurement and provided a stronger basis for sticking with effective past practices.

In other cases, we found Pilot Study Reports convincing and decided to include new questions in one or more of our 2008 and 2009 questionnaires. However, those questionnaires were of fixed length, and in the end, we ended

up not being able to include every question for which compelling evidence was provided. That said, all the Pilot Study Reports we received are posted on the ANES website and are, thus, available for ANES and other survey researchers to leverage in future questionnaires.

We deeply appreciate all the efforts of everyone who participated in the OC. We are particularly grateful to those OC participants, and everyone else, who took the time to write Pilot Study Reports.

CONTRIBUTIONS OF THE CHAPTERS IN THIS BOOK

All the Pilot Study Reports included in this book provide evidence about the likely value of specific questions in an ANES context. However, some of the recommendations made in the reports in this book were not implemented in the 2008 studies. Of the eighteen Pilot Study Reports included in this book, ten led to new questions being asked on ANES surveys in 2008. In the other eight cases, we learned a great deal from the reports, but for a variety of reasons (including the performance of the items in question, the amount of time required to implement the questions, and the fierce competition for limited questionnaire time), we were not able to include the pilot study questions in the panel or time series studies. Next, we review each of the reports and the decisions made based on each one.

Some of the Pilot Study Reports had substantial impact on the 2008 and 2009 questionnaires. For example, Althaus and Tewksbury explored a new way of measuring exposure to the news media. They proposed reformatting self-reported media exposure questions to address the growing number of ways that people receive political news. We followed much of their advice. The media exposure questions on the 2008–2009 Panel Study follow the Althaus–Tewksbury format. In the 2008 Time Series Study, a randomly selected half of the sample was asked the new questions, and the other half was asked the Time Series’ traditional media exposure questions. This “splicing” was necessary because the traditional media exposure questions were part of the ANES core battery of questions (a set of questions that scholars expect to appear regularly on ANES Time Series questionnaires). So the “core” questions were included to allow researchers to compare 2008 to the past, and the new questions were included to allow researchers to more accurately study trends over time after 2008.

Ansolabehere, Meredith, and Snowberg proposed questions to measure perceptions of economic phenomena. One result of this proposal is that the 2008 Panel Study included a question asking respondents to estimate the price of gas

in their state. The same question was asked in the 2008 ANES Time Series survey, in addition to another question that asked respondents to estimate the unemployment rate in their state.

Shani proposed a new way to measure respondents' interest in politics, how closely they pay attention to government and politics, and how often they pay attention to such matters. Although the Pilot Study Report's findings indicated that the new items did not outperform the traditional questions, the direct (and not double-barreled) wording of the new questions gave them an advantage in terms of the likelihood that respondents would understand, and respond to, the questions in a comparable manner. Therefore, a random half of the 2008 ANES Time Series sample received the traditional questions, and the other half was asked the new versions of the questions.

Martinez, Gainous, and Craig proposed including measures of ambivalence. A great deal of research in psychology and political science suggests that ambivalence may be worth measuring in ANES surveys. The 2006 Pilot Study afforded an opportunity to evaluate such measures regarding the federal government. While it is likely that people possess multiple and even conflicting attitudes about government, the Pilot Study Report indicated that although the newly proposed measures were less susceptible to question-order and response-order effects, they did not fare well in assessments of construct validity. Hence, these questions were not made a high priority in our 2008 studies.

Whitaker and Fulwider examined questions tapping perceptions of the extent of similarity among members of the major political parties. Interesting differences emerged. For example, self-identified "strong" Republicans perceived Republicans as being, on the whole, more similar to one another than did other Republicans. In other cases, the report produced no consistent finding that these items provided added predictive power with respect to turnout or vote choice. These findings were the basis for including an improved version of a perceived-partisan-similarity question in the time series questionnaire.

Judd, Van Boven, Huber, and Nunes proposed an innovative way to measure perceptions of attitude polarization in the public, a topic of great interest among political scientists in recent years. These measures performed well in a pilot study and were therefore included in a postelection wave of the panel study. However, because these questions took a substantial amount of time to implement, they were not included in the time series study.¹

Gershenson and Plane, and Uslander, wrote reports comparing various ques-

tions assessing who participants trust. We also commissioned additional analyses to compare different versions of the trust questions to one another. The aggregate results of these analyses suggested potential value in one of the new measurement approaches, so we used the traditional trust-measuring approach with a random half of the 2008 ANES Time Series respondents, and we used the new questioning approach with the other respondents.

Berinsky and Lavine evaluated questions measuring self-monitoring—the extent to which people feel comfortable expressing themselves in front of others. Albarracín, Wang, and Albarracín evaluated questions assessing respondents' confidence in their ability to defend their opinions in front of others. Both of these reports showed that these questions were helpful in tapping new dimensions of personality that have the potential to moderate important relationships between key attitudes and political behaviors. Both measures were therefore included in the 2008 ANES Time Series Study.

Schwartz, and Hitlin and Kramer, evaluated different ways of measuring values. Both of these means of understanding values build on earlier work by Schwartz. Both reports advance our understanding of the potentially interesting effects of measuring multiple value dimensions. However, the 2006 Pilot Study's values question battery was quite long. This fact, combined with the absence of evidence of stronger effects of values in explaining the ANES's key dependent variables resulted in these questions not being included in the 2008–2009 studies.

Dolan and Sanbonmatsu proposed questions to explore the impact of candidate gender on public thinking and action. These questions were especially useful because of Hillary Clinton's role in the 2008 election. Given that the Democratic nominee turned out to be African American rather than female, and that the basic logic of these questions could be used in both situations, we used the Dolan–Sanbonmatsu format to generate new measures of how a candidate's race affects citizens' perceptions of candidates. These measures were used in the 2008 ANES Time Series Study and the 2008–2009 Panel Study.

Mockabee, Wald, and Legee proposed a number of new questions about religion. These questions went beyond traditional questions about respondents' religious affiliations and into their beliefs about specific religious practices. Several of these questions were determined to provide new explanatory leverage and were included in the 2008 ANES Time Series Study.

Zigerell and Rice evaluated a new approach to measuring respondents' attitudes on abortion. The existing ANES question about abortion had been criticized for many reasons, and the 2006 Pilot Study included a set of questions that attempted to overcome these problems. Instead of asking one catchall question, the new version resembled the abortion question sequence from the

¹The visual interface prevented these measures from being included in the 2006 ANES Pilot Study, which involved telephone interviews. We therefore conducted a subsequent pilot test of these measures over the Internet. The Judd–Van Boven Pilot Study Report used these data.

General Social Survey, in that it asks a series of questions about abortion under various circumstances. The new questioning approach provided more-valid measures of respondents' attitudes. This approach provides new information about how and when various populations disagree on the politics of abortion. As a result, we asked a particularly effective part of the new abortion-question sequence in both the 2008–2009 Panel Study and the 2008 ANES Time Series Study. Half of the respondents in the time series study were asked the traditional question in order to permit longer-term comparisons of the old and new measures.

Matsueda and his coauthors focused on perceptions of how different social groups are treated by police. These investigators' report suggests that the items have strong construct validity and are related to perceived efficacy. These items' relations to turnout and vote choice, however, were small and nonsignificant. While we were convinced of the validity of the concept, the limited applicability and power of these factors in electoral contexts led us not to include these questions in the 2008–2009 surveys.

Durante and Putterman examined new questions on progressive taxation and on the taxing of inheritances. Their Pilot Study Report and our own analyses showed that responses to these questions had unique potential to help analysts explain vote choice and turnout. Consequently, these questions were included in the 2008 ANES Time Series questionnaire.

We are grateful to the authors of all these reports for their excellent work and for their contributions to the collective good of improving electoral surveys. The reports challenged us to think carefully about a wide range of issues. They also constituted a rigorous basis for making difficult choices. Readers may disagree with our judgments or with the conclusions reached by these reports. Further inquiry into a number, perhaps all, of these matters is certainly needed. We hope that by making the questionnaire-development procedure more transparent, we are making it easier for future scholars to use our collective records and reports to increase their knowledge and effectiveness.

CONCLUSION

The Online Commons was designed to enhance participation in the design of ANES questionnaires, to enhance the accountability of ANES decision makers in designing the questionnaires, and to enhance the transparency with which the decisions were made. As a result, it leaves a **comprehensive record of questionnaire-design-related content** to future scholars (and future designers of the ANES) so that they will be less likely to wonder (as we did often) why a particu-

lar measure was or was not included in the questionnaire. The approach was successful in yielding hundreds of new questions that were asked in the 2008–2009 studies, a level of innovation (inspired by a broad range of scholars from many disciplines) that has not been seen in the ANES since its very early years. We hope this approach will serve as a model for future PIs of this study and of others in the coming years.

We believe that open scholarly communication about questionnaire design can yield not only better questionnaires in the short run, but also better scholarship in the long run. The long-run benefits, however, will come only if a broad community of scholars takes an active role in research on critical aspects of questionnaire design. The more that people approach questionnaires in a strictly passive manner, the less progress we should expect in creating more effective questionnaires. Science surely benefits from the easy availability of “free” data sets and the increasing user-friendliness of statistical programs capable of running intricate analyses. But science also surely suffers from the lack of careful attention to data quality that such easy access to precollected data can engender.

We recognize that it is tempting for scholars to assume that data sets like the ANES surveys include measures with strong theoretical rationales and demonstrated empirical effectiveness. Indeed, much hard work has been done in recent decades to increase the likelihood of such circumstances. But the scientific community would be better off if more scholars were actively questioning the data sets that they use and offering constructive challenges to such instruments so that the quality of the data can steadily improve over time. With regard to this vision, the ultimate goal of the OC, the Pilot Study Reports, and this book is to serve as examples of how survey research might benefit when people offer systematic evidence and argument in evaluations of specific hypotheses about specific questions. We hope that you will find this book a constructive step in that direction.