Institute for Homeland Security Solutions
Applied research • Focused results

November 2009

Innovative Survey Methodologies for the Study of Attitudes Toward Terrorism and Counterterrorism Strategies

Introduction

Radicalization is a topic of growing interest to academics and policy makers in the United States. Scholars provide a range of definitions for the concept, from the movement toward increasingly extreme viewpoints to the gradual—or sometimes quick—acceptance of active extralegal behavior in support of political, religious, or social objectives (McCauley & Moskalenko, 2008). At the same time, they generally agree that radicalization is a process, and thus needs to be examined over time. Furthermore, it needs to be studied in the context of government policies and activities—both those that have been developed to counter radicalization and those that may inadvertently fuel it. Thus a better understanding of public attitudes toward both terrorism and counterterrorism strategies is central to any study of radicalization.

A study of public attitudes toward terrorism and counterterrorism strategies poses an array of challenges, ranging from selecting the appropriate target populations to developing instruments that provide meaningful measures of relevant attitudes, emotions, and behaviors. Surveys are a powerful and timely methodology for studying the attitudes, beliefs, and behaviors of the American public, particularly for capturing *changes* in public opinion both over time and in response to specific government policies and political events. Based on a review of the relevant survey methods literature, this report provides an overview of some of the most salient methodological and measurement considerations in developing a survey project to examine attitudes toward terrorism and counterterrorism strategies in the United States.

As with all survey projects, there will be inevitable constraints, and quality trade-offs based on those limitations to costs and time, but there are two additional challenges specific to this subject that demand attention in developing a survey on this topic. First, it is certain to be difficult accurately to measure the attitudes, emotions, and behaviors related to a willingness to engage in illegal and violent political action (or sympathy toward those possessed of such a willingness) because such questions are sensitive in nature. Concerns about confidentiality, privacy, or social desirability could make respondents less likely to answer questions or to answer them honestly. We devote considerable attention to survey design considerations relevant to this issue, highlighting innovative approaches developed by survey research professionals who have dealt effectively with measuring sensitive attitudes. Second, support for terrorism is an orientation not widely held in the general population. For this reason, researchers may decide to identify subgroups within the population that warrant more focused investigation. However, identification and selection of these more narrowly defined populations presents a unique set of challenges if the overarching goal is to draw generalizable inferences about these subgroups.

The review proceeds as follows. First, we provide a brief, introductory overview of the survey process, defining relevant terms and raising basic design considerations organized around sample, survey type, and mode of contact. Next, we proceed to a discussion of considerations about the measurement of attitudes, behaviors, and emotions relevant to the study of support for terrorism and counterterrorism strategies. That discussion provides an overview of general measurement considerations and specific measurement tools, with particular attention to techniques for measuring "sensitive" items.

Survey Design Considerations

We start with a brief overview of the survey process, which includes defining the survey objectives (specifying population, type of data, desired precision), determining who will be sampled, choosing the response mode, drafting the questionnaire, pretesting the instrument, implementing the survey (including any pre-notification, reminders, nonresponse follow-up), processing and analyzing the data, and, finally, reporting the results. It is important to recognize that each step in the survey process can introduce error into the resulting survey statistics and should be carefully designed to minimize total survey error subject to the constraints and goals of the project. Groves et al. (2004) provide a thorough discussion of the survey process and the various sources of error in the process, including respondent selection issues (coverage error, sampling error, and unit nonresponse error), response accuracy issues (item nonresponse error, measurement error due to interviewer, measurement error due to respondent), and survey administration errors.

Rather than walking through the various design possibilities for each of these steps, we focus primarily on instrument design, and on the ways in which *survey mode*—the format by

Institute for Homeland
Security Solutions
Applied research • Focused results

which the respondent is interviewed (e.g., face to face, by telephone, on the Web, or through the mail)—can interact with measurement decisions. In particular, social desirability effects can be exacerbated by the use of interviewers, suggesting that a self-administered survey mode is likely preferable for this sensitive topic. Because of the substantive interest in the impacts of policies and events, we also outline various measurement techniques for capturing attitudes and emotions that might change in response to government policies.

Sampling Considerations

Although we will not walk through all of the possible sampling design considerations (for more extensive reviews,see Cochran [1977]; Kish [1965]; Lohr [1998]), it is worth highlighting that this project as currently conceived could involve two target populations: the general American public (including noncitizens) and one or more smaller groups of theoretical interest within the population (perhaps right-wing extremists, radical environmentalists, or anti-Western Islamist radicals). To the extent that the project seeks to make inferences from the survey results to any given broader target population, it will be necessary to employ a probability-based sample. A convenience sample, in which the probability that an individual is selected into the sample cannot be computed, will not allow for generalizations, no matter its size (Groves et al., 2004). There might be value in a convenience sample if there is interest in an in-depth study of just the sampled individuals—for example, to trace their changes in opinion over time or implement a survey experiment—but it should again be emphasized that results from any analysis would not be generalizable beyond the particular individuals in the sample.

Unfortunately, for special populations there often is no reliable *sampling frame*—the set of units from which the sample will be drawn. Yet there are some techniques available for probability-based sampling of rare populations. One is to use disproportionate sampling, in which the population is divided into geographic strata and high-density strata are oversampled (Kalton & Anderson, 1986). Unfortunately, this approach does not help if the subpopulation is geographically dispersed, as would be the case if radical environmentalists, for instance, were of particular interest. An alternative approach that can provide a valid probability sample even with a dispersed population is to use a screening in which a large sample is selected and then the questionnaire is administered only if the selected respondent is a member of the subpopulation; this approach can be a one- or two-stage design (Clark, 2009). Though effective, this strategy can be expensive. For example, it would take a very large general population sample to reach even a small number of Muslim Americans. In 2007, the Pew Research Center for the People and the Press required initial interviews with 55,000 randomly sampled respondents before it could administer a survey to a representative subsample of 1,050 American Muslims (Pew Research Center, 2007). A lower-cost alternative is to use even partial or incomplete lists of the population if they are available (Kalsbeek, 2003). Of course, the accuracy of results generated using such samples is less than ideal and varies—generally unknowably—depending on the completeness of the lists used. For example, Elliot et al. (2009) report that surname lists for Hispanic populations in the US have about 20%

undercoverage and 10% overcoverage (with similar rates for Chinese and Japanese populations).

Cross-Sectional or Longitudinal Panel Survey Design

Another design consideration is whether a survey will be repeated with the same set of sample members (longitudinal panel) or whether it will be conducted only once with a set of sample members (cross-sectional). The nature of the research project goals will make some survey designs more preferable than others, because different designs carry distinct costs and benefits. A *cross-sectional* design allows for inference about the relationships between variables of interest (e.g., demographic characteristics, attitudes, behavior) but it does not allow researchers to investigate dynamics. A *repeated cross-sectional survey* measures the same items over time and is useful for identifying aggregate trends, but cannot identify individual-level change or, by extension, its causes. In a *panel survey*, the sample of individuals is interviewed multiple times. Panel surveys are useful for identifying and measuring individual-level change but can be more costly to conduct than cross-sectional surveys. Panel surveys also introduce challenges such as panel attrition and conditioning, although there are techniques for testing for such effects (e.g., independent cross-sections; accuracy benchmarks; see, for example, Kasprzyk, Kalton, Duncan, & Singh, 1989). If there is particular interest in gauging *changes* in individual attitudes—for instance, in response to a government counterterrorism policy—then a panel design is most appropriate.

Mode Considerations

Another fundamental design issue that must be considered together with decisions about sample frame and survey type is the mode of contact. Surveys can be administered through an array of different modes—telephone, mail, Internet, face-to-face, combinations of these, and so on. Each mode has advantages and disadvantages given the particular research question being investigated and the budgetary and time constraints of the project. As we will discuss in more detail below, a study of attitudes toward terrorism and counterterrorism strategies draws attention to the key distinction between interviewer-administered surveys and self-administered surveys. In short, the choice of mode is always important, but this consideration is particularly important in a study that may be attempting to measure sensitive attitudes and behaviors. Likewise, for some specialized populations, there may also be particular advantages or disadvantages associated with each of the modes. Survey designers must take into account effects introduced by the race or gender of the interviewer, errors that arise due to language and cultural issues, and the possibility that respondents may vary in their levels of literacy.

Face-to-face interviewing is often considered the gold standard for survey research. This mode generally has higher response rates and allows researchers to probe respondents' views in more depth, because interviewers can ask follow-up questions and press for more detail if responses are deemed incomplete (Biemer & Lyberg, 2003). It is typically considered the best mode for eliciting open-ended responses, and interviewers can also code qualitative

Institute for Homeland
Security Solutions
Applied research • Focused results

observations about respondents during the course of the interview. On the other hand, this mode is considerably more expensive (due to interviewer training and travel expenses, among others), especially for national samples. It can also be more susceptible to social desirability biases and other interviewer effects. For instance, in some cultures, male respondents might be hesitant to reveal their beliefs and behaviors to a female interviewer (Davis, Couper, Janz, Caldwell, & Resnicow, in press).

Telephone surveys have both the advantages and disadvantages of a live interviewer but are typically much lower in cost. Most contemporary telephone surveys involve computer-assisted telephone interviewing (CATI), a general term for an interview mode in which the administration of a telephone-based survey is augmented by information contained in computer databases. When a live interviewer is present, it appears that CATI is susceptible to the same interviewer effects observed in face-to-face surveys, although CATI is probably superior to face-to-face modes in its capacity to elicit and measure sensitive attitudes (Harmon et al., 2009; Kreuter, Presser, & Tourangeau, 2008).

For both face-to-face and telephone interviews, computer assisted self-administered interviewing (CASAI) allows for self-administration of questionnaires. For face-to-face modes, subjects answer questions presented in text or graphical form on a computer screen (typically the interviewer brings a laptop that can be used), or survey items are administered by a voice-digitized automated interviewer. In a telephone interview mode, an interviewer is used to screen and recruit the subject, but then the subject answers some or all of the questions using prerecorded, audible questions by pressing appropriate numbers on a telephone keypad (Mingay, 2000). In a recent study, Harmon et al. (2009) showed that telephone audio computer-assisted self-interviewing (T-ACASI) may have an additional ability to uncover the semicensored attitudes that are evident among interviewees in conventional telephone surveys. The authors found that T-ACASI subjects were more supportive of corporal punishment, traditional gender roles, and recreational marijuana use, and less supportive of racially integrated communities. Thus, variants of social desirability bias may extend into peripheral issue attitudes in live-interviewer survey modes. ACASI surveys carry the costs and benefits from the absence of an in-person interviewer. This mode may also allow for improved response rates from low-literacy research subjects, who could be more prevalent in some of the subgroups considered. Couper, Tourangeau, and Marvin (2009) have shown, however, that audio components in the in-person self-administration process may not be necessary. In a study of the general population, most respondents did not use the audio components available to them, rendering ACASI results largely analogous to text-based computer-assisted self-interview (CASI) in this population.

Another variation available for telephone interviews is interactive voice recognition (IVR) mode. This typically involves full interaction between a computer database and a telephone respondent, who enters response information by pressing number keys or providing brief voice responses. IVR eliminates the presence of a live interviewer, which comes with a cost to

response rate because refusals may be higher. This mode is not without advantages, however; it has been shown to reduce the prevalence of social desirability bias relative to other modes such as CATI (although IVR appears more vulnerable to this bias than Web-based administration; see Kreuter et al., 2008). Perhaps more importantly, the duration of an IVR survey must typically be kept quite short.

However responses are elicited, random-digit-dial (RDD) telephone surveys have had declining response rates since the early to mid-1990s, a behavioral shift that researchers attribute to the increasing prevalence of unsolicited telephone calls and the spread of technology that allows potential respondents to screen their calls. There are, however, a range of techniques available to increase response rates. De Leeuw, Callegaro, Hox, Korendijk, and Lensvelt-Mulders (2007) found that prenotification letters mailed to respondents increased both response rates and cooperation rates. Interviewer training (Groves and McGonagle 2001), voice characteristics (Groves et al. 2006), and persuasion techniques (De Leeuw & Hox, 2004; Dijkstra & Smith, 2002) also influence response rates, as does the use of incentives (see Cantor, O'Hare, & O'Connor, 2006; Singer, 2002). The RDD telephone mode has also encountered increasing coverage error as a result of the increasing number of cell phone–only households (Keeter, Kennedy, DiMock, Best, & Craighill, 2006). This may be of particular concern for studying attitudes toward terrorism and counterterrorism strategies, because cell phone–only households differ systematically from households with landlines. Tucker, Brick, and Meekins (2007) found that cell phone–only households are younger and more likely to be minorities.

Relative to the face-to-face modes of in-person and telephone interviews, Web-based survey administration is quite convenient for both interviewers and subjects. Self-administration can be conducted at home over a personal computer or researcher-provided unit, and experimental procedures can be embedded in the administration process. Kreuter et al. (2008) and Dillman (2006) have shown that the Web-based mode is quite resistant to social desirability biases. On the other hand, it can be difficult for researchers to verify that the person taking the survey is the desired respondent; the administration process can be anonymous. Also, Heerwegh and Loosveldt (2008) have shown that Web respondents are more likely to satisfice when entering responses to survey items. In this mode, the authors found respondents provided "don't know" responses at a higher rate, were less likely to provide differentiated responses across items (see also Fricker, Galesic, Tourangeau, & Yan, 2005), and were more likely to avoid responding to individual items altogether. Nonresponse may be higher for Web surveys than for telephone-based modes. In an experiment, Fricker et al. (2005) found that only about half of respondents who were contacted by telephone and assigned to a Web survey completed the main questionnaire, whereas nearly all of those assigned to the CATI mode completed the entire survey. Schlenger and Silver (2006) observe that some respondents may worry that Internet surveys will lead to their subsequent Internet activity being tracked. These downsides notwithstanding, there are, however, also a host of

Institute for Homeland Security Solutions
Applied research • Focused results

benefits that arise from the Internet mode for surveys, including relatively quick response times and automatic data compilation (Fowler, 2009).

The major pitfall of Web-based surveys is that identifying an appropriate sample frame can be difficult for two reasons: (1) Analysts who want to use data drawn from Web-based surveys to make inferences about the general population must confront the problem that computer non-owners differ in systematic ways from computer owners (Couper & Miller, 2008). (2) Unlike telephone surveys, the nature of the Internet means that "that frames of Internet users in a form suitable for sampling do not—and likely will not—exist" (Couper & Miller, 2008). Solutions to this problem come in two varieties. First, a pre-recruited probability-based sample can be generated by using techniques familiar to other modes (such as telephone-based RDD techniques or address-based sampling). Respondents who do not own computers can be provided with some means of Internet access to participate in the survey; this is the approach used by the survey research firm Knowledge Networks. Second, an "opt-in" or volunteer group of Internet users can be used, and then researchers can attempt to counter representational biases by employing various weighting methods (Couper & Miller, 2008). Because this second option is a non–probability-based approach, generalizations to a target population are suspect. It is not possible to calculate a response rate for a nonprobability sample, so Callegaro and DiSogra (2008) recommend that completion rates, absorption rates, and break-off rates should be reported in analyses of the second variety alongside other routine statistical measures.

Another survey mode, mail surveys, makes it easier to reach randomly sampled individuals, provided that researchers have obtained the respondent's correct address. This mode has demonstrated response rates as high or higher than other modes in some instances (Fowler, Gallagher, & Nederend, 1998); however, achieving these high response rates can require labor-intensive follow-up contacts by telephone or repeated mailings (Dillman, 2007). It can also be quite labor-intensive to enter the responses into a manageable database. Also, mail-based modes have limitations in terms of the complexity of the questionnaire that is used. For example, it is difficult or impossible to include branching questions or skip patterns, which could be particularly pertinent to a study focused on subpopulations. Finally, mail surveys typically take much longer to administer than do either Web- or telephone-based surveys, a significant downside if quick results are desired.

The modes described above do not comprise an exclusive list of the options that are available, but they are the modes most frequently employed in contemporary survey research. However, almost any media could be used to make contact with potential respondents, including personal digital assistants (PDAs), cell phones, and other emerging communication technologies (e.g., Facebook, Second Life). It is also possible to contact research subjects using more than one mode. Mixed-mode techniques can allow researchers to make contact with individuals who are difficult to reach by one mode but are more reachable by another (Fowler, 2009). For example, some individuals may not be willing to respond to telephone contact but would be willing to return a written survey mailed to their home address or to fill out

Institute for Homeland
Security Solutions
Applied research • Focused results

an online questionnaire. Use of a mixed mode can therefore improve the overall response rate, although researchers will need to investigate whether responses to survey items of interest systematically differ across the multiple modes employed in the study. De Leeuw (2005) notes that most concurrent mixed-mode approaches have negative consequences for data quality, with one obvious exception: in live interview sessions, a more private mode can be employed to administer a subset of questions to ensure greater self-disclosure and less social desirability bias.

Survey Instrument/Measurement Issues

Determining survey samples, type, and mode are fundamental considerations involving important tradeoffs. Yet beyond deciding who will be selected to respond to the survey, how regularly they will be contacted, and via what mechanisms, it is essential to have a clear plan for measuring those respondents' attitudes and behaviors. Researchers have devoted considerable effort to identifying the attributes of attitudes and to determining how they can best be measured. In this section we will discuss some of the key findings of that research.

General Measurement Considerations

In general, investigators must pay careful attention to question wording, syntax, and semantics in the construction of a survey instrument. If respondents are confronted with technical terms, vague question wordings, or unclear relative comparisons, they may make errors when responding or may fail altogether to answer survey items (Fowler, 2009). Cognitive pretesting is a common strategy employed to identify problems emerging from faulty question construction. Typically, preliminary groups of respondents drawn from the population of interest will respond to newly constructed survey items and then report their experiences to the researcher in a formal feedback session (Lessler & Forsyth, 1996). When question abnormalities cannot be easily articulated by survey respondents, observers in face-to-face modes can code such things as puzzled facial expressions or pauses before response (Fowler & Cannell, 1996; Schober & Conrad, 1997). Graesser, Cai, Louwerse, and Daniel (2006) have developed a computer program (QUAID) that identifies errors in question wording and may be of use to researchers looking for an objective tool when evaluating new survey items. In general, questions should be clear, simple, and precise (Fowler, 2009).

Translation and Cultural Sensitivity

Survey items can also have different meanings for respondents who are socialized in different cultures or speak different languages, an issue of particular concern if the survey on attitudes toward terrorism and counterterrorism strategies were to focus on some special populations. Some concepts may mean slightly different things to different cultures, or, in a more extreme case, some concepts that are important to one culture may not even exist in another. Self-interest, for example, is often assumed by English speakers to reasonably

Institute for Homeland
Security Solutions
Applied research • Focused results

govern behavior, yet other cultures can have different normative orientations toward the concept (Marin & Marin, 1991). Cultural problems are complicated when researchers must translate survey items for administration to respondents who speak different languages.

Translation is a complex process that involves language professionals' having to make a considerable number of subjective decisions. In addition to overcoming cultural differences, researchers need to keep in mind that even though some words and terms have analogues in other languages, the comparison may be somewhat imprecise. Marin and Marin (1991) point out that the word "respect" is associated with admiration and subordination to the typical English speaker, whereas the same word tends to invoke obedience, duty, and deference to a Spanish speaker. There are a variety of translation options available for researchers, offering varying levels of quality and affordability. From cheapest and least effective to expensive and most effective are simple direct translation, modified direct translation, back translation, and committee translation. (See De la Puente, Pan, & Rose, 2003, for a quick primer; Marin & Marin, 1991, go into this process in more exhaustive detail.) In a study of attitudes toward terrorism and counterterrorism strategies, non–English-speaking populations, such as Arab immigrants, may be a population of interest. Translation concerns could be paramount in a survey that includes these populations, because researchers need to ensure that both the language and concepts under investigation are precisely communicated and equivalently understood by researchers and respondents.

"Don't Know" Responses

Respondent confusion with questions can be difficult to detect, particularly if survey items allow subjects that hold clear directional attitudes to report a nonattitude. "Don't know" (DK) responses are often interpreted as nonattitudes, but they can occur for a variety of reasons, and researchers should be aware of this ambiguity if DK will be a response option for any particular survey item. Respondents might choose the DK option if they do not understand the question, if they reject the premises of the question or the response options available to them (Fowler, 2009), or if they wish to avoid socially undesirable responses (Berinsky, 2004). Fowler (2009) notes that respondents might also enter DK if they are bored with the survey process and eager to finish or could choose it by accident when they are capable of reporting a concrete preference.

An estimator has been developed in the contingent valuation literature to determine which of these accounts provides the most likely explanation for a DK response (Wang, 1997), and such techniques may be of use for researchers who wish to rigorously investigate why respondents are choosing the DK option. Experimental research has shown that the likelihood of a DK response varies based on question form (Schuman & Presser, 1981) and the degree to which interviewers will accept the DK response without further probing (Cannell, Oksenberg, & Converse, 1981; Sanchez & Morchio, 1992). Fowler (2009) suggests including initial screening questions before administering complex survey items to respondents who may be unfamiliar with the topic of the question.

Institute for Homeland
Security Solutions
Applied research • Focused results

Vignettes

When subjective attitudes are under investigation, different subjects may use the same words to report qualitatively different responses. For example, a study on attitudes toward terrorism and counterterrorism strategies could include questions inquiring whether certain policy changes would make respondents "very angry," and different subjects may have different conceptions about what "very angry" means. One innovation in question design that can resolve problems of this sort is the use of vignettes. King, Murray, Salomon, and Tandon (2004) observe that many survey items suffer from interpersonal incomparability, which has typically been referred to as differential item functioning (DIF). To cite another example, the word "middle-aged" might invoke an image of a 35-year old-person to one respondent and a 60-year-old person to another. King et al. (2004) and King and Wand (2007) have proposed the use of anchoring vignettes to mitigate these kinds of response concerns. Here, a series of simple vignettes are presented to survey respondents, who then determine how the brief passage matches up with their use of a particular concept or a relative term. King et al. (2004) study self-reported efficacy, and to that end they first match respondents' views of a fictional person's political predicament with subjective descriptions of that person's efficacy (the person has no say; some say; a lot of say, etc.). The researchers then adjust the respondent's self-reported efficacy score based on the meaning the respondent attached to the description when considering the vignettes. King and Wand (2006) update the procedure, offering both nonparametric and a variety of parametric estimators to researchers. They find considerable substantive differences in the adjusted efficacy scores: for example, Chinese respondents report higher levels of efficacy than Mexican respondents when researchers employ simple self-reports; when using King et al.'s (2004) anchoring vignettes, Mexican respondents are shown to have considerably higher efficacy ratings than Chinese respondents do.

Vignettes also can be creatively employed to solicit information that respondents may not reveal unless they are primed by particular images or considerations. In some circumstances, an attitudinal response to a particular scenario is more easily conveyed by a vignette than a direct survey item. For example, in a post-9/11 study of Dutch children, Muris, Mayer, vanEijk, and van Dongen (2008) found that native Dutch children were slightly more afraid of terrorism than Islamic Dutch children. In the same study, the children were exposed to hypothetical situations constructed by researchers seeking to determine whether children associate Middle Eastern men with the threat of terrorism. The authors found that ambiguous situations did not invoke fears of terrorism among native Dutch children, a finding arising from the use of vignettes that may have been more difficult to obtain through the use of attitude self-reports.

What Should Be Measured? Attitudes, Behaviors, and Emotion

Explicit Attitudes

There are numerous attitudinal and behavioral variables that have the potential to be influenced by counterterrorism policy. The other literature review prepared for the workshop

Institute for Homeland
Security Solutions
Applied research • Focused results

provides an outline of specific attitudes, emotions, and behaviors that have been measured in previous surveys on the topic; we identify here some of the measurement considerations that should be kept in mind in constructing a survey questionnaire. Perhaps most obviously, researchers may want to solicit *explicit* attitudes, or those attitudes consciously held by interview subjects. These attitudes should be fairly easy to obtain, provided that respondents are provided with clear survey items that they can effortlessly comprehend. In addition to identifying the *direction* of an attitude—for or against a particular concept, policy, or action— researchers must consider other dimensions that have been found to be relevant to determining if an attitude will be predictive of behavior (Krosnick, 1990). If there is interest in identifying a propensity to engage in radical political acts, for example, it may be important not only to measure the direction of an attitude but also to capture other dimensions of the attitude, including extremity, intensity, importance, stability, and ambivalence.

Extremity and Intensity

First, *extremity* is a concept best understood by its connection to direction. Responses that fall on the outer ends of a relative spectrum—as in reports of 1 or 7 in a traditional 7-point scale—are considered relatively extreme. It is often, but not always, the case that a respondent whose directional self-report is extreme holds an opinion with a large amount of *intensity*, that is,with a great deal of spiritedness, passion, or commitment. Intensity may be one of the most important ideas to the study of support for terrorism, because individuals who are inspired to engage in illegal activity may be animated by more than reasoned, directional attitudes. From a perspective of measurement, intensity may be difficult to capture in isolation. Surveys often solicit self-reported attitude intensity by including options such as "Strongly Agree" or "Somewhat Disagree" in response to a controversial or contestable statement included in an item's question stem. However, these types of responses are often conflated with attitude importance, which we consider next.

Importance

Attitude *importance* is the degree to which a respondent considers an attitude salient. It is possible for a subject to hold extreme and intense preferences on an attitude, yet consider the attitude unimportant in comparison to other competing attitudes. Krosnick (1990) argues that attitude importance may be the most critical dimension of an attitude. People who consider their own directional attitude important tend to divide the world into two groups: those who agree with them, and those who disagree with them. Importance is also relevant to persuadability, because those who consider an attitude to be important are less likely to reverse the direction of their opinion after being exposed to contrary evidence. One way to measure importance is simply to ask how an important an attitude is to the respondent. However, scholars have found that people tend to say (and, presumably, many believe that) most of their attitudes are important. Alternatively, open-ended questions that ask about important attitudes can be useful for identifying those that are truly accessible in someone's

Institute for Homeland Security Solutions
Applied research • Focused results

mind. However, open-ended questions have higher levels of nonresponse and can create other problems, including challenges in comparing responses across different subjects.

Researchers have developed a variety of techniques that force respondents to rate the relative importance of a litany of competing issues. One technique that can be used to measure the relative importance of different attitudes comes from the broad umbrella of paired comparison surveys; it is the analytic hierarchy process (AHP), which is useful when establishing priorities in a multiobjective environment or when deciding how to allocate scarce resources (Forman & Gass, 2001). Essentially, the AHP creates decision hierarchies, establishing weights for a variety of considerations the subject considers relevant to making a decision. MaxDiff is a technique developed by Louviere (Finn & Louviere, 1992; Louviere, 1993) that is a variant of the method of paired comparisons (David, 1969). Computer software can allow respondents to make pairwise comparisons between subsets of possible competing considerations in a highly complex decision environment. Researchers then use these comparisons (and transitivities implied by them) to estimate coherent preference orderings using Bayesian hierarchical modeling (Orme, 2005).

Stability

Stability refers to the degree to which a subject continues to hold an attitude over time; it is typically measured by asking a subject the same question in repeated interviews in a panel study. This dimension of attitudes is particularly relevant to the study of attitudes toward terrorism and counterterrorism strategies because one of the objectives might be to determine how government policies lead to changes in attitudes over time. If a particular government action is correlated with systematic changes in respondents' attitudes, it is important to distinguish between attitude changes that are the result of the policy change and attitude changes that are a consequence of simple attitude instability.

Ambivalence

Finally, *ambivalence* refers to the fact that many respondents hold competing considerations relevant to producing a response to a survey item. Sometimes, a respondent may hold considerations both in favor of and against a particular proposition and therefore will experience some degree of uncertainty or indecision in reporting a response. For example, when asked about an attitude toward abortion, many respondents may oppose the procedure for religious reasons, yet support abortion rights because of opposition to state intervention into the relationship between doctors and patients. Such respondents would experience ambivalence to a question asking whether they consider themselves pro-life or pro-choice. Most scholars recognize that, although certain decisions are complex, when a decision must be made, people may place relative weights on the considerations that are accessible to them at the time of the decision.

Measuring attitude accessibility is one of the core ways that researchers can assess the degree to which respondents are ambivalent about a response. Attitude accessibility is

Institute for Homeland
Security Solutions
Applied research • Focused results

typically measurable by *response latency,* or the length of time that elapses between when the interviewer poses the question and the subject begins to articulate a response. When an attitude is highly accessible, respondents tend to respond to questions quickly, whereas inaccessible attitudes lead to slower response times (Mulligan, Grant, Mockabee, & Monson, 2003). In recent research, scholars have measured response latency by using active timers, which are essentially computerized stopwatches activated by interviewers during the interview process. Because this method can require high levels of training and frequently leads to input errors (Bassili & Fletcher, 1991; Huckfeldt, Levine, Morgan, & Sprague, 1999), scholars have also developed automatic timers to measure response latency. These are often voice-activated timers that start when an interviewer finishes asking a question and stop when the respondent begins to answer the question. Although this technique overcomes some of the errors associated with interviewer-activated timing, it has issues of its own: frequently, automatic timers pick up background noise or respondent-generated sounds, such as throat-clearing or sighing, that do not constitute a genuine response (Mulligan et al., 2003). A third alternative is the use of "latent timers," which simply measure the time that elapses from the moment the question appears on the interviewer's computer monitor through the entirety of the respondent's answer. Mulligan et al. (2003) derive a method of accounting for interviewer and subject effects and show that latent timing methods are quite strongly correlated with active timing techniques.

Regardless of the method used to measure response latency, researchers should be aware that simply counting the time that elapses before a subject answers the question does not necessarily provide a clean measure of accessibility. Subjects may delay response due to confusion over question wording or because they are weighing some of the social desirability conditions discussed above. Mulligan et al. (2003) suggest these confounding factors can be minimized by constructing carefully worded questions and anticipating response situations that could be biased by social desirability.

Implicit Attitudes

Although subjects are often willing to report explicit attitudes openly and honestly, in some cases they are not. Subjects may fail to report such attitudes accurately because the attitudes are seen as socially undesirable or because the subject is worried that they contain information that could potentially be reported to the authorities. For example, individuals might be reluctant to report that they believe that the drinking age should be lowered, that marijuana should be decriminalized, or that long-term U.S. residents present in the country illegally should be granted clemency and allowed legal residency status, particularly if those views are linked to engaging in behaviors (or having friends or family engaged in behavior) that *are* illegal but that the respondents hold should not be. Finally, some researchers study *implicit* attitudes, or those associations held by subjects who are either unaware of them or unwilling to acknowledge them.

Institute for Homeland
Security Solutions
Applied research • Focused results

A variety of techniques have been developed by social psychologists to explore implicit attitudes and explicit attitudes that respondents are unwilling to reveal. For instance, scholars have monitored brain activity after exposure to stimuli (Phelps et al., 2000; Ronquillo et al., 2007). The implicit association test (IAT; Sriram & Greenwald, 2009) uses response latency— the length of time it takes respondents to make a judgment (discussed more below)—to measure associations between an attitude object and an automatic evaluative judgment about that object. However, the procedure has been criticized for failing to distinguish between the presence of an affective judgment and simple environmental exposure to the judgment of interest (Arkes & Tetlock, 2004; Karpinski & Hilton, 2001; Uhlmann, Brescoll, & Paluck, 2006). The affect misattribution procedure (AMP; Payne, Cheng, Govorun, & Stewart, 2005) has yet to face such criticism. The AMP involves exposing subjects to an image of theoretical interest (e.g., a photo of a spider, which typically elicits negative affect), and then quick exposure to a neutral image (e.g., Chinese calligraphy). Payne et al. (2005) found that affect toward the theoretical object is transferred to the neutral object, and respondents freely report their affect toward the neutral object, unaware of its association with the theoretical object. This technique has been used to measure implicit racism, but could likely be adapted to other sensitive items. The approach is also more efficient than many of the alternative approaches, because it can be administered in as little as 5 minutes in an Internet-based study. On the other hand, basing inferences about attitudes toward terrorism and counterterrorism strategies on responses that are not explicitly provided may raise unwanted political considerations in this project.

Reports of Past or Current Behavior

Much of our discussion has been focused on survey design considerations and problems pertaining to soliciting attitudes of various kinds, but we should note that many of the same issues and concerns apply to survey questions about behaviors. For example, the classic example of social desirability bias is voter overreporting—many respondents tell interviewers that they voted when they did not (Sigelman, 1982; Silver, Anderson, & Abramson, 1986; Traugott & Kattosh, 1979). More relevant to the study of attitudes toward terrorism and counterterrorism strategies, we might expect subjects to be reluctant to admit that they have participated in violent protests or engaged in illegal activity in the past (or are planning to engage in such activity in the future). In such circumstances, the considerations of survey design and mode are equally relevant in a study that asks respondents about behavior. Self-administered modes will reduce social desirability bias and lead to a greater willingness on the part of respondents to volunteer information that they might not reveal in front of a live interviewer.

Questions about illegal behaviors may be even more susceptible to respondent deception than socially undesirably attitudes because of concerns of confidentiality (Singer, Mathiowetz, & Couper, 1993). It is important both to give strong assurances of confidentiality and to have data collected and managed in a way that provides strong protections against breaches in

Institute for Homeland
Security Solutions
Applied research • Focused results

confidentiality (e.g., a disclosure analysis should be conducted to make sure all respondent identifiers are removed from the response file).

Emotion

Emotion has been a fruitful area of study for political scientists and psychologists in recent years, and analysis of a survey of attitudes toward terrorism and counterterrorism strategies may yield relationships between extremism and any of a number of distinct emotions. Affective intelligence, a theory of emotion from political psychology, identifies enthusiasm, fear, and anxiety as types of emotion that have consequences for political behavior (Marcus, Neuman, & MacKuen, 2000). Enthusiasm is expected to lead individuals to rely on pre-existing information and established habits when confronting a decision. Anxiety is hypothesized to lead to more conventionally rational decision strategies: individuals confronting threats are expected to engage in more exhaustive informational searches and to base decisions on utility-maximizing criteria. Empirical and experimental research has largely borne out these hypotheses for everyday political phenomena (Brader 2005; Brader, Valentino, & Suhay, 2008; Marcus et al., 2000). A recent study of responses to terrorism found distinct implications for anxiety and perceptions of threat (Huddy, Feldman, Taber, & Lahav, 2005). After the September 11th terrorist attacks, Americans who experienced anxiety were more risk averse and therefore more likely to oppose military action overseas and to distrust President George W. Bush. Those who perceived "threat" to the United States in the future were more supportive of military engagement and of President Bush in general.

A key methodological challenge in a rigorous study of emotion is identifying whether it has a causal effect or is a simple correlate of another variable of interest. Marcus et al. (2000) has been criticized by Joslyn (2001) and Brader (2005) because their findings do not necessarily show that self-reported emotional responses are causal determinants of a variety of political behaviors. It could easily be the case that cognitive decision making could be the cause of emotional response, rather than the reverse relationship hypothesized by many of the theorists discussed above. More convincing evidence for the causal impact of emotion comes from experimental research such as Brader's (2005) study of the mediating influence of emotion on the persuasiveness of political advertising. In the study of attitudes toward terrorism and counterterrorism strategies, researchers should consider whether emotion as conventionally measured (in survey self-reports) is sufficient to uncover the nature of the relationship between emotions and particular attitudes and behaviors of interest.

Specific Measurement Tools

Measurement Scales

One of the inherent difficulties in the study of any complex construct in public opinion is that it can be difficult or impossible to operationalize the theoretical construct into a single survey question. For example, researchers have found that they get a more valid and reliable

Institute for Homeland
Security Solutions
Applied research • Focused results

measure of ideology if they create a scale from a variety of issue questions than if they ask respondents to place themselves on a liberal-conservative scale (Treier & Hillygus, in press). Attitudes toward terrorism and counterterrorism strategies may be constructs that present analogous challenges. It might be difficult to directly measure a respondent's propensity to support terrorism, but there are a number of attitudes that could be related to the construct: perception of government intrusion and opinions about government conspiracies, among many others (see the other literature review for a more comprehensive survey of specific concepts).

Moskalenko and McCauley (2009) generated a battery of questions from which they estimate respondents' relative positions along the latent constructs of political activism and radicalism. The former construct refers to support for lawful political activism, whereas the latter construct measures the extent to which respondents support extralegal political activism. Unfortunately, the particular questions used might not be ideal. One assumption underlying this scale is that respondents were willing to provide truthful responses even though many of the items asked were sensitive. Respondents may have been uncomfortable answering the questions at all, or they may have misreported because of a desire to avoid the appearance of social impropriety. There are techniques for identifying if this was the case (Berinsky, 2005).

Hypothetical situations have also been used in surveys to assess complex constructs. For example, as discussed above, Muris et al. (2008) used hypothetical vignettes in their study of terrorism fears among Dutch children. At the same time, although hypotheticals may provide useful inferences in some circumstances, researchers should be aware of the benefits and limitations of this approach. One such limitation is that people are not skilled at predicting their responses to hypothetical situations they have not yet encountered (Fowler, 1995). To the extent that hypotheticals are to be included in a battery of questions, Fowler suggests that they should focus on scenarios respondents have experienced before. In addition, Fowler warns that respondents should not be asked to explain the causes behind decisions they have made or actions they have taken, nor should they be asked to report knowledge based on second-hand experience.

It may be that it is possible to avoid hypotheticals and to develop meaningful scales by using issues related to an attitude of interest. For example, although respondents may be reluctant to express anger toward Homeland Security policies they regard as invasive of their civil liberties, they may be willing to express anger toward more innocuous policy measures, such as swine flu vaccinations. Where attitudes share a common attitudinal component (such as distrust of government or opposition to regulation), attitudes toward one class of policy actions might serve as useful indicators toward others. Certainly, as with development of any new attitudinal measure, pretesting to ensure meaningful scales is essential.

Experimental Approaches

Social science researchers have developed experimental techniques designed to manage respondents' propensities for social desirability response bias. These methods allow

Institute for Homeland Security Solutions
Applied research • Focused results

for strong inferences about causation; by randomizing assignment into control and treatment groups, researchers can control for any number of intervening variables that otherwise confound analysis in survey research. Typically, researchers will insert an experimental element into the treatment group and observe the downstream consequences of this effect by comparing responses to identical questions administered to both control and treatment groups. Examples of this research design include Sniderman and Piazza's (1993) "mere mention" experiment, in which inclusion of the phrase "affirmative action" led to increased reports of dislike for blacks among white respondents in the treatment group.

Survey experiments have been shown to be particularly effective at measuring *framing* or *priming* effects, which could be useful for measuring how the public might differentially respond to various ways of framing a government action. Framing refers to the process by which an event or activity is presented so that it resonates with particular cognitive schema that respondents may depend upon to understand questions or events; priming refers to the process by which certain considerations relevant to a judgment are rendered more salient while others are made less salient. An experiment on the 2000 Census found that positive and negative framing devices about the census provided to treatment groups before administration had consequences for subsequent responses on the survey (Hillygus, Nie, Prewitt, & Pals, 2006). The experiment found that negatively primed subjects skipped a larger proportion of items than the control group across a wide range of question types, but particularly on those items covering sensitive topics. In studying attitudes toward terrorism and counterterrorism strategies, a policy such as requiring a permit for a protest could be introduced in terms emphasizing the value of respecting civil liberties or the importance of ensuring public safety, and support for the required permit could be assessed across the two different frames.

Although many scholars are enthusiastic about the improved causal inferences offered by the survey experiment, Gaines, Kuklinski, and Quirk (2007) caution that certain experimental designs can potentially contaminate inferences. When there are multiple experiments conducted within a single survey, the first experimental treatment that respondents encounter can have spillover effects during administration of the second experiment. More generally, Gaines et al. argue that all survey experiments should include control groups to protect against spillover effects. They also point out that the discovery of treatment effects may be valuable at the theoretical level, but significant experimental effects do not necessarily lead to substantive consequences for political behavior. In the study of attitudes toward terrorism and counterterrorism strategies, survey experiments may uncover behavior of interest to researchers, but scholars should be aware that the duration of an effect of interest is as potentially consequential as positive evidence of the effect to begin with. This cautionary note echoes the longstanding concerns about the external validity of experiments in social science research generally (McDermott, 2002; Morton & Williams, 2008). With these warnings in mind, scholars may want to include experiments to determine whether different ways of framing government action, or different levels of government action, lead to different responses.

List Experiments

A list experiment is a particular question type that can be used to measure sensitive attitudes and behaviors that might help to reassure the respondent about confidentiality. This tool has usually been employed to manage response effects due to racial hostility or gender bias (Gilens, Sniderman, & Kuklinski, 1998; Kuklinski, Cobb, & Gilens, 1997; Kuklinski, Sniderman, et al., 1997; Streb, Burrell, Frederick, & Genovese, 2008). In a typical (simple) version of this research design, subjects are randomly assigned to control and treatment groups. The control group is given a series of three or four statements (for example) and then asked how many of those statements make them angry (or elicit whatever response the researcher might be interested in). The treatment group is then queried on the same series of statements the control group responded to, although here an additional statement of interest is asked. In many cases, the statement of interest is intended to elicit the sensitive attitude in question. For example, in a study of racialized political attitudes in the South, Kuklinski, Sniderman, et al. (1997) asked the control and the treatment groups whether they were angry about corporations polluting the environment, out-of-control salaries for professional athletes, and increases in gas taxes. Then they also asked the treatment group whether they would be angry about "a black family moving next door." The idea here is that respondents are free to mention that the latter hypothetical makes them angry (that is, to include it among the count of the total items that anger them), because the presence of the control items gives them sufficient anonymity to reveal a socially undesirable attitude on the treatment item.

The strength of a list experiment is that researchers who employ it can get a more honest sense of the extent to which a controversial or undesirable attitude exists in a population of interest. Mean comparison tests can allow precise statements about the presence of the attitude within the populations under analysis. On the other hand, this approach does not give individual-level measures. Corstange (2009) has derived an estimator, LISTIT, that allows for individual-level inferences from carefully designed list experiments. The author's estimator allows probabilistic assessments about an individual's sensitive attitudes, but he acknowledges his estimator will produce larger standard errors than would be the case if respondents were willing to directly reveal the sensitive attitude in question. Glynn (2009) has also developed an estimation strategy for individual-level attitudes in list experiments, which is conditional on responses to related indirect questions.

Another problem unique to list experiments concerns "ceiling effects," or the problem of individuals responding in the affirmative (or negative) to all control questions. In such a circumstance, the control questions no longer cloak the response to the sensitive question in anonymity. To borrow from the example cited above, if respondents believe all three control questions make them angry, reporting to the interviewer that *four* questions made them angry is tantamount to acknowledging that the presence of a black family in their neighborhood makes them angry. In such a circumstance, social desirability biases may lead to misleading aggregate-level inferences. Glynn (2009) proposes pretesting control questions with a goal of

Institute for Homeland
Security Solutions
Applied research • Focused results

finding a negatively correlated mix of such questions. For example, if a U.S. researcher asked respondents whether the U.S. invasion of Iraq, federal tax increases, or new environmental regulations make them angry, it is unlikely that many respondents would respond affirmatively to all three questions. Glynn also offers useful advice on experimental design and provides an estimator to determine the minimum necessary sample size for a particular list experiment.

Conclusion

Surveys are a powerful tool for collecting information about the attitudes, characteristics, and behaviors of individuals. This is particularly true when the subject of interest extends across large populations, because survey methods are in many ways well suited for administration to (comparatively small) samples that permit generalization to such larger groups. Yet administering a survey in such a way as to collect meaningful,—and, particularly, generalizable—data is a challenging task involving many judgment calls: every decision made has the potential to introduce error into the resulting survey statistics, that is, into the value and meaning of the findings. There are at least three challenges of particular concern that should be considered in making survey design decisions related to studying public attitudes toward terrorism and counterterrorism strategies: (1) selecting an appropriate sampling frame or frames; (2) accurately collecting information about sensitive topics; and (3) ensuring that the attitudinal data collected offer meaningful insights into likely behaviors. This literature review highlights some of the tools available to address these challenges and to produce a survey project that will provide for a more complete understanding of public attitudes toward terrorism and counterterrorism strategies in the United States.

Authors

Brad Bishop, MA, Duke University

Alexandra Cooper, PhD, Duke University

D. Sunshine Hillygus, PhD, Duke University

References

Arkes, H., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or "Would Jesse Jackson 'fail' the implicit association test?" *Psychological Inquiry, 15,* 257–278.

Bassili, J. N., & Fletcher, J. F. (1991). Response-time measurement in survey research: A method for CATI and a new look at nonattitudes. *Public Opinion Quarterly, 55,* 331–346.

Berinsky, A. (2004). *Silent voices: Opinion polls and political representation in America.* Princeton, NJ: Princeton University Press.

Institute for Homeland Security Solutions
Applied research • Focused results

Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to survey quality*. Hoboken, NJ: Wiley.

Brader, T. (2005). Striking a responsive chord: How political ads motivate and persuade voters by appealing to emotions. *American Journal of Political Science, 49,* 388–405.

Brader, T., Valentino, N. A., & Suhay, E. (2008). What triggers public opposition to immigration? Anxiety, group cues, and immigration threat. *American Journal of Political Science, 52,* 959–978.

Callegaro, M., & DiSogra, C. (2008). Computing response metrics for online panels. *Public Opinion Quarterly, 72,* 1008–1032.

Cannell, C. F., Oksenberg, L., & Converse, J. M. (1981). Research on interviewing techniques. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 389-437). San Francisco: Jossey-Bass.

Cantor, D., O'Hare, B. C., & O'Connor, K. S. (2006, January). *The use of monetary incentives to reduce nonresponse in random digit telephone surveys.* Paper presented at the Second International Conference on Telephone Survey Methodology (TSM II), Miami, FL.

Clark, R. G. (2009). Sampling of subpopulations in two-stage surveys. *Statistics in Medicine, 28,* 3697–3717. Retrieved from http://dx.doi.org/10.1002/sim.3723

Cobb, M., & Kuklinski, J. (1997). Changing minds: Political arguments and political persuasion. *American Journal of Political Science, 41,* 88–121.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York: Wiley.

Corstange, D. (2009). Sensitive questions, truthful answers? Modeling the list experiment with LISTIT. *Political Analysis, 17,* 45–63.

Couper, M. P., & Miller, P.V. (2008). Web survey methods. *Public Opinion Quarterly, 72,* 831–835.

Couper, M. P., Tourangeau, T., & Marvin, T. (2009). Taking the audio out of audio-CASI. *Public Opinion Quarterly, 73,* 281–303.

David, H. A. (1969). *The method of paired comparisons*. London: Charles Griffin.

Davis, R. E., Couper, M., Janz, N., Caldwell, C., & Resnicow, K. (in press). Interviewer effects in public health surveys. *Health Education Research.* Published online September 17, 2009. doi:10.1093/her/cyp046

De la Puente, M., Pan, Y., & Rose, D. (2003, November). An overview of a proposed Census Bureau guideline for the translation of data collection instruments and supporting materials. *Proceedings of the Federal Committee on Statistical Methodology Research Conference.* Retrieved from http://www.fcsm.gov/03papers/delaPuente_Final.pdf

De Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. The *Journal of Official Statistics, 21,* 233–255.

De Leeuw, E. D., & Hox, J. J. (2004). I am not selling anything: 29 experiments in telephone introductions. *International Journal of Public Opinion Research, 16,* 464–473.

De Leeuw, E., Callegaro, M., Hox, J., Korendijk, E., & Lensvelt-Mulders, G. (2007). The influence of advance letters on response in telephone surveys: A meta-analysis. *Public Opinion Quarterly, 71,* 413–443.

Dijkstra, W., & Smith, J. H. (2002). Persuading reluctant recipients in telephone surveys. In R. M. Groves, D. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 121–134). New York: Wiley.

Dillman, D. A. (2006). Why the choice of survey mode makes a difference. *Public Health Reports, 121,* 11–13.

Dillman, D. A. (2007). *Mail and Internet surveys (*2nd ed.). Hoboken, NJ: Wiley.

Elliott, M. N., McCaffrey, D., Perlman, J., Marshall, G. N., & Hambarsoomians, K. (2009). Use of expert ratings as sampling strata for a more cost-effective probability sample of a rare population. *Public Opinion Quarterly, 73,* 56–73.

Finn, A., & Louviere, J. J. (1992). Determining the appropriate response to evidence of public concern: The case of food safety. *Journal of Public Policy and Marketing, 11,* 12–25.

Forman, E. H., & Gass, S. I. (2001). The analytic hierarchy process: An exposition. *Operations Research, 49,* 469–486.

Fowler, F. J., Jr. (1995). *Improving survey questions: Design and evaluation.* Newbury Park, CA: Sage.

Fowler, F. J., Jr. (2009). *Survey research methods* (4th ed.). Thousand Oaks, CA: Sage.

Fowler, F. J., & Cannell, C. F. (1996). Using behavioral coding to identify cognitive problems with survey questions. In N.A. Schwartz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 15–36). San Francisco: Jossey-Bass.

Fowler, F. J., Gallagher, P. M., & Nederend, S. (1998). Comparing telephone and mail responses to the CAHPS survey instrument. *Medical Care, 37* (3, Suppl.), MS41–MS49.

Fricker, S., Galesic, M., Tourangeau, R., & Yan, T. (2005). An experimental comparison of Web and telephone surveys. *Public Opinion Quarterly, 69,* 370–392.

Gaines, B. J., Kuklinski, J. H., & Quirk, P. J. (2007). The logic of the survey experiment reexamined. *Political Analysis, 15,* 1–20.

Gilens, M., Sniderman, P. M., & Kuklinski, J. H. (1998). Affirmative action and the politics of realignment. *British Journal of Political Science, 28,* 159–183.

Glynn, A. N. (2009). On the design of statistical truth serum: Getting more from the list experiment with sample size calculation, design principles, and the probabilistic imputation of socially undesirable responses (Working paper). Retrieved from http://scholar.iq.harvard.edu/aglynn/publications/what-can-we-learn-statistical-truth-serum-design-and-analysis-list-experiment

Graesser, A. C., Cai, Z., Louwerse, M. M., & Daniel, F. (2006). Question Understanding Aid (QUAID): A Web facility that tests question comprehensibility. *Public Opinion Quarterly, 70,* 3–22.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74,* 1464–1480.

Groves, R. M., Couper, M. P., Presser, S., Singer, E., Tourangeau, R., Acosta, G. P., et al. (2006). Experiments in producing nonresponse bias. *Public Opinion Quarterly*, *70,* 720–736.

Groves, R. M., Fowler, F. J., Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey methodology*. Hoboken, NJ: Wiley.

Groves, R. M., & McGonagle, K. (2001). A theory-guided interviewer training protocol regarding survey participation. *Journal of Official Statistics, 17,* 249–265.

Harmon, T., Turner, C. F., Rogers, S. M., Eggleston, E., Roman, A. M., Villarroel, M. A., et al. (2009). Impact of T-ACASI on survey measurements of subjective phenomena. *Public Opinion Quarterly, 73,* 255–280.

Heerwegh, D., & Loosveldt, G. (2008). Face-to-face versus Web surveying in a high-Internet-coverage population: Differences in response quality. *Public Opinion Quarterly, 72,* 836–846.

Hillygus, D. S., Nie, N. H., Prewitt, K., & Pals, H. (2006). *The hard count: The political and social challenges of census mobilization*. New York: Russell Sage Foundation.

Huckfeldt, R., Levine, J., Morgan, W., & Sprague, J. (1999). Accessibility and the political utility of partisan and ideological orientations. *American Journal of Political Science, 43,* 888–911.

Huddy, L., Feldman, S., Taber, C., & Lahav, G. (2005). Threat, anxiety, and support for antiterrorism policies. *American Journal of Political Science, 49,* 593–608.

Joslyn, M. (2001). [Review of the book *Affective intelligence and political judgment*]. *American Political Science Review, 95,* 1003–1004.

Kalsbeek, W. D. (2003). Sampling minority groups in health surveys. *Statistics in Medicine, 22,* 1527–1549.

Kalton, G., & Anderson, D. (1986). Sampling rare populations. *Journal of the Royal Statistical Society, Series A, 149,* 65–82.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the implicit association test. *Journal of Personality and Social Psychology, 81,* 774–788.

Kazprzyk, D., Duncan, G., Kalton, G., & Singh, M. P. (1989). *Panel surveys.* New York: Wiley.

Keeter, S., Kennedy, C., DiMock, M., Best, J., & Craighill, P. (2006). Gauging the impact of growing nonresponse from a national RDD telephone survey. *Public Opinion Quarterly, 70,* 759–779.

King, G., Murray, C. J. L., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, *98,* 191–207.

King, G., & Wand, J. (2007). Comparing incomparable survey responses: Evaluating and selecting anchoring vignettes. *Political Analysis, 15,* 46–66.

Kish, L. (1965). *Survey sampling.* New York: Wiley.

Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and Web surveys: The effects of mode and question sensitivity. *Public Opinion Quarterly, 72,* 847–865.

Krosnick, J. A. (1990). Government policy and citizen passion: A study of issue publics in contemporary America. *Political Behavior*, *12,* 59–92.

Kuklinski, J. H., Cobb, M. D., & Gilens, M. (1997). Racial attitudes and the "New South." *Journal of Politics, 59,* 323–349.

Kuklinski, J. H., Sniderman, P. M., Knight, K., Piazza, T., Tetlock, P. E., Lawrence, G. R., et al. (1997). Racial prejudice and attitudes toward affirmative action. *American Journal of Political Science, 41,* 402–419.

Lessler, J. T., & Forsyth, B. H. (1996). A coding system for appraising questionnaires. In N.A. Schwartz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 259–291). San Francisco: Jossey-Bass.

Lohr, S. L. (1998). *Sampling design and analysis.* New York: Brooks/Cole.

Louviere, J. J. (1993). *The best-worst or maximum difference measurement model: Applications to behavioral research in marketing.* Paper presented at the behavioral research conference of the American Marketing Association, Phoenix, Arizona.

Institute for Homeland Security Solutions
Applied research • Focused results

Marcus, G. E., Neuman, W. R., & MacKuen, M. (2000). *Affective intelligence and political judgment.* Chicago: University of Chicago Press.

Marin, G., & Marin, B. V. (1991). *Research with Hispanic populations.* Newbury Park, CA: Sage.

McCauley, C., & Moskalenko, S. (2008). Mechanisms of political radicalization: Pathways toward terrorism. *Terrorism and Political Violence, 20,* 415–433.

McDermott, R. (2002). Experimental methodology in political science. *Political Analysis, 10,* 325-342.

Mingay, D. M. (2000, May). *The strengths and limitations of telephone audio computer-assisted self-interviewing (T-ACASI): A review.* Paper presented at the annual meeting of the American Association of Public Opinion Research, Portland, OR.

Morton, R. B., & Williams, K. C. (2008). Experimentation in political science. In J. M. Box-Steffensmeier, H. E. Brady, & D. Collier (Eds.), *The Oxford handbook of political methodology* (pp. 339–356). New York: Oxford University Press.

Moskalenko, S., & McCauley, C. (2009). Measuring political mobilization: The distinction between activism and radicalism. *Terrorism and Political Violence, 21,* 239–260.

Mulligan, K., Grant, J. T., Mockabee, S. T., & Monson, J. Q. (2003). Response latency methodology for survey research: Measurement and modeling strategies. *Political Analysis, 11,* 289–301.

Muris, P., Mayer, B., van Eijk, S., & van Dongen, M. (2008). "I'm not really afraid of Osama Bin Laden!" Fear of terrorism in Dutch children. *Journal of Child and Family Studies, 17,* 706–713.

Orme, B. (2005). Accuracy of HB estimation in MaxDiff experiments (Sawtooth Software research paper series). Sequim, WA: Sawtooth Software. Retrieved from http://www.sawtoothsoftware.com/education/techpap.shtml

Payne, K., Cheng, C. M., Govorun, O., & Stewart, B. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89,* 277–293.

Pew Research Center for the People and the Press (2007). *Muslim Americans: Middle class and mostly mainstream.* Retrieved from http://pewresearch.org/pubs/483/muslim-americans

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., et al. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience, 12,* 729–738.

Institute for Homeland Security Solutions
Applied research • Focused results

Portney, P. R. (1994). The contingent valuation debate: Why economists should care. *Journal of Economic Perspectives, 8,* 3–17.

Ronquillo, J., Denson, T. F., Lickel, B., Lu, Z., Nandy, A., & Maddox, K. B. The effects of skin tone on race-related amygdala activity: An fMRI investigation. *Social Cognitive and Affective Neuroscience, 2,* 39–44.

Sanchez, M. E., & Morchio, G. (1992). Probing "don't know" answers: Effects on survey estimates and variable relationships. *Public Opinion Quarterly, 56,* 454–474.

Schlenger, W. E., & Silver, R. C. (2006). Web-based methods in terrorism and disaster research. *Journal of Traumatic Stress, 19,* 185–193.

Schober, M. F., & Conrad, F. G. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly, 61,* 576–602.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys.* New York: Academic Press.

Sigelman, L. (1982). The nonvoting voter in voting research. *American Journal of Political Science, 26,* 47–56.

Silver, B., Anderson, B., & Abramson, P. (1986). Who overreports voting? *American Political Science Review, 80,* 613–624.

Singer, E. (2002). The use of incentives to reduce nonresponse in household surveys. In R.M. Groves, D. Dillman, J.L. Eltinge, and R.J.A. Little (Eds.), *Survey nonresponse* (pp. 163–178). New York: Wiley.

Singer, E., Mathiowetz, N. A., & Couper, M. P. (1993). The impact of privacy and confidentiality concerns on survey participation: The case of the 1990 Census. *Public Opinion Quarterly, 57,* 465–482.

Siram, N., & Greenwald, A. G. (2009). The brief implicit association test. *Experimental Psychology, 56,* 283–294.

Sniderman, P. M., & Piazza, T. (1993). *The scar of race.* Cambridge, MA: Harvard University Press.

Streb, M. J., Burrell, B., Frederick, B., & Genovese, M. A. (2008). Social desirability effects and support for a female American president. *Public Opinion Quarterly, 72,* 76–89.

Traugott, M., & Kattosh, J. (1979). Response validity in surveys of voting behavior. *Public Opinion Quarterly, 43,* 359–377.

Treier, S. A., & Hillygus, D. S. (in press). *The nature of political ideology in the contemporary electorate. Public Opinion Quarterly.*

Institute for Homeland
Security Solutions
Applied research • Focused results

Tucker, C., Brick, J. M., & Meekins, B. (2007). Household telephone service and usage patterns in the United States in 2004: Implications for telephone samples. *Public Opinion Quarterly, 71,* 3–22.

Turner, C., Forsyth, B., O'Reilly, J., Cooley, P., Smith, T., Rogers, S., et al. (1998). Automated self-interviewing and the survey measurement of sensitive behaviors. In M. Couper, R. Baker, J. Bethlehem, C. Clark, J. Martin, W. Nicholls II, & J. O'Reilly (Eds.), *Computer assisted survey information collection* (pp. 455–474). New York: Wiley.

Uhlmann, E. L., Brescoll, V. L., & Paluck, E. L. (2006). Are members of low status groups perceived as bad, or badly off? Egalitarian negative associations and automatic prejudice. *Journal of Experimental Social Psychology, 42,* 491–499.

Wang, H. (1997). Treatment of "don't know" responses in contingent valuation surveys: A random valuation model. *Journal of Environmental Economics and Management, 32,* 219–232.

Institute for Homeland Security Solutions
Applied research • Focused results