THE EFFECTS OF METACOGNITION IN SURVEY RESEARCH
EXPERIMENTAL, CROSS-SECTIONAL, AND CONTENT-ANALYTIC EVIDENCE

MATTHEW D. SWEITZER
HILLARY C. SHULMAN*

Abstract Three studies examined the role of language difficulty in public opinion questions. Guided by feelings-as-information theory (FIT; Schwarz 2011), the first study presents an experiment (N = 1,018) in which the language difficulty of public opinion questions was varied. Findings suggest that language difficulty affected metacognitive experiences, which in turn affected reports of political interest, political efficacy, *don't know* responding, and ideological constraint. Study 2 (N = 1,817) presents cross-sectional evidence from publicly available data that also indicates question-language difficulty influences *don't know* responding. Given these findings, study 3 (N = 8,090) presents a content analysis that reveals significant systematic variability in language difficulty within polling questions across 10 polling firms in 2016. Contextualizing these findings within a FIT framework, we contend that variability in language difficulty differentially and systematically affects participants' metacognitive experiences while responding to public opinion questions. Given that metacognitive experiences affect survey response, language difficulty ought to be more carefully considered when drafting opinion questions. To this end, the data presented in these studies can be used to aid question construction by providing numeric guidelines, using widely available measures that contextualize the relative difficulty of survey language. It is also recommended that items assessing metacognitive experiences be included in survey

MATTHEW D. SWEITZER is a doctoral student in the School of Communication at Ohio State University, Columbus, OH, USA. HILLARY C. SHULMAN is an assistant professor in the School of Communication at Ohio State University, Columbus, OH, USA. The authors thank the editors and three anonymous reviewers for their comments and suggestions. *Address correspondence to Hillary Shulman, Ohio State University, School of Communication, 154 N. Oval Mall, Columbus, OH 43210, USA; email: hillaryshulman@gmail.com.

doi:10.1093/poq/nfy034

research to account for variance in this measure. At a time when polling data are ubiquitous yet polling accuracy is being called into question, it is critical to identify sources of unmeasured error within polling data.

Public opinion polls strive to solicit valid opinions from a representative sample. Given that the quality of survey response is predicated on the quality of the survey itself, an abundance of scholarly attention has been devoted to the construction of these instruments (e.g., Schuman and Presser 1996; Schwarz 1999; Presser et al. 2004; Holbrook, Cho, and Johnson 2006; Krosnick, Malhotra, and Mittal 2014). Yet, despite the proliferation of work in this area, there has been a paucity of research linking metacognition to survey design. To remedy this gap, this paper presents three studies utilizing different methodologies to explore the effect of language difficulty on metacognition and survey response. The first, guided by feelings-as-information theory (FIT; Schwarz 2011), provides experimental evidence to evince how language difficulty and metacognition affect survey response. The second study uses publicly available data to examine whether noteworthy findings from study 1 can be replicated. Through content analysis, study 3 aims to contextualize the potential impact of the findings from studies 1 and 2. Taken together, this work underscores the need to pay close attention to the difficulty of language employed within survey instruments.

Research on survey design has long acknowledged that factors inherent to a survey but tangential to its purpose can affect outcomes in unintended ways (Schwarz 1999). Some of these factors include question order (e.g., Bishop, Oldendick, and Tuchfarber 1984), response alternatives (e.g., Mondak 2001; Tourangeau, Maitland, and Yan 2016), language comprehension (e.g., Graesser et al. 2006), survey mode (e.g., Shulman and Boster 2014; Clifford and Jerit 2016), and question wording (e.g., Pan and Kosicki 1993; Druckman 2001; Krosnick, Malhotra, and Mittal 2014). In this vein, this research examines language difficulty. Because survey organizations strive to improve question comprehension through simplifying language, whether this practice unknowingly affects information processing and survey outcomes is the purpose of these investigations.

Information Processing and Survey Response

To explicate metacognitive experiences, we first introduce information accessibility, which refers to the information people retrieve from their memory when asked to render an opinion (Tourangeau and Rasinski 1988). Information accessibility has garnered attention across domains due to its prominence in memory-based models of information processing. Memory-based models contend that when a question is posed, people search their memory for all

applicable pieces of information. The information retrieved is accessible, whereas information not retrieved is less accessible. The importance of information accessibility based on this premise is that only accessible information can be used to inform preferences. Relating this process to survey research, Zaller and Feldman (1992) argue that the best way to explain survey response is accessibility, such that people respond based on what considerations are evoked.

Although a lot of what is known regarding information processing and survey response is predicated on information accessibility, a similarly influential part of this process is less often considered. Metacognitive experiences accompany information accessibility (Schwarz 1998), yet refer to a distinct cognitive process. To understand this distinction, a discussion of primary and secondary cognitions is useful (Petty et al. 2007). Primary cognitions refer to the declarative information recalled when someone is asked their opinion (i.e., information accessibility). Secondary cognitions, or metacognition, refer to thinking about thinking, or how thoughts associated with the information retrieval process inform preferences. Metacognitive experiences, one type of secondary cognition, refer to "the subjective experience of the ease with which people process information" (Alter and Oppenheimer 2009, p. 219; Schwarz 2010). These experiences reflect affective judgments related to the experiential information that comes to mind. If information comes to mind easily, it is more likely to be tagged as correct and used, whereas when information is difficult to retrieve, people are less confident in its utility (Schwarz 2006, 2011). Because these affective responses influence judgments (Winkielman et al. 2003), it is important to consider both accessible information and metacognitive experiences when soliciting preferences. As Alter and Oppenheimer (2009) argue, "Researchers who fail to recognize the effects of [metacognitive] experiences on cognition might unwittingly introduce confounds into their studies, undermining the phenomena they seek to identify" (p. 232).

Feelings-as-Information Theory

Feelings-as-information theory (FIT; Schwarz 2011) contends that people use metacognition in the same way as informational content. When information processing is experienced as easy, people feel positively about the information at hand, and become more likely to rely on this information in the formation of preferences (Schwarz 2006). Guided by the idea that an easy metacognitive experience "feels good" (Schwarz 2006, 2010), these experiences have been shown to affect preferences in many ways (see Petty et al. 2007; Alter and Oppenheimer 2009).

Here, the connection between metacognition research and public opinion research is through language comprehension. In survey research, comprehension is defined as participants' ability to understand linguistically and

conceptually the language and intent of the question (Holbrook, Cho, and Johnson 2006). Factors that may impede comprehension include: unfamiliar and difficult terms, vague terms, vague or unfamiliar referents, complex syntax, or cognitive load (Graesser et al. 2006). Given that participants' ability to comprehend a question is a necessary condition for providing reliable and/or valid responses, survey designers attempt to create questions that are as comprehensible as possible.

The relationship between language difficulty and metacognition has also been explored. For example, Oppenheimer (2006) had participants read either an original college admittance essay or a highly complex version in which words from the original were replaced with the longest entry in a thesaurus. Consistent with expectations, those in the complex condition encountered a more difficult metacognitive experience than those who read the original essay, and consequently negatively evaluated the essay. Thus, differences in language difficulty affect metacognition. This implies that as researchers improve question comprehension, they may be inadvertently inducing variance in survey takers' metacognitive experiences.

Metacognitive experiences have been found to positively influence reports of liking, confidence, and perceptions of knowledge (Schwarz 1998, 2010, 2011; Petty et al. 2007; Alter and Oppenheimer 2009). These findings have been replicated using critical variables from the political science literature, including interest, internal efficacy, and perceived knowledge (Shulman and Sweitzer 2018a, 2018b). These variables are vital to political scientists because they positively associate with important outcomes (e.g., Zaller 1992; Delli Carpini and Keeter 1996; Prior 2010). Given that language difficulty affects metacognition and that these experiences systematically impact frequently utilized political variables, conclusions drawn from prior research may be confounded by variability in metacognitive experiences.

There are also practical implications that arise from unintentionally inducing metacognitive experiences alongside language difficulty. In an experiment conducted by Shulman and Sweitzer (2018a), participants exposed to an easy-language condition reported an easier experience and, in turn, attitudes that more strongly cohered to their political ideology than those in the difficult-language condition. They also found that as experiences were eased, survey responses became less ambivalent. These findings evince that language difficulty and metacognitive experiences impact the substantive nature of opinion data. This underscores the importance of identifying mechanistically how language difficulty affects metacognition and, in turn, survey results. Though previous research has documented these relationships using student samples, the current studies extend these ideas in order to understand whether experimental effects replicate to the general population (study 1), whether notable findings from study 1 replicate in publicly available data (study 2), and the degree to which variability in language difficulty exists in a broad array of polling questions through content analysis (study 3).

Study 1

HYPOTHESES

Four hypotheses guide study 1, predicting that metacognitive experiences will mediate the relationship between language condition and (H1) political interest, (H2) internal political efficacy, and (H3) the frequency of *don't know* responses. Specifically, easier experiences should positively associate with interest and efficacy and negatively associate with *don't know* responding. H4 predicts that the language condition will moderate the relationship between ideology and metacognitive experiences; as experiences become easier, the relationship between ideology and political attitudes will strengthen as a result. This expectation reflects our assumption that an easier experience should help with the evocation of more reliable attitudes. If one were to assume that true political preferences are ideologically guided, then making these positions easier to access through language should strengthen the relationship between ideology and political attitudes (Shulman and Sweitzer 2018a).

Method

PARTICIPANTS

Participants (N = 1,018) were recruited, and compensated, through Qualtrics' online panel.[1] Participants were 61.8 percent female (age: M = 50.71, SD = 15.70), with 81.7 percent identifying as Caucasian, 7.4 percent as African American, 3.0 percent as Asian or Asian American, 0.5 percent as Native American, 5.6 percent as Hispanic, 0.4 percent as other, and 1.4 percent as mixed. Participants had to be at least 18 years old and eligible to vote in US elections.

PROCEDURE

Participants were randomly assigned to either an easy- (n = 514) or a difficult-language condition (n = 504). All eligible participants were presented with five public opinion questions and three metacognitive experience questions for three issues (24 questions in total; see supplementary material online). Questions were introduced in the same order across conditions. Participants responded to these items using a scale ranging from (1) *strongly disagree* to (7) *strongly agree*, along with a *don't know* option. The three political issues, in survey order, were the economy, education, and infrastructure policy.

1. Qualtrics uses opt-in panels. The AAPOR participation rate for this study is approximately 3.73 percent.

Following these questions, participants were asked about their political interest, efficacy, and ideology.

LANGUAGE CONDITION CREATION

The language conditions were designed to include easy or difficult words using a thesaurus (Oppenheimer 2006; Shulman and Sweitzer 2018b; tables 1, S1 and S2). This manipulation was verified using the *koRpus* package (Michalke 2016) in *R*, including the Flesch Reading Ease scale (FRE; Flesch 1948). Consistent with expectations, the easy condition (M = 59.37, SD = 5.38) included denotatively easier words than the difficult condition (M = 21.14, SD = 4.99, t (10) = 11.82, $p < 0.001$, d = 7.48).[2]

METACOGNITIVE EXPERIENCES

Metacognitive experiences were measured from three items that followed opinion questions on each issue, with higher scores reflecting an easier experience. These items asked whether: 1) the language used in the questions felt difficult; 2) the information presented felt new; or 3) it was easy to provide opinions when answering these questions (table S3). Importantly, these items capture individual differences in one's experience while answering the questions and do not address more objective issues of comprehension (Shulman and Sweitzer 2018a, 2018b). Moreover, these items were written to reflect how participants *felt* while answering the questions to capture the experiential quality inherent to metacognitive processes. Items were aggregated across issues to produce a nine-item scale (M = 5.15, SD = 1.14, $\alpha = 0.88$, range 1–7). Participants in the easy condition reported easier experiences (M = 5.42, SD = 0.97) than those in the difficult condition (M = 4.88, SD = 1.23, t (1016) = 7.71, $p < 0.001$, d = 0.49). As expected, issue-specific scales were positively associated with FRE for the economy, r (1016) = 0.22, $p < 0.001$; education, r (1016) = 0.22, $p < 0.001$; and infrastructure issues, r (1016) = 0.21, $p < 0.001$.

MEASURES

Political interest was measured using a four-item scale (M = 4.73, SD = 1.57, $\alpha = 0.93$). Internal political efficacy was measured using Niemi, Craig, and Mattei's (1991) four-item scale (M = 4.48, SD = 1.37, $\alpha = 0.85$, table S3).

Don't know responding was determined by participants' selection of the *don't know* response within the 15 attitude items. This option allowed for the distinction between a lack of opinion and nonresponse (Mondak 2001). In

2. These statistics represent the aggregate of 12 collapsed sub-conditions (six each collapsed to "easy" and "difficult"). See the supplementary material online (tables S1 and S2).

Table 1. Example political attitude questions

Economy easy	Economy difficult
*U.S. bank regulators should raise interest rates to try to keep the prices of everyday items about the same.	*The U.S. Federal Reserve Bank should increase interest rates in an effort to inhibit price inflation on conventional merchandise.
*The struggling world market would be better off if more leaders agreed to trade goods fairly with others.	*The international state of economic stagnation could be counteracted if more leaders sign equitable commerce agreements with others.
Lawmakers help new businesses by reducing the number of rules and policies that are needed to do business.	Legislators can promote new business ventures by reducing the amount of corporate parameters and regulations required to operate.
Company presidents create jobs and they should not be asked to pay a lot of money in taxes.	Chief Executive Officers produce employment opportunities for people and should not be consigned to a higher tax bracket.
I believe that a smart economic plan is to give less U.S. government money to funding public assistance programs.	I believe that a sound fiscal strategy is to appropriate less federal revenue towards funding of public assistance programs.
Education easy	**Education difficult**
*I support college scholarships that make it easier for students of certain backgrounds to attend school.	*I support collegiate scholarships earmarked towards making a university education accessible for students of certain backgrounds.
U.S. public school teachers should be paid based on their students' performance, not how long they've worked at the school.	U.S. public educators should be remunerated based upon metrics of pupils' performance, rather than as a function of their tenure.
State colleges should not use standardized test scores when they decide which students should be admitted.	Admissions committees at publicly funded universities should not use standardized test scores for their admissions decisions.
Classes in the creative arts are not important to a well rounded education.	Curricula focusing on creative arts are inessential to a holistic approach to instruction.
*School choice programs that allow students to go to any school they want are a bad way to address the problems seen in public schools.	*School voucher programs that entitle students to attend the institution of their choosing are an ineffective way to address the problems inherent in public schools.

Continued

Table 1. *Continued*

Infrastructure easy	Infrastructure difficult
The U.S. government should be willing to provide money in an effort to support public projects.	The federal government should be willing to allocate resources in an effort to subsidize public projects.
I would support some sort of government program that tries to increase the number of jobs within the area of construction.	I would support some sort of government initiative aimed at facilitating a higher level of employment within the domain of infrastructure.
The transportation system in the U.S. is currently in need of much repair.	The infrastructure system within the U.S. is currently in a state of dilapidation.
*Our main energy sources use technology that is out of date.	*Our primary energy system utilizes technologies that can be considered antiquated.
*It is the job of local government to make sure that all buildings located in the community are up to code.	*It is the responsibility of local municipalities to ensure that all construction within the district adheres to safety standards and regulations.

NOTE.—*Indicates items that were reverse coded for testing hypothesis 4. Please see the online supplementary materials for response options and Flesch Reading Ease scores.

total, this option was selected 638 times (4.2 percent). Political attitudes were measured by averaging responses across the 15 opinion questions ($M = 3.92$, $SD = 0.66$, $\alpha = 0.55$). To create this scale, six of the political attitude items were recoded so that higher scores reflect more conservative preferences (tables 1, S1, and S2). This was done to maintain consistency with the ideology measure, which was operationalized using the seven-point ANES political ideology measure ranging from (1) *very liberal* to (7) *very conservative* ($M = 4.16$, $SD = 1.89$).

Results

Analyses were conducted using Hayes's (2016) macro PROCESS. To address hypotheses, Hayes's simple mediation model (model 4, 95 percent bias-corrected bootstrap CIs based on 10,000 resamples) was used to test whether language condition affected metacognitive experiences, and whether these experiences influenced reports of (H1) interest, (H2) efficacy, and (H3) *don't know* responses. For H1, significant indirect effects were obtained as hypothesized, $B = 0.23$, $SE = 0.04$, $R^2 = 0.09$, 95 percent CI [0.16, 0.31], suggesting that easier language eased metacognitive experiences ($B = 0.54$, $SE = 0.07$,

$t = 7.71$, $p < 0.001$), which in turn led to higher reports of political interest ($B = 0.42$, $SE = 0.04$, $t = 9.87$, $p < 0.001$). Similarly, for H2, significant indirect effects were obtained as hypothesized, $B = 0.18$, $SE = 0.03$, $R^2 = 0.07$, 95 percent CI [0.12, 0.25], suggesting that easier language eased metacognitive experiences ($B = 0.54$, $SE = 0.07$, $t = 7.71$, $p < 0.001$), which in turn led to higher reports of internal political efficacy ($B = 0.33$, $SE = 0.04$, $t = 8.82$, $p < 0.001$). H3 was also supported with significant indirect effects in the hypothesized direction, $B = –0.28$, $SE = 0.06$, $R^2 = 0.10$, 95 percent CI [–0.41, –0.17], suggesting that an easier experience was negatively associated with *don't know* responses, $B = –0.52$, $SE = 0.05$, $t = –9.88$, $p < 0.001$. Additionally, we ran an item-level logit analysis ($n = 15{,}270$) regressing *don't know* responses on reading ease (figure 1). As expected, the log-likelihood of selecting *don't know* significantly decreased with each unit increase in FRE; $B = –0.013$, $SE = 0.002$, $Z = –7.15$, $p < 0.001$, providing response-level support of H3.[3]

H4 was tested using Hayes's (2016) PROCESS model 3 (three-way moderation, 95 percent bias-corrected bootstrap CIs based on 10,000 resamples). Overall, this model was significant, $F(7, 1002) = 71.08$, $R^2 = 0.33$. Specifically, language condition ($B = 0.87$, $SE = 0.42$, $t = 2.08$, $p < 0.05$) and ideology ($B = 0.11$, $SE = 0.05$, $t = 2.07$, $p < 0.05$) were significant predictors of political attitudes, whereas metacognitive experiences were not ($B = –0.09$, $SE = 0.05$, $t = –1.94$, $p = ns$). The only significant interaction effect was ideology by metacognitive experiences ($B = 0.02$, $SE = 0.01$, $t = 2.07$, $p < 0.05$). As illustrated by the conditional effects, for those assigned to the easy condition, and who self-reported the highest (i.e., easiest) metacognitive experiences, ideology was a significantly

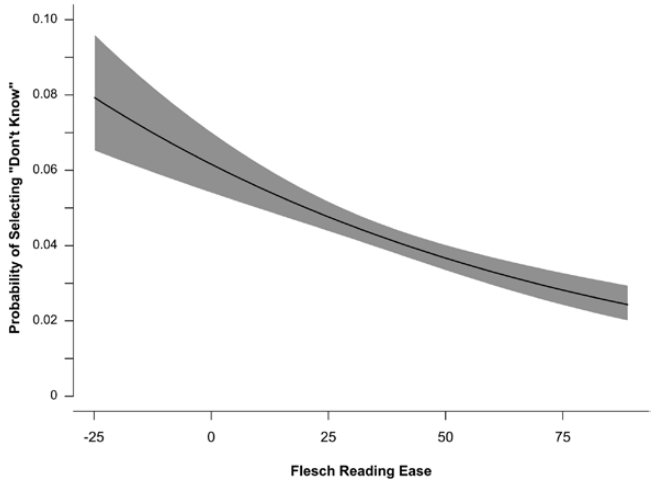

Figure 1. Probability of selecting *don't know* by FRE—Qualtrics sample. Shaded area indicates the 95 percent confidence interval of the estimate.

3. See additional analysis 1 in the supplementary material online regarding this analysis.

stronger predictor of political attitudes ($B = 0.24$, $SE = 0.02$, 95 percent CI [0.20, 0.27]) than for those assigned to the difficult condition who reported the lowest (i.e., most difficult) experiences ($B = 0.11$, $SE = 0.02$, 95 percent CI [0.07, 0.16]). This finding is consistent with H4 and suggests that using easier language to promote an easier metacognitive experience improves participants' ability to respond to opinion questions in ideologically consistent ways. This finding carries important implications for survey quality and measurement reliability.

Discussion

These results demonstrate that variability in metacognitive experiences, induced through language difficulty, affects the reporting of critical variables. Specifically, we found that when language was simplified and a metacognitive experience was facilitated, participants reported more interest in politics, higher political efficacy, were less likely to report *don't know*, and reported preferences more consistent with their ideology. Taken together, these findings proffer support for considering the impact of language difficulty and metacognition in public opinion polling.

Although the aforementioned findings are theoretically and empirically noteworthy, it is important to test whether these trends replicate in existing polling data. As such, study 2 tests the relationship between language difficulty and *don't know* responses in an analysis of publicly available data.

Study 2

In order to generalize findings from study 1, we set out to identify the effect of language difficulty using polling data from other organizations. To do so, we gathered all available datasets for surveys conducted during 2016 using the data aggregating service, the Roper Center for Public Opinion Research (Roper n.d.). After reading the codebooks, it became apparent that the only relationship that could feasibly be tested from study 1 was the effect of language difficulty on *don't know* responses (H3).[4] Thus, the hypothesis guiding study 2 predicted that language ease would be inversely related to *don't know* responding.

Method

There were three inclusion criteria for this study: (1) the dataset was available from Roper; (2) the survey was administered during 2016; and (3) *don't know* responses were differentiated from nonresponse. Questions pertaining to demographic information, behaviors, or knowledge were omitted because a

4. Political interest and efficacy measures were not widely available. Additionally, attitude constraint could not be measured without recoding opinion questions such that responses fall on a left- to right-leaning scale. Unfortunately, this analysis was not possible given the size, scope (e.g., nonpolitical questions), and variability of response scales included in this sample.

don't know response to these types of questions has a different meaning than in response to attitude/opinion measures. These criteria resulted in a sample of 1,817 questions, from 70 surveys (# of questions: $M = 25.96$, $SD = 28.78$, range = 2–218), conducted by 16 polling firms (table 2).

MEASURES

Language difficulty was measured using the FRE scale (Flesch 1948), using the *koRpus* package in *R* (Michalke 2016). Although language difficulty was "fairly difficult," this measure included considerable variability ($M = 59.89$, $SD = 29.61$, range = –133.59 to 121.22) and therefore was standardized for analysis.

Don't know responding was measured using each response as the unit of analysis. A *don't know* response was coded as 1, whereas other responses were coded as 0. We did not distinguish between *don't know* and *no opinion* responses, in order to include as much data as possible. Of the 1.86 million

Table 2. Descriptive statistics of polling firms included in study 2

Polling firm	M_{FRE}	SD_{FRE}	Number of valid instruments	Number of valid questions
ABC/*Washington Post*	59.40	17.47	2	125
American Enterprise Institute	61.88	14.48	1	36
AP/National Opinion Research Center	56.02	38.99	12	326
Fannie Mae	81.12	11.93	12	132
Kaiser Family Foundation	57.93	26.83	11	370
KFF/CNN	66.63	22.88	1	50
NORC	58.88	39.42	1	218
NPR	80.30	21.16	1	18
NY Times Upshot/Siena	58.37	13.23	6	56
Phi Delta Kappa International	70.86	17.46	1	44
Pew Research Center	37.50	23.29	5	28
Politico	63.56	15.52	1	35
Paul Simon Public Policy Institute	55.53	20.25	1	39
Rutgers Center for Public Opinion Research	74.49	13.36	1	9
Siena Research Institute	56.28	27.83	13	283
Washington Post/University of Maryland	56.88	17.37	1	48
Total	59.89	29.61	70	1,817

NOTE.—Language difficulty scores calculated on Flesch's (1948) Reading Ease scale using the *koRpus* (Michalke 2016) package in the *R* statistical program. Higher scores suggest an easier metacognitive experience, per study 1.

survey responses, 3.47 percent were reported as *don't know*. Although this rate is lower than study 1 (4.20 percent), this difference is likely due to methodological differences across firms.

RESULTS

To test H1, we estimated a logistic mixed-effects model regressing *don't know* responses on standardized FRE scores, clustering standard errors around a participant identifier to account for subject-level variance in the propensity to report *don't know*. Consistent with H1, results indicate a significant negative relationship between reading ease and the log-likelihood of responding *don't know*, $B = -0.04$, $SE = 0.01$, $Z = -7.45$, $p < 0.001$, OR = 0.96, 95 percent CI = [0.95, 0.97] (figure 2).

DISCUSSION

Using publicly available data, results from study 2 replicated the finding that question language difficulty, operationalized through FRE scores, increases the probability of a *don't know* response. Study 2's successful replication of experimental results supports the notion that language difficulty affects polling data in undesirable ways. As such, it is important to consider the distribution of FRE scores across an even broader range of opinion questions. This is accomplished in study 3 through a content analysis of 10 polling firms' instruments. By understanding the degree to which language difficulty varies, and the systematic correlates of such variance, practical solutions for survey designers can be proffered.

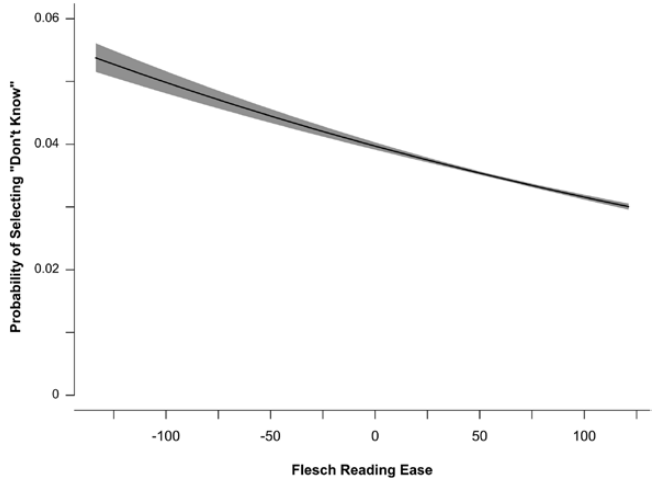

Figure 2. Probability of selecting *don't know* by FRE—Roper. Unstandardized for interpretability. Shaded area indicates the 95 percent confidence interval of the estimate.

Study 3

This content analysis follows a rich history of research focused on the effects of question language on survey response (e.g., Cantril 1944; Payne 1951; Pan and Kosicki 1993; Schwarz 1999; Druckman 2001; Graesser et al. 2006; Holbrook, Cho, and Johnson 2006; Krosnick, Malhotra, and Mittal 2014). Although work in this vein focuses on how variations in question wording affect the opinions reported, this analysis takes a different approach by focusing on the variability that exists across the questions themselves. This approach addresses a broader issue in the field of public opinion; namely, do systematic differences exist in the way polling organizations construct their questions? Given that it has been illustrated how language difficulty affects responses, study 3 takes a meta-approach to this issue by examining the nature of the questions themselves. To our knowledge, an investigation into the substantive differences between opinion questions *across firms* has not been conducted.

In particular, this analysis investigates whether "house effects" (e.g., Smith 1978) exist regarding the difficulty of language used in survey instruments by examining whether polling firms systematically differ in the difficulty of the language they employ. In addition to house effects, there may also be factors related to the aims or scope of a poll that affect question construction. For example, polls aimed at registered voters may employ more difficult language than polls targeting the general population due to presumptions of knowledge. Relatedly, language differences may exist in polls interested in community samples versus national samples. Language difficulty may also be differentially associated with different topics, as the language that survey writers use to explain an idea may vary. Investigating these matters is critical from a measurement standpoint because if systematic differences exist, but go unaccounted for, they present a threat to the validity of these instruments in the ways demonstrated by studies 1 and 2. In the same vein as other research that examines exogenous factors that lessen measurement quality, this analysis explores the degree to which metacognitive experiences may be affected by survey construction. By identifying patterns of language difficulty, future work can attend to these concerns during the survey construction process by having a more informed sense of how an FRE score might affect data.

RESEARCH QUESTIONS

This analysis is guided by one hypothesis and four research questions: H1 predicts that language ease will be inversely related to *don't know* responding. The four research questions include: (RQ1) How does public opinion poll language difficulty compare to the experimental conditions from study 1? (RQ2) Does language difficulty vary systematically between firms? (RQ3) Does language difficulty vary alongside factors related to the sample of the survey? and (RQ4) Does language difficulty vary based on topic?

Method

PROCEDURE

Instruments were collected from the 10 highest-rated polling firms according to FiveThirtyEight's ratings (Silver 2016). These ratings were used to ensure that the firms included were of high quality in the lead-up to the 2016 election, and provide a nice contrast in relation to the firms included in study 2. In addition to rating, firms were included only if full instruments were available for every survey conducted in 2016. With these criteria, the following firms were included (in order of rating): Monmouth University Polling Institute (MU 2017), SurveyUSA (SUSA 2017), Siena Research Institute (SRI n.d.), Muhlenberg College Polling Institute (MC n.d.), Fairleigh Dickinson University (FDU 2016), the Public Policy Institute of California (PPIC 2017), Marquette University Law School (MUL 2017), Quinnipiac University (QU n.d.), Public Policy Polling (PPP 2016), and Suffolk University Political Research Center (SU; Suffolk University and *USA Today* n.d.). The minimum rating among these polls was a "B+" (tables 3 and S4).

A total of 455 survey instruments were collected, composed of 8,090 questions. To prepare data for analysis, interviewer instructions occurring in parentheses or brackets (e.g., "[rotate choices]" or "{pronounced: 'kay-sick'}") were removed, as these are not read to participants.

MEASURES

Language difficulty was measured using the FRE scale (Flesch 1948) calculated using the *koRpus* (Michalke 2016) package in *R*. Overall, the polling questions included "fairly difficult" (Flesch 1948) language (M = 58.30, SD = 23.75; table 3), though there was considerable variability.

Unlike in study 2, *don't know/no opinion* was measured using each question's cumulative response as the unit of analysis because only topline results were available for these instruments. Therefore, a question's *don't know* response was measured as the percentage of respondents who reported *don't know*. Within this sampling frame, only some of the questions distinguished nonresponse from *don't know* responses and could be included in this analysis (n = 4,881 questions from eight firms, table 3). On average, the percentage of respondents reporting *don't know* was higher than the rates from studies 1 and 2 (M = 12.30 percent), though there was much greater variability in this sample (SD = 12.34).

Three *sample characteristics* were coded: voter likelihood, region, and sample size. For *voter likelihood*, three levels were coded: "general population" (n = 880), "registered voters" (n = 6159), and "likely voters" (n = 1,052). Although polling firms differed in their classifications of "likely voters" versus "registered voters," this variable was coded in accordance with

Table 3. Descriptive statistics of polling firms included in study 3

Polling firm	M_{FRE}	SD_{FRE}	538 score*	Number of instruments	Number of questions	Number of valid questions	$M_{DK\ Rate}$	$SD_{DK\ Rate}$
Fairleigh Dickinson University	59.99	22.70	A	36	236	144	6.96%	7.43%
Marquette University Law School	57.21	19.21	A	10	671	450	4.78%	4.36%
Monmouth University Polling Institute	53.79	17.51	A+	76	892	597	6.16%	5.43%
Muhlenberg College Polling Institute	50.52	24.24	A	2	61	35	11.63%	9.21%
Public Policy Institute of California	51.34	17.92	A	8	375	277	6.91%	5.73%
Public Policy Polling	58.97	24.74	B+	144	3,124	2,188	18.23%	14.65%
Quinnipiac University	64.38	17.98	A-	13	566	0	–	–
Siena Research Institute	61.39	25.55	A	39	790	0	–	–
Suffolk University Political Research Center	46.77	28.73	B+	6	231	206	9.75%	6.75%
SurveyUSA	59.40	27.81	A	121	1,144	984	9.14%	8.48%
Total	58.30	23.75	–	455	8,090	4,881	12.30%	12.34%

NOTE.—Language difficulty scores calculated on Flesch's (1948) Reading Ease scale using the *koRpus* (Michalke 2016) package in the *R* statistical program. Higher scores suggest an easier metacognitive experience, per study 1. *See Silver (2016) for poll ratings; these grades are based on a projection of the pollster's accuracy in future elections relative to other polling firms.

the phrasing from the survey instruments to avoid researcher bias. *Region* was coded into three categories: national (n = 1,552), state/regional (e.g., Iowa or the Midwest; n = 5,797), and community samples (e.g., Chicago, or the New York 26th Congressional District; n = 742). *Sample size* was coded as the number of participants who met study criteria (M = 982.94, SD = 475.39).

Question topic was measured using a structural topic model (STM). An STM is a method of unsupervised machine learning that parses textual data to categorize documents based on words that are *frequent* within documents of a classification, yet are relatively *exclusive* to that set of documents (FREX). These FREX words are used to interpret which topics are covered by documents in each of the categories specified by the model. The model was created using the *stm* (Roberts, Stewart, and Tingley 2017) package in *R*. To eliminate extraneous collinearity among topic classifications, survey questions underwent two additional cleaning procedures: stemming, which removes suffixes to reduce words to their referent concept, and the removal of stop words (e.g., "it" or "of"). A k-search diagnostic procedure was run to determine the number of topics to model. Diagnostic statistics converged at k = 7 topics.[5]

FREX words for each of the seven topic categories are illustrated in figure 3. We interpreted these categories as follows: Topic (1), "Economic Issues" (n = 516), are words that concern issues pertaining to minimum wage and state-level policies (e.g., casinos); Topic (2), "Issue Involvement" (n = 1,569), contains questions about the salience of the issues they refer to; Topic (3), "Referenda" (n = 908), are words pertaining to questions regarding initiatives on local ballots, such as "cigarett" and "nfl"; Topic (4), "Campaign Communication" (n = 573), includes words from questions pertaining to communication from candidates' campaigns; Topic (5), "Supreme Court" (n = 701), includes words like "scalia" and "confirm"; Topic (6), "Runners Up" (n = 2,406), includes the names of candidates who participated in the primaries; and Topic (7), "Attitudes" (n = 1,418), contains questions about the valence of opinions like "favor" and "unfavor."

Results[6]

We estimated a linear model regressing the percentage of respondents reporting *don't know* on FRE scores. Consistent with H1, results indicate a negative relationship between reading ease and the proportion of respondents reporting *don't know*, F (1, 4879) = 26.15, $p < 0.001$, $R^2_{adj} = 0.01$; Ease: B = –0.04, SE = 0.01, t = –5.11, $p < 0.001$. Though this coefficient may seem small, 0.04 percent fewer respondents reporting *don't know* per unit increase in FRE can be quite impactful over the broad range of FRE observed in the dataset (see figure 4).

5. Additionally, we created models of k = 6 and k = 8, neither of which significantly improved interpretability of the FREX words. See figure S6 for diagnostic information.

6. Graphical representations of results from each of the analyses can be found in the supplementary material online.

Topic 1: Economic Issues	Topic 2: Issue Involvement	Topic 3: Referenda	Topic 4: Campaign Communication	Topic 5: Supreme Court	Topic 6: Runners-Up	Topic 7: Attitudes
minimum	concern	cigarett	contact	confirm	libertarian	unfavor
wage	press	revenu	close	scalia	cruz	neutral
direct	serious	gender	martin	suprem	kasich	opinion
deleg	import	transport	campaign	perform	rubio	havent
hour	old	graduat	commerci	refus	marco	favor
wrong	threat	nfl	tell	session	sander	impress
casino	factor	school	bother	court	gari	regardless

Figure 3. Top seven FREX words by topic. STM was estimated using $k = 7$ topics of all survey questions ($N = 8{,}090$).

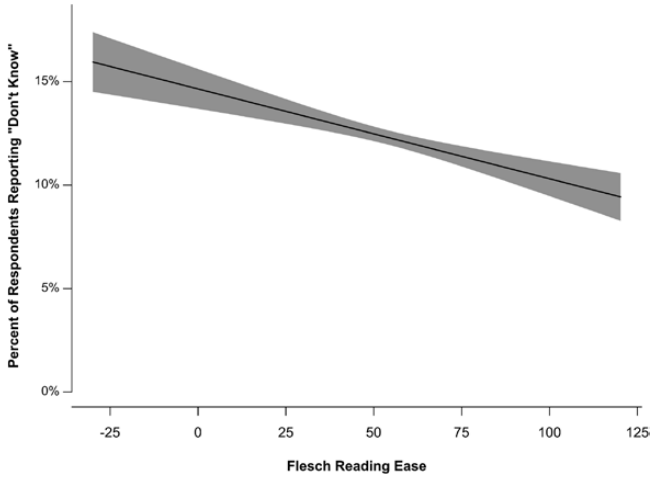

Figure 4. Percent of respondents selecting *don't know* by FRE—Five Thirty Eight. Shaded area indicates the 95 percent confidence interval of the estimate.

In order to contextualize the language difficulty of the experimental conditions in study 1, we compared the language in our experimental questions with actual polling questions. We found that the mean of the polling firm questions ($M = 58.30$, $SD = 23.75$) was closer to the questions in the easy condition ($M = 59.51$, $SD = 11.87$) compared to the difficult condition ($M = 20.97$, $SD = 16.26$). The range and standard deviations, however, suggest that polling-firm language was much more variable than the experimental language. This implies that the experimental conditions were representative of actual polling questions given that the range of

the experimental conditions (easy: 28.50 to 88.83; difficult: –24.85 to 50.61) fell within the bounds of the values observed in public opinion polls (–68.96 to 121.22). Additionally, an examination of the distributions of language difficulty from the experiment and the public opinion questions (figure S1) illustrates a flatter distribution in the polling firm sample, indicating more variability than the experimental manipulation. Taken together, these comparisons imply that experimental results may underestimate the influence of metacognitive experiences on polling data, given that the range was comparatively restricted in the experimental manipulation. These trends suggest that metacognition is likely posing a threat to the quality of survey data that has, thus far, gone unmeasured.

POLLING FIRM

To assess whether systematic differences in language difficulty exist between polling firms, a regression model was estimated excluding the Fairleigh Dickinson University (FDU) poll for reference. The omnibus model indicated significant differences *between* polling firms, F (9, 8081) = 20.72, $p < 0.001$, $R^2_{adj} = 0.02$ (figure S2). Compared to FDU, MU (B = –6.19, SE = 1.72, t = –3.60, $p < 0.001$), MC (B = –9.45, SE = 3.38, t = –2.80, $p < 0.01$), PPIC (B = –8.64, SE = 1.95, t = –4.42, $p < 0.001$), and SU (B = –13.20, SE = 2.18, t = –6.06, $p < 0.001$) contained significantly more difficult language. On the other hand, QU used significantly easier language (B = 4.41, SE = 1.82, t = 2.42, $p < 0.05$). The remaining firms (MUL, PPP, SRI, and SUSA) did not statistically differ from FDU. These results indicate the presence of a "house effect." Given this effect, all further analyses include polling firm as a control variable.

SAMPLE CHARACTERISTICS

To assess whether FRE scores vary based on sample characteristics such as voter likelihood and region, two logistic regression models were estimated.[7] For the voter likelihood model, F (11, 8079) = 27.52, $p < 0.001$, $R^2_{adj} = 0.04$ (figure S3), questions targeting registered voters included significantly more difficult language than questions targeting likely voters (B = –2.17, SE = 1.02, t = 2.13, $p < 0.05$). Moreover, questions targeting likely voters were significantly more difficult than questions targeting the general population (B = –11.41, SE = 1.50, t = –7.59, $p < 0.001$). These results suggest that polling questions appear to be constructed differently depending on the anticipated voting behaviors of the target sample.

For the regional model, F (11, 8079) = 26.79, $p < 0.001$, $R^2_{adj} = 0.03$ (figure S4),[8] results suggest that questions crafted for state/regional samples

7. The first model was estimated with the likely voters and the polling firm FDU as the referent groups, whereas the second model included general population voters and FDU.

8. The first model was estimated with national sample and the polling firm FDU as the referent categories, whereas the second model referents were the community sample and FDU.

included significantly more difficult language than those crafted for national samples ($B = –2.46$, $SE = 0.94$, $t = –2.63$, $p < 0.001$). Moreover, questions targeting national samples included significantly more difficult language than community samples ($B = –8.55$, $SE = 1.44$, $t = –5.95$, $p < 0.001$). Thus, it appears that language difficulty varies as a function of the region being targeted.

The final characteristic compared was sample size. The overall regression model was significant, $F(10, 8080) = 19.22$, $p < 0.001$, $R^2_{adj} = 0.02$ (figure S5). Sample size was a significant predictor in the model, $B = –0.002$, $SE = 0.001$, $t = –2.37$, $p < 0.05$, indicating that as sample size increased, question language became more difficult.[9]

TOPIC

Overall, the logistic regression model indicated significant differences by topic category, $F(15, 8075) = 140.40$, $p < 0.001$, $R^2_{adj} = 0.21$ (figure S7).[10] Based on this model, the topics that employed the most difficult language included Referenda ($B = –7.58$, $SE = 1.18$, $t = –6.42$, $p < 0.001$), Supreme Court ($B = –9.15$; $SE = 1.24$, $t = –7.36$, $p < 0.001$), Runners Up ($B = –15.23$, $SE = 1.04$, $t = –14.67$, $p < 0.001$), and Attitudes ($B = –22.01$, $SE = 1.11$, $t = –19.85$, $p < 0.001$). Each of these topics utilized significantly more difficult language than questions surrounding the Economy. Campaign Communication questions were found to be only marginally (one-tail) more difficult than Economic questions ($B = –2.17$, $SE = 1.29$, $t = –1.67$, $p = 0.09$), whereas questions pertaining to Issue Involvement were significantly easier ($B = 8.27$, $SE = 1.09$, $t = 7.60$, $p < 0.001$). These findings indicate systematic variability across political topics.

Discussion

The purpose of study 3 was to examine the extent to which language difficulty varied across surveys administered during 2016. The results indicate that language difficulty both within and across polls varied considerably more than the experimental conditions in study 1. Additionally, findings from this study further support the claim that language difficulty contributes to *don't know* responses. In our effort to better contextualize how FRE values vary within the polling landscape, this content analysis uncovered several systematic differences in language difficulty scores. Specifically, factors such as polling firm, voter likelihood, region, sample size, and question topic were associated with significant differences in language difficulty. Importantly, these effects

9. Sample sizes range from 300 to 2,268.

10. In this logistic regression model, topic 1 ("Economic Issues") and FDU served as the referent categories.

persisted even when house effects were controlled. These findings imply that, whether knowingly or not, the polling industry utilizes normative strategies when crafting questionnaires to suit their population of interest, as well as to ascertain opinions on specific topics.

Importantly, these findings underscore the need for consistency in language difficulty across surveys if drawing comparisons between these polls is desirable. Although the effect sizes across these analyses (with the exception of question topic) are relatively small, effects can add up and hinder the predictive utility of polling. The purpose of these investigations is to argue that by understanding the implications of language difficulty, this source of measurement error can be prevented. We hope that the distribution of FRE scores across firms, topics, populations, and scope presented here helps in this regard.

General Discussion

The contributions of these studies are both practical and theoretical. Practically, if the purpose of polling is to provide a precise characterization of public opinion, then polling firms and researchers must consider language difficulty when drafting survey instruments to combat measurement errors in self-reports. Although the easing of language and corresponding metacognitive experiences affords greater measurement validity, so too would efforts to decrease variability of language difficulty by "reining in" outliers. The computational tools employed here offer empirical insight as to which questions might be problematic by providing guidance into how FRE scores vary and how this variance affects critical outcomes such as *don't know* responding. To combat measurement issues related to language difficulty and metacognition, we offer two recommendations. The first is to consider language difficulty in the early stages of survey construction by comparing FRE scores to the distributions we provide here. This practice should not only be conducted with the intention of making language as easy as possible, but should also be implemented to improve consistency between questions on the same survey. If consistency is impossible, we suggest that polling firms include the metacognitive measures utilized here to empirically account for language differences through statistical control. Although reading ease measures offer an objective measure of ease, our metacognition items offer a subjective measure, which would help discern when questions are perhaps too difficult for respondents. Thus, researchers need not abandon strategic practices for a given sample, but should be able to anticipate or measure how survey language differentially affects the metacognitive experiences of respondents.

The second contribution is the ability to theoretically link language difficulty to the cognitive process of metacognition, and to demonstrate how this process affects theoretically and practically relevant outcomes. Specifically, our findings corroborated FIT postulates (Schwarz 2011) regarding the influence

of metacognitive experiences on political interest and efficacy (study 1), the number of *don't know* responses (studies 1, 2, and 3), and the reporting of ideologically (in)consistent attitudes (study 1). As survey questions and the associated metacognitive experiences became more difficult, the measures and subsequent analyses became less precise. This relationship is inherently problematic for the polling industry, given that wide variations in language difficulty exist (studies 2 and 3), and that this variability can be traced back to strategic decisions, not random error, on the part of polling firms (study 3). Given that polling firms aim to reduce *don't know* responses and improve attitude reliability, the findings obtained here should be interpreted as problematic yet preventable.

Despite advancements, there were also limitations. Regrettably, not all of the relationships from the experiment could be replicated in studies 2 and 3 because individual-level datasets were not consistently available. Thus, we were forced to sacrifice our ability to replicate findings from study 1 to prioritize a more ecological sampling frame. We did, however, replicate the relationship between FRE and *don't know* responses across all samples, providing empirical support that language difficulty does consistently, and problematically, affect this dimension of survey response. Additionally, although findings across all three studies suggest that language difficulty impairs metacognitive experiences, it is possible that other cognitive processes are also affected by this difficulty. To evidence this possibility, in our experiment we unexpectedly found statistically significant differences in reported attitudes, t (1014) = 2.75, $p < 0.01$, $d = 0.17$, based on assignment to language condition. Specifically, those in the easy condition reported more liberal-leaning values ($M = 3.87$, $SD = 0.63$) than those in the difficult condition ($M = 3.99$, $SD = 0.68$), despite no significant differences in self-reported ideology by condition, t (1010) = 0.71, $p = 0.48$. This finding underscores the importance of including metacognitive experience measures on surveys. Doing so allows researchers to parse out metacognitive effects from other cognitive processes likely affected by survey language.

In conclusion, the purpose of this research was to theoretically apply the cognitive process of metacognitive experiences to public opinion research (Schwarz 2010, 2011). Our results suggest that utilizing easier language eases political information processing. With guidance from the FRE measure, we hope that survey designers become more aware of their linguistic decisions and that this awareness functions to make survey language even easier and more consistent. If this is the case, then practice can lead to even more desirable outcomes than those obtained here. By expanding our understanding of how people respond to opinion questions and by identifying factors that affect data in unaccounted-for ways, better surveys can be produced and better data can be obtained.

Supplementary Data

Supplementary data are freely available at *Public Opinion Quarterly* online.

References

Alter, Adam L., and Daniel M. Oppenheimer. 2009. "Uniting the Tribes of Fluency to Form a Metacognitive Nation." *Personality and Social Psychology Review* 13:219–35.

Bishop, George F., Robert W. Oldendick, and Alfred J. Tuchfarber. 1984. "Interest in Political Campaigns: The Influence of Question Order and Electoral Context." *Political Behavior* 6:159–69.

Cantril, Hadley. 1944. *Gauging Public Opinion*. Princeton, NJ: Princeton University Press.

Clifford, Scott, and Jennifer Jerit. 2016. "Cheating on Political Knowledge Questions in Online Surveys: An Assessment of the Problem and Solutions." *Public Opinion Quarterly* 80:858–87.

Delli Carpini, Michael X., and Scott Keeter. 1996. *What Americans Know About Politics and Why It Matters*. New Haven, CT: Yale University Press.

Druckman, James N. 2001. "The Implications of Framing Effects for Citizen Competence." *Political Behavior* 23:225–56.

Fairleigh Dickinson University. 2016. *PublicMind Research–Calendar Year 2016*. Accessed March 22, 2017. Available at https://view2.fdu.edu/publicmind/2016/.

Flesch, Rudolph. 1948. "A New Readability Yardstick." *Journal of Applied Psychology* 32:221–33.

Graesser, Arthur C., Zhiqiang Cai, Max M. Louwerse, and Frances Daniel. 2006. "Question Understanding Aid (QUAID): A Web Facility that Tests Question Comprehensibility." *Public Opinion Quarterly* 70:3–22.

Hayes, Andrew F. 2016. *PROCESS* (version 2.16). Computer software. Accessed March 12, 2017. Available at http://www.processmacro.org/download.html.

Holbrook, Allyson, Young I. K. Cho, and Timothy Johnson. 2006. "The Impact of Question and Respondent Characteristics on Comprehension and Mapping Difficulties." *Public Opinion Quarterly* 70:565–95.

Krosnick, Jon A., Neil Malhotra, and Urja Mittal. 2014. "Public Misunderstanding of Political Facts: How Question Wording Affected Estimates of Partisan Differences in Birtherism." *Public Opinion Quarterly* 78:147–65.

Marquette University Law School. 2017. *Results and Data*. Accessed March 22, 2017. Available at https://www.law.marquette.edu/poll/category/results-and-data/.

Michalke, Meik. 2016. *koRpus*: An R Package for Text Analysis (version 0.06–5). Computer software. Accessed March 25, 2017. Available at https://cran.r-project.org/web/packages/koRpus/.

Mondak, Jeffrey J. 2001. "Developing Valid Knowledge Scales." *American Journal of Political Science* 45:224–38.

Monmouth University Polling Institute. 2017. *Monmouth University Poll Reports*. Accessed March 22, 2017. Available at https://www.monmouth.edu/polling-institute/reports/.

Muhlenberg College Polling Institute. n.d. *Surveys*. Accessed March 22, 2017. Available at http://www.muhlenberg.edu/main/aboutus/polling/surveys/.

Niemi, Richard G., Stephen C. Craig, and Franco Mattei. 1991. "Measuring Internal Political Efficacy in the 1988 National Election Study." *American Political Science Review* 85:1407–13.

Oppenheimer, Daniel M. 2006. "Consequences of Erudite Vernacular Utilized Irrespective of Necessity: Problems with Using Long Words Needlessly." *Applied Cognitive Psychology* 20:139–56.

Pan, Zhongdang, and Gerald M. Kosicki. 1993. "Framing Analysis: An Approach to News Discourse." *Political Communication* 10:55–75.

Payne, Stanley L. B. 1951. *The Art of Asking Questions: Studies in Public Opinion*, vol. 3. Princeton, NJ: Princeton University Press.

Petty, Richard E., Pablo Briñol, Zakary L. Tormala, and Duane T. Wegener. 2007. "The Role of Meta-Cognition in Social Judgment." In *Social Psychology: Handbook of Basic Principles*, 2nd ed., edited by Arie W. Kruglanski and E. Tory Higgins, pp. 254–84. New York: Guilford Press.

Presser, Stanley, Mick P. Couper, Judith T. Lessler, Elizabeth Martin, Jean Martin, Jennifer M. Rothgeb, and Eleanor Singer. 2004. "Methods for Testing and Evaluating Survey Questions." *Public Opinion Quarterly* 68:109–30.

Prior, Markus. 2010. "You've Either Got it or You Don't? The Stability of Political Interest over the Life Cycle." *Journal of Politics* 72:747–66.

Public Policy Institute of California. 2017. *Statewide Surveys*. Accessed March 22, 2017. Available at http://www.ppic.org/main/allpubs.asp?sort=type#a12.

Public Policy Polling. 2016. *2016 Archive*. Accessed March 22, 2017. Available at https://www.publicpolicypolling.com/polls-archive/archive-2016/.

Quinnipiac University. n.d. *Recent Releases*. Accessed March 22, 2017. Available at https://poll.qu.edu/.

Roberts, Molly, Brandon Stewart, and Dustin Tingley. 2017. *stm: R Package for Structural Topic Models* (version 1.2.2). Computer software. Accessed March 25, 2017. Available at https://cran.r-project.org/web/packages/stm/.

Roper Center for Public Opinion Research. n.d. *Dataset Catalog*. Accessed February 16, 2018. Available at https://www.ropercenter.cornell.edu/CFIDE/cf/action/catalog/.

Schuman, Howard, and Stanley Presser. 1996. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. Thousand Oaks, CA: Sage.

Schwarz, Norbert. 1998. "Accessible Content and Accessibility Experiences: The Interplay of Declarative and Experiential Information in Judgment." *Personality and Social Psychology Review* 2:87–99.

———. 1999. "Self-Reports: How the Questions Shape the Answers." *American Psychologist* 54:93–105.

———. 2006. "Feelings, Fit, and Funny Effects: A Situated Cognitive Perspective." *Journal of Marketing Research* 43:20–23.

———. 2010. "Meaning in Context: Metacognitive Experiences." In *The Mind in Context*, edited by Batja Mesquita, Lisa F. Barrett, and Eliot R. Smith, pp. 105–25. New York: Guilford.

———. 2011. "Feelings-as-Information Theory." In *Handbook of Theories of Social Psychology*, edited by Paul A. M. Van Lange, Arie W. Kruglanski, and E. Tory Higgins, pp. 289–308. Thousand Oaks, CA: Sage.

Shulman, Hillary C., and Franklin J. Boster. 2014. "Effect of Test-Taking Venue and Response Format on Political Knowledge Tests." *Communication Methods and Measures* 8:177–89.

Shulman, Hillary C., and Matthew D. Sweitzer. 2018a. "Advancing Framing Theory: Designing an Equivalency Frame to Improve Political Information Processing." *Human Communication Research* 44:1–21.

———. 2018b. "Varying Metacognition Through Public Opinion Questions: How Language Can Affect Political Engagement." *Journal of Language and Social Psychology* 37:224–37.

Siena College. n.d. *Political–Siena Research Institute News Archive*. Accessed March 22, 2017. Available at https://www2.siena.edu/centers-institutes/siena-research-institute/political-polls/.

Silver, Nate. 2016. *FiveThirtyEight's Pollster Ratings*. Accessed March 21, 2017. Available at https://projects.fivethirtyeight.com/pollster-ratings/.

Smith, Tom W. 1978. "In Search of House Effects: A Comparison of Responses to Various Questions by Different Survey Organizations." *Public Opinion Quarterly* 42:443–63.

Suffolk University and *USA Today*. n.d. *Polling Archives*. Accessed March 22, 2017. Available at http://www.suffolk.edu/academics/1094.php.

SurveyUSA. 2017. *SurveyUSA Poll Results*. Accessed March 22, 2017. Available at http://www.surveyusa.com/PollHistory.aspx.

Tourangeau, Roger, Aaron Maitland, and H. Yanna Yan. 2016. "Assessing the Scientific Knowledge of the General Public: The Effects of Question Format and Encouraging or Discouraging Don't Know Responses." *Public Opinion Quarterly* 80:741–60.

Tourangeau, Roger, and Kenneth A. Rasinski. 1988. "Cognitive Processes Underlying Context Effects in Attitude Measurement." *Psychological Bulletin* 103:299–314.

Winkielman, Piotr, Norbert Schwarz, Tedra A. Fazendeiro, and Rolf Reber. 2003. "The Hedonic Marking of Processing Fluency: Implications for Evaluative Judgment." In *The Psychology of Evaluation: Affective Processes in Cognition and Emotion*, edited by Jochen Musch and Karl C. Klauer, pp. 189–217. Mahwah, NJ: Lawrence Erlbaum Associates.

Zaller, John. 1992. *The Nature and Origins of Mass Opinion*. New York: Cambridge University Press.

Zaller, John, and Stanley Feldman. 1992. "A Simple Theory of the Survey Response: Answering Questions versus Revealing Preferences." *American Journal of Political Science* 36:579–616.