# Measuring Political Knowledge and Not Search Proficiency in Online Surveys

Matthew DeBell

Institute for Research in the Social Sciences, Stanford University

30 Alta Road, Stanford, CA 94305, USA

Tel. 703-598-5983

debell@stanford.edu

## Author Note

Matthew DeBell is Director of Stanford Operations for the American National Election Studies and Senior Research Scholar at the Institute for Research in the Social Sciences at Stanford University.

## Funding

Data collection for Studies 1 and 2 was funded by the John E. Fetzer Memorial Trust (Jon A. Krosnick, Principal Investigator). Data collection for Studies 3 and 5 was funded by a grant from the National Science Foundation to Stanford University (grant no. SES-1835022; Shanto Iyengar, Principal Investigator), and Study 5 was also funded by a gift to Stanford University from Facebook. Data collection for Study 4 was funded by a grant from the National Science Foundation to the University of Michigan (grant no. SES-1835971; Ted Brader, Principal Investigator).

## Acknowledgments

The author thanks the principal investigators for making the studies possible, Jennifer Jerit for comments, and Matthew K. Berent for managing data collection for studies 1 and 2. Findings and interpretations are the author's alone.

**Abstract**

Most online survey questions testing political knowledge are susceptible to measurement error when participants look up the answers. This paper reports five studies of methods to detect and prevent this common source of error. To detect lookups, "catch questions" are more reliable than self-reports, because many participants lie rather than admit looking up answers. Strongly worded instructions reduced lookups by about two-thirds, while the triple combination of instructions, requesting a promise not to look up answers, and adaptive feedback (asking participants who look up an answer to stop doing so) reduced the percentage of respondents looking up an answer by a further half, to 3%. For office recall knowledge items, photo-based open-ended questions eliminated lookups and had similar validity to traditional text-based versions, making them a good choice when a visual format is viable.

*Keywords*: political knowledge, web surveys, self-reported cheating, visual knowledge tests

# MEASURING POLITICAL KNOWLEDGE AND NOT SEARCH PROFICIENCY IN ONLINE SURVEYS

Quite a few people look up the answers when online surveys test their political knowledge. Such lookups invalidate the knowledge data for these survey participants and reduce the comparability of survey results between the internet mode and others, such as face-to-face interviews, where far fewer participants look up answers. To improve measurement validity and reduce mode differences, researchers interested in political knowledge need methods to avert lookups. This paper reports the results of five studies that have tested new methods to measure political knowledge online while preventing the measures from being confounded by participants who look up the answers. The contribution of these studies has been to test whether self-reports of lookups are accurate by using "catch questions," to test commitment and adaptive feedback mechanisms to discourage respondents from looking up answers, to pair visual political knowledge tests with a catch question to measure lookups, and to compare the validity of open-ended visual questions to conventional text-based questions. Methods to effectively prevent lookups are identified.

## Background

Political knowledge has been a widely used construct in public opinion and political behavior literature for decades (e.g., Converse, 1964; Delli Carpini & Keeter 1996; Gilens, 2001; Lupia, 2016). With major studies including the American National Election Studies, the Pew Research Center's American Trends Panel, and the Cooperative Election Study relying on self-administered Internet questionnaires, where there is no interviewer to police the process, a new threat to the validity of political knowledge measurement has emerged: respondents can easily look up the answers instead of basing their responses on their prior knowledge. Researchers

sometimes call this "cheating" (e.g. Clifford & Jerit 2016; Jensen & Thomsen, 2014; Vezzoni & Ladini, 2017), particularly when participants have been asked not to look up the answers (though lookups usually do not violate any established norm or promise).

When respondents look up answers and researchers interpret correct answers as indicating prior knowledge, it is obvious that this artificially inflates knowledge scores and reduces the validity of knowledge measures (Clifford & Jerit, 2014; 2016; Jensen & Thomsen, 2014; Smith et al. 2020), even though the ability to look up political information is also a valuable civic skill (Kleinberg & Lau, 2019). If respondents look up the answers it also invalidates assumptions that researchers make about the response process, which could call their conclusions into question. For example, Prior and Lupia (2008) designed an innovative study in which respondents to an online survey were paid $1 for each correct answer to a political knowledge question. Compared to a control group that was not paid, respondents who were offered money and were time-limited to answer within one minute answered more questions correctly. This was evidence that increasing respondents' motivation to think carefully about a survey question improved the quality of measurement—unless, of course, the money motivated people to look up the answers instead of to think. Current evidence about respondent behavior indicates we need to take this possibility more seriously.

**Extent of the problem**

Several recent studies have discovered that it is common for online survey participants to look up the answers to political knowledge questions. The frequency varies across samples and detection methods. Although lookups have appeared rare among in-person-intercept-sample participants subject to monitoring (Gooch & Vavreck, 2019) and among some paid workers on Amazon Mechanical Turk (MTurk) (Berinsky et al., 2012; but *cf*. Motta et al., 2017), the balance

of evidence points strongly to the conclusion that many respondents perform lookups in many online studies.

One approach to identifying lookups has been indirect, comparing the results from in-person interviews (where the interviewers presumably prevent lookups) to results from self-administered studies. Such studies have found that knowledge appears much higher in online modes than interviewer-administered modes. In the 2016 American National Election Study (ANES), 23% of face-to-face respondents correctly answered that a Senator is elected for 6 years, compared to 45% online (Guggenheim et al., 2019), even though both modes used probability samples with comparable sample frames and relatively high response rates. Differences in nonresponse patterns may confound such comparisons, so differences are not necessarily due to lookups. Experimental study designs can help reduce such confounds. Fricker et al. (2005) and Clifford and Jerit (2014) randomized between online and interviewer-administered modes and found higher knowledge scores among online respondents.

Another approach to identifying lookups has been direct: ask respondents if they looked up the answers. In the Clifford and Jerit (2014) study, 11% of online respondents admitted to browsing the web while completing the survey, compared to 1% of in-person respondents. The authors' interpretation was that the difference was due to lookups, and this was supported by lower criterion validity for the knowledge items in the online condition. In an online survey of a national sample in Denmark, 22% of participants reported obtaining outside assistance to answer political knowledge questions (Jensen and Thomsen 2014). On MTurk, about 7% of one sample admitted to having looked up answers, and in some student samples, where participants may be more motivated to appear knowledgeable, self-reported lookups rates were similar or even higher: 24 to 41% (Clifford & Jerit, 2016).

The problem may be worse than the direct approach suggests because these studies' ability to indicate the lookup frequency are limited by their reliance on self-reports. Respondents may hesitate to admit to having looked up answers due to the social desirability of appearing knowledgeable and appearing to have complied with the researcher's expectations. Self-reports may therefore be biased toward an under-estimate of look-ups. Smith et al. (2020) reported similar results from self-reports and catch questions in a study of undergraduates. However, no prior study has compared self-reports to other, non-self-reported and potentially more accurate indicators of lookups in a general population sample, and no prior study has compared the results at the level of a single question, so the reliability of self-reports is largely unknown.

An alternative to reliance on self-reports is to use a new inferential measure of lookups called a "catch question" (Motta et al., 2017). A catch question is a question so difficult that anyone who answers correctly is inferred to have looked up the answer. In an MTurk study, 25% of respondents who were asked such questions evidently looked up answers, and on SurveyMonkey (an opt-in panel) 11% answered a catch question correctly (Motta et al., 2017). Another study found 25% of participants from an opt-in panel (YouGov) with demographic characteristics weighted to match the U.S. adult citizen population correctly answered one or two catch questions (DeBell, 2019).

**Theory and methods to reduce lookups**

Lookups may be particularly motivated by self-deceptive enhancement (SDE) (Clifford & Jerit, 2016; Schulman & Boster, 2014; Style & Jerit, 2021). In the survey context, SDE refers to socially desirable responses that serve the respondent's self-esteem, in contrast to socially desirable responses that promote a positive appraisal of the respondent by others (known as "impression management"). Some evidence suggests that the latter form of social desirability

may be slightly less prevalent than SDE in online surveys (Booth-Kewley et al., 2007). Instead, online respondents look up answers so they can feel capable and avoid uncomfortable feelings caused by being unable to answer correctly, as matters of intellect are particularly triggering for SDE (Paulhus & Oliver, 1998).

In addition to SDE, interest in the subject matter may also motivate some lookups; if respondents are curious to know the answer to a factual question, they may want to look it up. (One respondent commented, "I have to tell you, I really wanted to look up the answers to those last questions. To see if I knew the answers.") Political interest is thought to motivate effort to answer political knowledge questions correctly (Robison, 2015), and effort can manifest as searching for answers online. Interest motivating such effort may be instrumental, such as to answer a question correctly for SDE or to be able to apply knowledge outside the context of a questionnaire, or it may be intrinsic, as intellectual curiosity valuing knowledge for its own sake. Prior literature has not tested any methods to limit lookups based on an intellectual curiosity motive, but respondents who are more interested in the survey topic are more likely to look up answers (Gummer & Kunz, 2019). If interest is intrinsic, telling participants they will see the correct answers after giving their best guess should reduce their motivation to look up the answers.

Lookup propensity is mediated by valuing the traits that lookups tend to accentuate, such as being politically knowledgeable (Clifford & Jerit, 2016), and can be further mediated by contexts that make these motives more or less salient. Lookup propensity can be moderated by the difficulty or cost of looking up an answer and by countervailing social expectations such as following a researcher's instructions or the norm against cheating on tests.

The literature has tested four mechanisms to reduce lookups: instructions (e.g., Motta et al., 2017), requested commitments (e.g., Clifford & Jerit, 2015, 2016), time limits (e.g., Clifford & Jerit, 2016), and questions that make it more difficult to look up the answers (e.g., Munzert & Selb, 2015). Instructions and requested commitments not to look up answers rely on the norms of following instructions and keeping one's word to reduce lookup propensity. Time limits and questions designed to make looking up answers more labor-intensive rely on making the unwanted behavior more difficult.

Instructions to not look up answers reduced lookups (as detected with experiments using catch questions) from 25% to the 5% to 13% range on MTurk and from 11% to 8% on Survey Monkey (Motta et al., 2017). The instruction in these studies was, "Please answer each of the following questions. As you do so, we ask that you please *do not look up answers online*." A survey experiment with the YouGov opt-in sample tested the following instruction: "We are interested in how much information about certain subjects gets out to the public. No one knows all the answers to the next few questions. When you are not sure, please just give your best guess. Please do not look up the answers. We want to see what people already know or can guess." This reduced lookups on catch questions from 25% to 15% (DeBell, 2019).

Feedback provides instructions to respondents in response to their specific behaviors during a survey. It has been used in some contexts, such as respondents answering too quickly to have given any thought to their answer (Conrad et al., 2017), or answering incorrectly during a psychological test (e.g., Nosek et al., 2005). However, adaptive feedback has not previously been tested in the context of measuring political knowledge. People sometimes need to be told more than once before they follow directions, so it is no surprise that some survey participants do not carefully read or follow instructions (Oppenheimer et al., 2009). For such respondents, a single

instruction may be less likely to prevent lookups than an instruction repeated immediately after a lookup.

A commitment mechanism asks respondents to state that they will not look up the answers. For example (Clifford & Jerit, 2016), "It is important to us that you do NOT use outside sources like the Internet to search for the correct answer. Will you answer the following questions without help from outside sources?" This has been expected to be more effective than an instruction alone because people usually want to be consistent with their promises. A commitment mechanism does reduce *self-reported* lookups (Clifford & Jerit, 2015, 2016). However, as noted above, respondents may hesitate to admit to having looked up the answer, and the commitment mechanism could reduce their willingness to make this admission. Prior general-population studies have not compared self-reports to other measures of look-ups, nor have they evaluated the effectiveness of commitment mechanisms using methods other than a self-report.

All else being equal, elapsed time to answer a question may be longer if a participant looks up the answer than if he or she answers from prior knowledge. Some researchers have considered imputing "lookup" status when answers take much longer than usual. But this approach is unsatisfactory because researchers cannot know if respondents taking a long time are doing so because they are looking up the answer, pondering the question, or taking a break. Many studies have imposed time limits, such as 30 or 60 seconds (e.g., Bullock et al., 2015; Prior & Lupia, 2008; Strabac & Aalberg, 2011). One concern with this approach is that participants who perform lookups may do so very quickly, with a majority of lookups in one study occurring in 33 to 37 seconds (DeBell, 2019). Another concern is that time limits confound the measurement of knowledge with the psychological stress of being asked to perform a timed

task. Timers have only very rarely been tested, but when they were, they did not appear effective at reducing lookups (Clifford & Jerit, 2016).

Some questions make it easier to look up the answer than others. An open-ended question on the ANES was, "What job or political office is held by Angela Merkel?" A few keystrokes to a search engine revealed the answer. Other question formats make it considerably more difficult to look up the answer online. One approach is to use a photo of the subject person instead of stating her name. Although it is possible to perform image searches online, this is more cumbersome and not as widely used as text searches.

Picture-based tests of political knowledge have given somewhat similar results to verbal tests, but the two formats are not necessarily interchangeable (Munzert & Selb, 2015; Prior, 2014). Women, older people, the less educated, and people with a visual cognitive style do better on visual than verbal tests (Prior, 2014; Stiers & Hooghe, 2021). Questions that combine a visual and verbal prompt – for example, asking "What job or political office is held by Angela Merkel?" while showing her picture – may produce more valid knowledge measurement than the name alone, because many people consume news in visual formats such as television and recall information visually. Verbal-only measures put some people at a disadvantage, while visual-only measures put others at a disadvantage. Most previous studies of visual questions used multiple choice response options (e.g., Prior, 2014; Strabac & Alaberg, 2011), which still make it relatively easy to look up the answers, and Munzert and Selb's (2015) comparison of visual and verbal questions in a German sample did not provide a mechanism to clearly identify answer lookups. To thwart lookups, picture-based knowledge tests can be open-ended. No prior study has tested photo-based knowledge items with an open-ended response format.

**Hypotheses for improved methods**

Building on prior literature, this study tests several research questions for the first time: if a commitment mechanism is effective when tested using reliable catch questions; if self-reports of lookups are accurate when gauged against catch questions; if adaptive feedback is effective; if addressing respondents' curiosity reduces lookups; if picture-based knowledge tests with open-ended questions reduce lookups compared to text-based tests; and if open-ended picture-based tests have similar results and validity compared to text-based tests. Hypotheses—some exploratory—are as follows.

*H1, instruction effect: Lookups will be reduced by an instruction not to look up answers*. Prior studies have demonstrated the instruction effect (e.g. Motta et al., 2017). The test of the instruction effect is incidental to the test of the commitment effect, below.

*H2, commitment effect: An instruction coupled with a request that the participant promise to guess will reduce lookups more than an instruction alone*. Prior studies demonstrated the commitment effect using self-reports (Clifford & Jerit, 2015, 2016). We test it using more reliable catch questions.

*H3, accurate admission: Substantially all (e.g., >90%) participants who look up answers and were not asked not to do so will admit having done so*. Prior research has never tested this assumption (which amounts to treating false denials as ignorable), but it is implicit in studies that rely on self-reports to measure the frequency of lookups (e.g., Clifford & Jerit, 2016; Jensen & Thomsen, 2014). Establishing the accuracy of such reports would bolster such literature and validate the method for future studies. Testing the exploratory *H3* requires setting a quantitative threshold for accuracy, which requires a subjective judgment about how much false reporting is ignorable. I suggest 90% accuracy is reasonable: for instance, if the true lookup rate were 30%, and fewer than 90% of these were reported, the reported lookup rate would be less than 27%, and

the reporting error would be more than 3 points – enough to rival or exceed sampling error in many studies, and a concern. Conversely, accuracy above 90% might indeed make false denials ignorable.

*H4, feedback effect: Among participants who looked up answers, the rate of lookups on a subsequent question will be lower for those receiving feedback (asking them not to do that again) than for those who do not receive feedback.* The feedback condition reinforces the instruction by giving it a second time when needed. Although not hypothesized before data collection, we also examine the joint effects of the instruction, commitment, and feedback mechanisms.

*H5, curiosity motive*: *Fewer participants look up the answers to a knowledge question after they have been informed they will be told the answers following the quiz*. This is untested in prior literature but warranted by the finding that interest in the subject matter motivates some lookups.

*H6, pictures prevent lookups: Fewer participants look up the answers to an office recall question that describes the subject with a picture than look up the answer to a conventional text-based question*. We suppose that respondents are unlikely to use image searching tools to look up the answer to a question where the subject matter is described with a picture. Unlike prior research, we use open-ended questions so that multiple choice options do not provide readily searchable answer options.

*H7, pictures produce different knowledge gaps than text, by gender, age, and education*. Documenting any heterogeneity in the question format's effect on measured knowledge is important if photo questions are to be adopted. Prior (2014) found that education was more strongly tied to verbal than visual political knowledge, that women performed better relative to

men on visual political tests, and that older people performed better relative to younger people on visual tests. We expect to extend these findings to questions with the open-ended format.

*H8, picture item validity vis-à-vis interest in politics: When political knowledge is measured using pictures of people, the association between political knowledge and interest in politics will be equivalent or stronger than the association when knowledge is measured using traditional office recall items*. The association between political knowledge and interest in politics is an established finding in the political knowledge literature (e.g. Delli Carpini & Keeter, 1996), so this is a criterion by which the validity of picture-based knowledge questions can be assessed. (If lookups are motivated by interest in the subject matter (Gummer & Kunz, 2019) and lookups are easier using traditional questions, then this validation estimate may be upwardly biased for the traditional question, which would increase the chance of erroneously rejecting the hypothesis.)

*H9, picture item validity vis-à-vis liberal-conservative placement: A knowledge score based on an online visual office recall test has equivalent or superior validity to a knowledge score based on a text-based office recall test, where validity is defined as the association between the test score and the score from another set of political knowledge items based on liberal-conservative placements of parties and candidates*. The existence of strong associations among different political knowledge questions is a standard expectation based on prior literature (e.g. Delli Carpini & Keeter, 1996), so H9 addresses the convergent validity of picture-based knowledge questions.

*H10, picture item validity as voting moderator: A knowledge score based on a visual office recall test has equivalent or superior validity to a knowledge score based on a text-based office recall test, where validity is measured by the strength of the knowledge score as a*

*moderator of the effect of birthright citizenship policy preferences on voting for Donald Trump*. In a study with a local sample in Belgium, Stiers and Hooghe (2021) found support for the validity of visual knowledge tests by measuring an effect of visual knowledge on voting behavior that was similar to or larger than the effect for textual knowledge. A repeated finding in the literature is that knowledge moderates the relationship between policy preferences and candidate choice (e.g, DeBell, 2013; Goren, 1997), so the detection of this type of moderation is a criterion by which to assess the validity of picture-based knowledge questions.

## Methods & Data

Data come from five studies. Each tested a subset of hypotheses shown in Table 1. Study 1 is described in detail below. Studies 2 through 5 are described more briefly by reference to Study 1's instrumentation. The chronological order of the studies was 2, 3, 1, 4, 5; the studies are numbered as they are because Study 1 was designed first. It also was the most comprehensive and it is easier to understand the contents when Study 1 is described first. All data and questionnaires are publicly available from referenced sources. Statistical code and output are in online supplemental materials: https://osf.io/kzsta/.

[TABLE 1 HERE]

Study 1 (DeBell 2021a) used a non-probability sample of 1,556 qualifying Americans aged 18 or older who volunteered to complete surveys over the Internet in exchange for payments. The sample was provided by Dynata, which randomly selected members of their opt-in panel using sampling rates that accounted for the population distribution and prior panelist response rates to yield a responding sample that approximates U.S. adult population distributions for age, gender, race, Hispanicity, education, family income, and census region. The data are not weighted, as weighting is not expected to improve accuracy with this kind of sample (MacInnis

et al. 2018). There were 75,722 panelists invited by email, of whom 2,822 opened the questionnaire. Qualifying participants completed the questionnaire, passed an attention check, and said they had not answered the questionnaire before. Data were collected on the Qualtrics platform from March 5 to March 19, 2020. The survey was an omnibus containing questions on a variety of topics, with a designed median length of 20 minutes. The political knowledge items were designed to take approximately 4 minutes. The study allows tests of all hypotheses except H5.

Participants were assigned to a 3 × 2 × 2 experiment design that independently randomized conditions for instructions, feedback, and the use of pictures.

Each participant was randomly assigned to one of three *instruction groups*. Instruction group 0 (the control group) received no instruction. Instruction group 1 received an instruction not to look up the answers to questions, as follows:

> **We are interested in the guesses people make when they do not know the answer to a question. We will ask you several questions. Some may be easy, but others are meant to be so difficult that you will have to guess.**
>
> **In fact, for some of these questions, if you answer correctly, we will know that you probably looked up the answer.**
>
> **Please do not look up the answers you do not know. Instead, please just make your best guess.**

A comparison of groups 0 and 1 tests for an instruction effect (H1). Instruction group 2 received the same instruction not to look up the answers and was also asked to make a promise not to do so: “Will you please promise to try your best without looking up any answers? Or do you not want to make that promise?” Comparison of groups 1 and 2 tests for a commitment effect (H2). Hypothesis 3, that people reliably admit to having looked up answers when they

have done so, will be tested for the set of respondents assigned to receive no instruction and to receive an instruction.

Each participant was randomly assigned to one of two *feedback group*s. Feedback group 0 received no feedback about their answers. In feedback group 1, a participant who correctly answered the first catch question, to which we assume no one knows the answer without having looked it up, was given feedback to discourage them from looking up more answers. The first open-ended catch question asked, "In what year did the Supreme Court of the United States decide Geer v. Connecticut?" This is an obscure case that even lawyers are usually unfamiliar with. If the respondent answered correctly and was in the feedback-receiving group, they were asked, "You are right! Did you look up the answer to that question, or did you already know it yourself?" Then, regardless of their response (because the answer that they already knew it is assumed to be a lie), they were told, "Please do not look up the answers yet! For the next questions, please make your best guess without any help. After you do that, we will show you the right answers." After this, respondents were asked a second open-ended catch question: "In what year was the Alaska Purchase Treaty signed?" Among respondents in the no-instruction group, the feedback hypothesis (H4) is tested by comparing the proportion of correct answers to the second catch question in the two feedback groups.

Each participant was randomly assigned to one of two *picture groups*. The control group received four textual political knowledge questions without pictures. These questions from the ANES asked the position held by a named individual: "What job or political office is held by (NAME)?" The questions asked about Mike Pence, Nancy Pelosi, Angela Merkel, Lemanu Peleti Mauga, and John Roberts. Responses were open-ended. Pence, Pelosi, Merkel, and Roberts were, respectively, Vice President of the United States, Speaker of the U.S. House of

Representatives, Chancellor of Germany, and Chief Justice of the United States. Mauga was Lieutenant Governor of American Samoa and was included as a catch question.

The office recall questions about Pence, Pelosi, Merkel, and Roberts were asked before the instructions or feedback, so they were not affected by those conditions. The picture group was shown a photo and asked, "What job or political office is held by this person?" Participants in both groups typed their answers to the open-ended questions, and responses were coded by a computer script (included in online supplemental materials) of the type developed and validated by DeBell (2013) and used by ANES to code items of this type, which credited answers as correct when they included key text, such as "VP," "veep," or "Vice President" for Pence.

A comparison of the catch question results for the two picture groups tests Hypothesis 6, that using pictures prevents lookups.

Hypothesis 7 is tested by comparing the associations of age, gender, and education with knowledge scores from the text- and photo-based knowledge measures calculated using the number of correct answers to the Pence, Pelosi, Merkel, and Roberts items.

To test hypotheses 8, 9, and 10, concerning the validity of knowledge items using text or pictures, knowledge scores are tested for criterion or convergent validity for the two groups. For all three hypotheses, "equivalence" of results means the effect size of a difference is tiny and is of no substantive importance, such as $r < .05$. For H9 the convergent knowledge measure is based on placing Donald Trump to the right of Hillary Clinton and the Democratic Party, and placing Clinton and the Democratic Party to the left of the Republican Party.

Study 2 (DeBell, 2021b) used the same design and sample source as study 1, though no individual was allowed to complete both studies. Data were collected from 1557 qualifying

participants from November 21 through 25, 2019, to allow tests of hypotheses 1, 2, and 6 through 10 using the same questions administered in Study 1.

Study 3 was the ANES 2019 Pilot Study (ANES, 2020a). It used a non-probability sample of 3,000 U.S. citizens aged 18 or older who volunteered to complete surveys over the Internet in exchange for points redeemable for gift cards. The sample was provided by YouGov. Data were collected from December 20 through 31, 2019. The survey was an omnibus containing questions on a variety of topics, with a median length of 34 minutes. The questionnaire, data, and documentation for all ANES studies are available online. The data are weighted and the opt-in sample is composed of people selected from a panel using a sample-matching procedure expected to produce more reliable population estimates than many other opt-in samples. Study 3 allows tests of hypotheses 1 and 3, the instruction effect and accurate admission. Instructions not to guess (ANES, 2020a) were worded differently from studies 1, 2, and 4.

Study 4 was the ANES 2020 Exploratory Testing Survey (ANES, 2020b). It used a non-probability sample of 3,080 U.S. citizens aged 18 or older who volunteered to complete surveys over the Internet in exchange for payments. The sample was sourced from three vendors: Bovitz, Dynata, and Toluna. Data were collected on the Qualtrics platform from April 10 through 18, 2020. The survey had a designed median length of 30 minutes. The political knowledge items were designed to take approximately 2 minutes. This questionnaire allows tests of hypotheses 1, 4, and 5 (instruction effect, feedback effect, and curiosity motive).

Study 5 was the ANES 2020 Social Media Study (ANES, 2021). It used a weighted probability sample of 5,750 U.S. citizens age 18 or older who were recruited to join a survey panel and to complete surveys over the Internet on a regular basis. The overall response rate (accounting for initial panel recruitment, panel attrition, as well as response to this survey) was

3.2%. The sample was provided by NORC at the University of Chicago and data were collected on an instrument programmed by NORC. The survey had a median length of 23 minutes and the knowledge questions were designed to take about 2 minutes. The study allows tests of hypotheses 6, 7, and 8: photos prevent lookups, form affects group differences, and photo validity is equivalent to text-based item validity. In this study, unlike the others, the text-based knowledge question (but not the photo question) was preceded by a request not to look up the answers, and knowledge questions asked about the Speaker of the House, the Chief Justice, and a catch question.

## Results

Sampling errors and significance tests reported for studies 1 through 4 apply to experimental randomization, but not to population estimates, as these are not probability samples. Studies 3 and 5 are weighted; the others are unweighted. All analyses use simple random sample methods for the calculation of sampling errors, and for studies 1–4 these should be interpreted in the context of non-probability samples.

*Hypothesis 1* was that there is an *instruction effect*: asking respondents not to look up the answers will reduce number of people who do it. Lookups after the instruction are indicated by a correct answer to a catch question.

Results for tests of this hypothesis (and for all effects of treatments on the percentage of respondents looking up answers – *H1, H2, H4, H5, H6*) are shown in Table 2. In Study 1, without instruction, the first catch question was answered correctly by 18.1% of participants, indicating they looked up the answer. With an instruction not to look up the answer, 5.6% were correct, for a reduction of 12.5 points ($p < .001$). For the second catch question, 23.5% answered

correctly with no instructions compared to 7.0% with instructions (difference 16.5 points, $p <$ .001).

[TABLE 2 ABOUT HERE]

In Study 2, the same instruction reduced the number of correct answers from 16.7% to 5.9%, (difference 10.8 points, $p < .001$). In Study 3, differently worded instructions reduced correct answers from 22.0% to 8.9% (difference 13.1 points, $p < .001$). In Study 4, instructions (matching Study 1 and 2) reduced correct answers from 29.1% to 10.8% (difference 18.3 points, $p < .001$) for the first catch question. For the second catch question in Study 4, instructions reduced correct answers from 19.2% to 7.4% (difference 11.8 points, $p < .001$). All six tests in all four studies support hypothesis 1, with large effects.

*Hypothesis 2* was that there is a commitment effect where a request not to search coupled with a request for a promise not to do so is more effective than the request alone. In studies 1 and 2, respectively, 98.6% and 95.0% of participants made the promise when asked.

In Study 1, with an instruction and the request for commitment, the Geer catch item was answered correctly by 1.9% of participants. Compared to 5.6% answering correctly with the instruction but without the commitment request, this difference is significant (difference 3.7, $p =$ .002). However, for the second catch item, respondents were more likely to answer correctly after having been asked for a commitment than with the instruction alone (8.1% compared to 7.0%; $p = .641$). In Study 2, with instruction and the request for commitment the Geer item was answered correctly by 4.8% of respondents. Compared to the 5.9% answering correctly with the instruction but without the commitment request, this difference is not significant (difference 1.1, $p = .438$). Mixed results mean no clear support for a commitment effect was found.

*Hypothesis 3* was that substantially everyone who looks up the answers will admit it. In Study 1, 30% of participants who looked up the answer to the Geer question admitted having done so (*s.e.*= 4.0, $N = 134$; not shown in tables), 19% lied and denied having looked up the answer, and 51% did not answer the question. Study 3 found that when no instructions were given, 75% of those who looked up answers admitted it, while 25% gave a false denial (*s.e*. 2.4, $N = 332$; this gives 95% confidence that fewer than 80% admit to their lookup). When instructed not to look up answers only 56% of those who disobeyed instructions admitted it (*s.e.* = 4.3, $N = 132$) (difference 19 points, *s.e*. = 6.1, $t = 3.16$, $p = .002$). Hypothesis 3 is rejected (for any standard where "substantially everyone" means 80% or more); many people lie about looking up answers.

*Hypothesis 4* was that feedback is effective: if we ask respondents who have looked up an answer to stop doing so, many will stop. To test this hypothesis in Study 1 we compare the percentage of respondents looking up answers on the second catch question (the year of the Alaska Purchase), after feedback was available. Table 2 shows these lookup rates. With no instructions, 23.5% of participants looked up the answer. Feedback reduced the lookup rate to 13.4% for the subsequent question (difference of 10.1 points, $p = .003$), but this was less effective than instructing participants not to look up answers in the first place (7% – a further reduction of 6 points, *s.e*. 2.67, $t = 2.40$, $p = .017$).

Combining instruction with the request for a commitment or with feedback had no apparent effect compared to instruction alone (with resulting lookup rates of 8%). However, the combination of all three elements – instruction, the commitment request, and feedback after a lookup – reduced the lookup rate to 3% (a reduction of 5 points, *s.e*. = 2.0, $t = 2.57$, $p = .011$.) In terms of the relative contribution of the mechanisms, giving instructions alone reduced the

lookup rate by over two-thirds (e.g. in Study 1 from a baseline of 24% to 7%), and although adding feedback or a commitment request singly made no detectable difference, adding both of these reduced lookups by a further half (i.e., from 7% to 3%).

Study 4 tested the feedback hypothesis slightly differently, with the Mauga catch question. With no instructions, 19% of participants looked up the answer. Feedback reduced the lookup rate to 9% (difference of 9.8 points, $p < .001$), also supporting H4.

*Hypothesis 5* was that acknowledging participants' curiosity by telling them they will be shown the correct answers reduces lookups. In study 4, instructions to not look up the answers randomly included or excluded the statement that "After you have guessed, we will show you the right answers." Without this statement, the lookup rate was 10.8%. With this statement, the lookup rate was two points lower (8.8%; difference 2.0, $p = .129$), but not significantly different, rejecting Hypothesis 5.

*Hypothesis 6* was that the picture format generates fewer lookups than the traditional text question format, where lookups are measured by correct answers to the catch question about the Lt. Gov. of American Samoa. In Study 1, with the traditional text question, 9.7% of participants answered the question correctly, compared to 1.1% for the photo question (difference 8.6, $p < .001$). In Study 2, 10.2% answered the text question correctly, compared to 0.0% for the photo question (difference 10.2, $p < .001$). In Study 5, 3.3% answered the text question correctly, compared to 1.6% for the photo question (difference 1.7, $p < .001$). All three studies support Hypothesis 6 and find photos reduce lookups to a large degree.

*Hypothesis 7* was that question format affects group differences in political knowledge for age, gender, and education. At issue is whether measured heterogeneity in knowledge within these three categories depends on the use of photo or text formats. We test this by measuring the

association between knowledge and gender, age, and education, separately for verbal and visual knowledge, and comparing the two modes. Results from Studies 1, 2, and 5 uniformly reject the hypothesis; no significant differences are detected, which is a favorable indicator that using a photo format will not seriously distort knowledge results for population subgroups. Results from 18 regression analyses are summarized in Table 3 and shown fully in online Appendix Table A2 (https://osf.io/kzsta/). Comparisons of means for groups revealed that people with less than a high school credential did significantly better with visual than verbal questions, but other comparisons by gender, age, and party identification found negligible question type differences (online Appendix Table A3).

[TABLE 3 HERE]

Hypotheses 8, 9, and 10 were about the validity of the photo format compared to the text format. *Hypothesis 8* was that the association between political knowledge and interest in politics will be equivalent or stronger in the photo format compared to traditional text.

In Study 1, in the text format, the association between knowledge and interest in politics is $r = .36$, regression $b = 0.40$ (*s.e.* = 0.037, $t = 10.81$, $p < .001$). In the photo format, the association is $r = .37$, regression $b = 0.41$ (*s.e.* = 0.038, $t = 10.75$, $p < .001$). The difference between the two is not significant (difference = 0.01, *s.e.* = 0.053, $t = 0.21$, $p = .84$). In Study 2, in the text format the association between knowledge and interest in politics is $r = .40$, regression $b = 0.45$ (*s.e.* = 0.037, $t = 12.12$, $p < .001$); in the photo format, $r = .47$, regression $b = 0.52$ (*s.e.* = 0.035, $t = 14.55$, $p < .001$). The difference between the two is not significant (difference = 0.07, *s.e.* = 0.051, $t = 1.37$, $p = .170$). In study 5, in the text format the association between knowledge and interest in politics is $r = .36$, $b = 0.25$ (*s.e.* = 0.012, $t = 21.12$, $p < .001$); in the photo format, $r = .39$, regression $b = 0.28$ (*s.e.*= 0.013, $t = 22.23$, $p < .001$). The difference

between the two is marginally significant (difference = 0.03, *s.e.* = 0.018, $t = 1.87$, $p = .06$) and favors the photo format. In all three studies, hypothesis 8 is supported; both question formats find very similar associations between political knowledge and interest.

*Hypothesis 9* was that compared to text items, picture items would be equivalently or more strongly associated with an alternative political knowledge measure based on liberal-conservative placement of parties and candidates.

In Study 1, in the text format, the association between knowledge and liberal-conservative placement accuracy is $r = .34$, regression $b = .30$ (*s.e.* = 0.29, $t = 10.30$, $p < .001$). In the photo format, the association is $r = .40$, regression $b = .34$ (*s.e.* = 0.028, $t = 11.89$, $p < .001$). The difference is not significant (difference = 0.04, *s.e.* = 0.04, $t = 0.99$, $p = .32$).

In Study 2, in the text format, the association between knowledge and liberal-conservative placement accuracy is $r = .55$, regression $b = 0.50$ (*s.e.* = 0.027, $t = 18.34$, $p < .001$). In the photo format, the association is $r = .47$, regression $b = .40$ (*s.e.* = 0.027, $t = 14.69$ $p < .001$). The difference is very small ($r = .08$) but statistically significant and lower for the photo rating (difference = 0.10, *s.e.* = 0.038, $t = 2.57$, $p = .010$), so hypothesis 9 is rejected.

*Hypothesis 10* was that picture items would perform as well or better than text items as a moderator of the effect of birthright citizenship policy preferences on voting for Trump. This is tested by comparing results from two logistic regression models in which voting for Trump is the dependent variable and the independent variables are political knowledge, opinion about ending birthright citizenship, and an interaction term for the first two independent variables. The interaction term indicates the effect of knowledge as a moderator of the effect of the policy preference on support for Trump. Models are run separately for the respondents in the text and

picture groups. If the coefficient for the interaction term in the photo condition is larger or not different from the coefficient in the text condition, the hypothesis is supported.

In both Studies 1 and 2 the included variables are significant predictors of voting for Trump but the models based on text or photo questions do not significantly differ. In Study 1, the coefficient for the interaction term in the model for the text group, $b$, = 2.07 (*s.e.* = 0.846), is larger than the coefficient for the interaction term in the model for the picture-based question, $b$ = 1.50 (*s.e.* = .892), but the difference is not statistically significant ($p$ = .642). Similarly, in Study 2, there is no difference detected between the interaction terms for the text and photo questions. Hypothesis 10 is supported. Full model results are shown in Table 4. (This result for *H10* used logistic regression, but the same support is found with OLS regression (not shown), which avoids possible confounds in the comparison of logistic coefficients—see Allison, 1999, on comparing logistic coefficients; see Hellevik, 2009, arguing that OLS is valid and perhaps preferable for analysis of binary outcomes).

[TABLE 4 HERE]

A qualitative summary of these findings is provided in Table 5.

[TABLE 5 HERE]

## Discussion & Conclusion

In recent practice the most common approach to the lookup problem may have been to ignore it; many studies have administered political knowledge tests as part of online questionnaires without taking any steps to prevent or discourage respondents from looking up the answers. For example, in 2016 the ANES administered the same kind of political knowledge test online that it began administering in face-to-face interviews decades ago. The state of the art now

indicates the old face-to-face methods are not well suited for online questionnaires, but effective methods are available to limit lookups online.

The current studies reinforce the lesson that answers to text-based political knowledge questions are highly susceptible to measurement error due to participants often looking up the answers instead of answering based on their prior knowledge. The current studies also deliver findings supporting three recommendations to improve research methods involving political knowledge. These recommendations, elaborated below, are (1) where feasible, to use photo-based questions to prevent lookups, and (2) where photo-based questions are not feasible, to use a combination of instructions, commitment requests, and feedback to discourage lookups, and (3) to use catch questions to detect lookups.

First, visual office-recall items collecting open-ended answers effectively eliminate lookups, take no more time than text-based questions, and the answers are easy to code automatically. Most importantly, they produce data with similar validity by several measures; although Hypothesis 9 regarding liberal-conservative identification was rejected in one study, the effect size was small enough ($r$=.08) that the photo and text formats might be considered equivalent. Given their resistance to lookups, photos therefore appear to be a preferable format for questions amenable to a visual presentation. Also, people with less than a high school credential do significantly better with visual prompts than with text, so the visual format may reduce an unnecessary bias that confounds political knowledge with verbal skills. Otherwise, the use of photo- or text-based questions had no detected effects on the associations between age and knowledge, gender and knowledge, or education and knowledge (*H7*).

Second, to effectively stop lookups in text-based questions, instructions can be combined with both a commitment from participants and adaptive feedback to ask respondents who do look

up answers (as revealed by a catch question, or potentially with a script in the questionnaire page) to stop doing so. In the current Study 1, lookups were nearly eliminated by this technique. This method is not quite as effective as using photos, and takes longer to administer, but by emphasizing norms of following instructions and keeping one's word, and by using questions designed to make it difficult to look up the answers, lookups can be largely prevented.

Third, to detect lookups, self-reports are subject to significant under-reporting. Using catch questions of the type described by Motta et al. (2017), these studies indicate that literature that has relied on self-reported lookups has been relying on an assumption that may be about 75% accurate. Future studies that need to identify participants who looked up answers may get a more accurate estimate using catch questions.

In addition to the above findings with positive application to practice, other findings caution against certain practices as suboptimal or ineffective. Commitment requests were effective in only one of three tests, and the combination of instructions and feedback was no more effective than instructions alone; it was only the triple combination of instruction, commitment request, and feedback that produced a significant benefit beyond instruction alone. Also, a test of the curiosity motive rejected that hypothesis (*H5*); either intellectual curiosity is not a significant factor in many lookups, or telling participants that the correct answers would be revealed later was not enough to quell their curiosity.

Limitations of the current studies are of note. All used samples in the USA and four of the five were non-probability samples, limiting generalizability, in part because impression management varies culturally (Lalwani et al., 2006). Lookup behavior could differ between users of different devices, such as smartphones and desktop computers, because devices affect the

convenience of online search, but this analysis did not investigate device effects. This study also did not examine multiple-choice questions.

For continuing research, several directions seem valuable. For lookup detection, catch questions are very useful but their validity as indicators of lookups on other items is open to doubt because they are intrinsically more difficult than other questions. More difficult questions are more likely to elicit lookups, at least among participants with higher levels of education (Gummer & Kunz, 2019). Catch questions could therefore over-estimate lookups that would occur in their absence. As an alternative or supplement to catch questions, webpage scripts can detect and report when users switch to different applications or browser tabs (see Diedenhofen & Musch, 2017; Höne et al., 2020). Such scripts can identify likely lookup behavior and, if integrated with adaptive instrumentation, can also be used to trigger feedback. The weakness of these tools is that they generally cannot conclusively determine whether a lookup was performed, but they are very promising for further study.

For lookup prevention, the wording of instructions, feedback, and commitment requests has yet to be optimized. It is interesting that the instruction effects observed here were larger than the effects in some other studies, which could be attributable to the use of wordings that stated the researchers' ability to infer that some correct answers would be lookups (analogous to a "genuine pipeline" approach; see Hanmer et al., 2014). Wording experimentation is warranted. Also, the studies had mixed results for the commitment effect alone, which suggests the potential for further research to reveal the importance of different wordings, samples, or contexts for the effectiveness of such requests. Finally, though five studies were reported here and most hypotheses were tested two or three times, the key finding that a triple combination of

instructions, commitment, and feedback produced the fewest lookups was tested in only one study, so replication would be valuable.

## References

Allison, P.D. (1999). Comparing logit and probit coefficients across groups. *Sociological Methods and Research*, *28*, 186-208. https://doi.org/10.1177/0049124199028002003

ANES. (2020a). ANES 2019 Pilot Study [dataset and codebook]. American National Election Studies (www.electionstudies.org). Dataset version: February 4, 2020.

ANES. (2020b). ANES 2020 Exploratory Testing Study [dataset and codebook]. American National Election Studies (www.electionstudies.org). Dataset version: July 27, 2020.

ANES. (2021). ANES 2020 Social Media Study Initial Release (Pre-election only). [dataset and codebook]. American National Election Studies (www.electionstudies.org). Dataset version: January 28, 2021.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*, *20*, 351-368. https://doi.org/10.1093/pan/mpr057

Bullock, J. G., Gerber, A.S., Hill, S. J., & Huber, G. A. (2015). Partisan bias in factual beliefs about politics. *Quarterly Journal of Political Science*, *10*, 519-578. http://dx.doi.org/10.1561/100.00014074

Clifford, S., & Jerit, J. (2014). Is there a cost to convenience? An experimental comparison of data quality in laboratory and online studies. *Journal of Experimental Political Science*, *1*, 120-131. https://doi.org/10.1017/xps.2014.5

Clifford, S., & Jerit, J. (2015). Do attempts to improve respondent attention increase social desirability bias? *Public Opinion Quarterly*, *79*, 790-802. https://doi.org/10.1093/poq/nfv027

Clifford, S., & Jerit, J. (2016). Cheating on political knowledge questions in online surveys: An assessment of the problem and solutions. *Public Opinion Quarterly*, *80*, 858-887. https://doi.org/10.1093/poq/nfw030

Conrad, F.G., Couper, M.P., Tourangeau, R., and Zhang, C. (2017). Reducing speeding in web surveys by providing immediate feedback. *Survey Research Methods*, *11*, 45-61. https://doi.org/10.18148/srm/2017.v11i1.6304

Converse, P. (1964). The nature of belief systems in mass publics. In D. E. Apter (ed.), *Ideology and its discontents* (pp. 206-261). New York: Free Press of Glencoe.

DeBell, M. (2013). Harder than it looks: Coding political knowledge on the ANES. *Political Analysis, 21*, 393-406. https://doi.org/10.1093/pan/mpt010

DeBell, M. (2019). Effect of instruction on respondents looking up the answers to political knowledge questions on the ANES 2018 Pilot Study. Memorandum, April 2, 2019.

DeBell, M. 2021a. Political Knowledge Study 1 [dataset and questionnaire]. Available at https://osf.io/kzsta/

DeBell, M. 2021b. Political Knowledge Study 2 [dataset and questionnaire]. Available at https://osf.io/kzsta/

Delli Carpini, M. X., & Keeter, S. (1996). *What Americans know about politics and why it matters*. New Haven, CT: Yale University Press.

Diedenhofen, B., & Musch, J. (2017). PageFocus: Using paradata to detect and prevent cheating on online achievement tests. *Behavior Research Methods*, *49*(4), 1444-1459. https://doi.org/10.3758/s13428-016-0800-7

Fricker, S., Galesic, M., Tourangeau, R., & Yan, T. (2005). An experimental comparison of web and telephone surveys. *Public Opinion Quarterly*, *63*, 370-392. https://doi.org/10.1093/poq/nfi027

Gilens, M. (2001). Political ignorance and collective policy preferences. *American Political Science Review, 95*, 379-396. https://doi.org/10.1017/S0003055401002222

Gooch, A., & Vavreck, L. (2019). How face-to-face interviews and cognitive skill affect item non-response: A randomized experiment assigning mode of interview. *Political Science Research and Methods*, *7*, 143-162. https://doi.org/10.1017/psrm.2016.20

Goren, P. (1997). Political expertise and issue voting in presidential elections. *Political Research Quarterly*, *50*, 387-412. https://doi.org/10.2307/448963

Gummer, T., & Kunz, T. (2019). Relying on external information sources when answering knowledge questions in web surveys. *Sociological Methods and Research*, published online November 4, 2019, https://doi.org/10.1177/0049124119882470

Guggenheim, L., McClain, C., Nardis, Y., & Brader, T. (2019). *Comparing the face-to-face and web modes in the ANES 2016 Time Series Study*. ANES technical report. https://electionstudies.org/wp-content/uploads/2019/09/ANES2016_ModeReport.pdf

Hanmer, M.J., Banks, A.J., & White, I.K. (2014). Experiments to Reduce the Over-Reporting of Voting: A Pipeline to the Truth. *Political Analysis*, 22, 130-141. https://doi.org/10.1093/pan/mpt027

Hellevik, O. (2009). Linear versus logistic regression when the dependent variable is a dichotomy. *Quality & Quantity*, *43*, 59-74. https://doi.org/10.1007/s11135-007-9077-3

Höhne, J. K., Schlosser, S., Couper, M. P., & Blom, A. G. (2020). Switching away: Exploring on-device media multitasking in web surveys. *Computers in Human Behavior*, *111*, 106417. https://doi.org/10.1016/j.chb.2020.106417

Jensen, C., & Thomsen, J. P. F. (2014). Self-reported cheating in web surveys on political knowledge. *Quality & Quantity*, *48*, 3343-3354. https://doi.org/10.1007/s11135-013-9960-z

Jerit, J. (2019). *Political knowledge items & the catch questions*. Presentation to the Board of the American National Election Studies, March 22, 2019, Austin, Texas.

Kleinberg, M. S., & Lau, R. R. (2019). The importance of political knowledge for effective citizenship: Differences between the broadcast and internet generations. *Public Opinion Quarterly*, *83*, 338-362. https://doi.org/10.1093/poq/nfz025

Lalwani, A. K., Sharon, S., & Johnson, T. (2006). What is the relation between cultural orientation and socially desirable responding? *Journal of Personality and Social Psychology*, *90*, 165-178. https://doi.org/10.1037/0022-3514.90.1.165

Lupia, A. (2016). *Uninformed: Why people know so little about politics and what we can do about it*. New York: Oxford University Press.

MacInnis, B., Krosnick, J.A., Ho, A.S., and Cho, M. (2018). The accuracy of measurements with probability and nonprobability survey samples: A replication and extension. *Public Opinion Quarterly*, 82, 707-744. https://doi.org/10.1093/poq/nfy038

Motta, M. P., Callaghan, T. H., & Smith, B. (2017). Looking for answers: Identifying search behavior and improving knowledge-based data quality in online surveys. *International Journal of Public Opinion Research*, *29*, 575-603. https://doi.org/10.1093/ijpor/edw027

Munzert, S., & Selb, P. (2015). Measuring political knowledge in web-based surveys: An experimental validation of visual versus verbal instruments. *Social Science Computer Review*, *35*, 167-183. https://doi.org/10.1177/0894439315616325

Nosek, B.A., Greenwald, A.G., and Banaji, M. (2006). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, 31, 166-180. DOI: https://doi.org/10.1177/2F0146167204271418

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*, 867-972. https://doi.org/10.1016/j.jesp.2009.03.009

Paulhus, D.L., & John, O.P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality*, *66*, 1025-1060. https://doi.org/10.1111/1467-6494.00041

Prior, M. (2014). Visual political knowledge: A different road to competence? *The Journal of Politics*, *76*, 41-57. https://doi.org/10.1017/S0022381613001096

Prior, M., & Lupia, A. (2008). Money, time, and political knowledge: Distinguishing quick recall and political learning skills. *American Journal of Political Science*, *52*, 169-183. https://doi.org/10.1111/j.1540-5907.2007.00306.x

Robison, J. (2014). Who knows? Question format and political knowledge. *International Journal of Public Opinion Research*, *27*, 1-21. https://doi.org/10.1093/ijpor/edu019

Smith, B., Clifford, S., & Jerit, J. (2020). How internet search undermines the validity of political knowledge measures. *Political Research Quarterly*, *73*, 141-155. https://doi.org/10.1177/1065912919882101

Stiers, D., & Hooghe, M. (2021). A picture is worth a thousand words. A visual test of political knowledge: Does it reduce the effect of age and gender? *International Journal of Public Opinion Research*, *33*(1), 137-146. https://doi.org/10.1093/ijpor/edz052

Strabac, Z., & Aalberg, T. (2011). Measuring political knowledge in telephone and web surveys: A cross-national comparison. *Social Science Computer Review*, *29*, 175-192. https://doi.org/10.1177/0894439310371340

Style, H., & Jerit, J. (2021). Does it matter if respondents look up answers to political knowledge questions? *Public Opinion Quarterly, 84,* 760-775. https://doi.org/10.1093/poq/nfaa038

Vezzoni, C., & Ladini, R. (2017). Thou shalt not cheat: How to reduce internet use in web surveys on political knowledge. *Italian Political Science Review/Rivista Italiana Di Scienza Politica, 47*(3), 251-265. https://doi/10.1017/ipo.2016.25