How Internet Search Undermines the Validity of Political Knowledge Measures

Brianna Smith
Assistant Professor
U.S. Naval Academy

Scott Clifford
Associate Professor
University of Houston

Jennifer Jerit
Professor
Stony Brook University

Forthcoming at *Political Research Quarterly*

Abstract

Political knowledge is central to understanding citizens' engagement with politics. Yet, as surveys are increasingly conducted online, participants' ability to search the web may undermine the validity of factual knowledge measures. Recent research shows this search behavior is common, even when respondents are instructed otherwise. However, we know little about how outside search affects the validity of political knowledge measures. Using a series of experimental and observational studies, we provide consistent evidence that outside search degrades the validity of political knowledge measures. Our findings imply that researchers conducting online surveys need to take steps to discourage and diagnose search engine use.

Keywords: political knowledge, online surveys, validity

Political knowledge is considered the “currency of citizenship” because it helps people process new information and link their values and interests to their attitudes. A common way to operationalize this concept is with questions asking individuals to recall specific facts from memory (Zaller 1992). Given the centrality of knowledge in studies of public opinion and political behavior, numerous debates have occurred over measurement, from the use of a “Don’t Know” response option (Luskin and Bullock 2011; Miller and Orr 2008; Mondak 2001) and the proper coding of open-ended questions (Gibson and Caldeira 2015), to item difficulty (Ahler and Goggin 2017) and the differential functioning of items across demographic groups (Abrajano 2015). As more surveys are self-administered over the internet, a new measurement challenge has arisen: respondents, who typically complete surveys at their own pace and without interviewer interaction, can use search engines to look up information.[1]

Outside search has been detected in several subject populations with search rates reaching as high as 41%. Evidence of search engine use comes from a variety of sources, including studies showing higher factual knowledge scores in the online condition of randomized mode experiments (Burnett 2016; Clifford and Jerit 2014), lower scores when respondents are randomly assigned instructions not to search (Clifford and Jerit 2016; Motta, Callaghan, and Smith 2016; Vezzoni and Ladini 2017), correct answers to extremely difficult and obscure open-ended questions (“catch” questions; Motta, Callaghan, and Smith 2016), and the outright admission of searching (Clifford and Jerit 2016; Jensen and Thomsen 2014). Outside search is

[1] In some studies, this behavior is labeled “cheating.” We reserve that term for instances in which a respondent defies survey instructions not to look up answers, and use the phrase “online search” or “search engine use” in all other situations.

common even in high-quality surveys that have explicitly instructed respondents not to look up answers. For example, 15% of respondents in the American National Election Studies 2018 Pilot Study looked up the answer to at least one of two catch questions seconds after being asked not to do so (among those who received no instructions about search engine use, 25% looked up at least one answer). Despite the prevalence of search engine use, previous work has not definitively answered the question of whether search degrades the measurement properties of knowledge scales.

We contribute to the literature with a series of studies—experimental and observational—that employ a range of criterion measures and varying subject populations.[2] In the first study, we examine the validity of knowledge measures in experiments where subjects are either discouraged from seeking outside assistance on factual questions or told that it is acceptable to do so. In the second study, we examine search engine use in a national online sample of adults (instructed *not* to look up answers) and conduct a similar analysis of the measurement properties of knowledge scales. Findings from both studies suggest that the ability to *recall* information is tied to theoretically-linked concepts such as political interest, while the ability to find that information is related to the effort a person devotes to answering survey questions. Consequently, when respondents seek outside assistance, knowledge scales have weaker associations with criterion outcomes such as political engagement and ideological constraint. Finally, simulations reveal that the validity of knowledge scales is degraded at rates of search behavior that have been observed in past studies (i.e., at "baseline" levels of search engine use). Scholars who are interested in measuring political knowledge should take efforts to minimize and diagnose search behavior. This recommendation can be applied in other contexts such as the

[2] Replication data are available as part of Supplemental Materials at http://prq.sagepub.com.

assessment of science knowledge (Kahan 2016), verbal knowledge (Alwin 1991), and reasoning ability (Bialek and Pennycook 2017).

What Knowledge Scales Measure, With and Without Outside Search

There is growing consensus that political awareness "is best represented with data from survey batteries that measure factual political knowledge" (Mondak 2001, 224). Numerous studies support this claim, showing that the well informed differ from the less informed in a myriad of ways that relate to opinion quality (Althaus 2003; Jacoby 2006; Kam 2005; Lau and Redlawsk 2001). Kinder summarized this literature when he observed that the well informed "are more likely to express opinions in the first place. They are more likely to possess stable opinions—real opinions, opinions held with conviction. They are more likely to use ideological concepts correctly, to cite evidence in political discussions, and to process information sensitively. They are better at retaining new information" (2006, 207). According to this view, multi-item knowledge batteries assess general differences in the amount and richness of knowledge citizens bring to issues of politics (Kinder and Kalmoe 2017, 134-5).

Knowledge scales are intended to capture a latent disposition rather than awareness of a specific set of facts (for a related discussion, see Kahan 2016). Looking up the answer(s) may improve a person's knowledge score, but this search behavior does not necessarily correspond with the latent ability the scale is designed to measure.[3] This situation has a useful analogue in

[3] Kam and Trussler (2016: 791) make a parallel argument regarding the ability of researchers to manipulate this disposition: "requiring subjects to memorize a set of political facts may boost their scores on a political information test but is unlikely to induce subjects to behave like political sophisticates."

the literature on “Don’t Know” responses. As Mondak (2001) and others (Lizotte and Sidman 2009; Mondak and Anderson 2004) have argued, survey protocols encouraging “Don’t Know” responses weaken the validity of knowledge measures because certain personality factors (e.g., self-confidence, risk-taking) are related to guessing. Assuming that respondents with these characteristics guess correctly a percentage of the time, their scores will be inflated relative to non-guessers. While higher scores due to guessing may reveal “partial” knowledge, such scores also reflect the personality characteristics that make people more likely to guess. In a similar way, when respondents seek outside assistance on knowledge questions, their resulting score reflects their learned (i.e., stored) political knowledge as well as the *effort* they are willing to expend to satisfy the demands of the survey, either by using the internet or consulting someone for help. Search engine use may arise out of a desire to be an attentive survey taker, genuine curiosity, or even technological savviness—but these traits are not necessarily indicators of one’s level of political engagement. Thus, outside search can cause knowledge scales to become confounded with other respondent characteristics.

A more optimistic view is that search behavior does not introduce confounds, but merely inflates scores and increases the prevalence of ceiling effects. Even under this perspective, search engine use will harm the descriptive validity of knowledge questions and potentially undermine other aspects of validity. Because knowledge scales often consist of only four or five items (Delli Carpini and Keeter 1996), outside search could push many respondents to the maximum value on the scale and, in turn, weaken the correlation between political knowledge and other variables. Given the state of the existing literature, however, it is unclear whether either or both problems plague the data collected in online surveys.

Despite clear evidence that online administration increases the frequency of outside search, there is only scattered evidence regarding the consequences of this behavior. To our knowledge, just two studies have examined whether search affects the relationship between education and political knowledge and they reach opposite conclusions. Vezzoni and Ladini (2017) experimentally reduced outside search in a sample of Italian adults by manipulating knowledge instructions and report no difference between conditions in the relationship between knowledge and education. In contrast, Jensen and Thomsen (2014) examined the association between education and knowledge among Danish adults who reported outside search and those who did not. They find that the positive effect of education on political knowledge disappears among those who engaged in search.[4]

The empirical evidence regarding other criterion outcomes is sparse. Clifford and Jerit (2014) report a weaker correlation between political interest and knowledge among participants who were randomly assigned to complete a questionnaire online in comparison to a lab condition. However, this difference could be the result of other factors that vary across modes (e.g., level of distraction). Finally, Clifford and Jerit (2016) explore whether predictive validity *improves* among respondents who received a randomized set of instructions not to look up answers. They found a stronger relationship between knowledge and the complexity of open-ended responses when outside search was discouraged. Notwithstanding the suggestive patterns

[4] Gooch (2015) examines the relationship between education and general cognitive ability in a mode experiment involving a face-to-face condition and a self-administered condition. However, the study took place in the CBS research facility in the MGM Grand Hotel in Las Vegas, making the self-administered condition conceptually different than a traditional online survey.

from these two studies, the question of whether outside search undermines the validity of knowledge measures has not been extensively investigated—i.e., across a range of outcome measures and commonly-used subject populations. Moreover, existing evidence is unclear on whether outside search merely weakens the relationship between knowledge and other variables (due to ceiling effects), or whether outside search inadvertently strengthens the relationship between knowledge with other variables (due to confounding effects).

In the analyses that follow, we explore this topic in terms of convergent, discriminant and predictive validity. Convergent validity pertains to whether a measure is related to conceptually-related variables. Because the existing literature has shown that a person's level of interest in and attention to politics is the strongest predictor of his or her level of knowledge (Delli Carpini and Keeter 1996), we examine whether outside search weakens the relationship between knowledge and political interest. Discriminant validity assesses whether measures that should *not* be related to one another in fact show no relationship. As our test, we investigate whether search behavior causes knowledge scores to become confounded with the effort a respondent is willing to put into the survey. Because searching for an answer is more effortful than guessing, search behavior should cause knowledge scores to become more strongly related to survey effort. Support for this hypothesis would demonstrate that outside search does not merely weaken the relationship between knowledge and other variables due to ceiling effects; instead, search causes knowledge to become confounded with a different factor altogether.[5] Finally, predictive validity concerns

[5] Effort in an online survey may vary for a variety of reasons—e.g., some respondents may be less agreeable, face more distractions, or have less interest in the specific topics of the survey—none of which should be strongly related to a person's latent political sophistication.

the extent to which a measure is causally related to outcomes that are specified by the existing literature. In this part of the analysis, we examine the predictive power of knowledge scales in areas where the literature leads us to expect positive relationships between political awareness and specific outcomes, such as political participation, ideological constraint, complexity of open-ended responses, and news comprehension. We determine if outside search compromises the explanatory power of political knowledge scales. Taken together, our analyses provide some of the most extensive evidence regarding the effect of search behavior on the validity of knowledge scales.

Empirical Evidence

We examine how outside search affects the measurement properties in two different types of analyses. In Studies 1a and 1b, we *manipulate* the behavior of interest with randomly assigned instructions that either discouraged respondents from online lookup or indicated that it was permissible to look up answers. In Study 2, we *measured* search behavior as it naturally occurs in an online survey and analyze the relationship between political knowledge and criterion measures among respondents who did and did not look up the answers. Although each approach has limitations, Studies 1 and 2 tell a consistent story about the effect of search engine use on the validity of knowledge scales. The final set of analyses employ simulations to examine the predictive power of knowledge (i.e., coefficient strength) across levels of outside search observed in past research.

Studies 1a & 1b: Experimental Data and Measures

In the spring and fall of 2017 we administered two experiments on undergraduate students recruited from required courses at a large public university in the southern United

States.[6] Respondents received extra credit for completing the study. Each experiment was administered as part of an online survey, which took about 20 minutes to complete. Study 1a recruited 1,170 participants in the spring, while Study 1b recruited 880 participants in the fall. Participants in Study 1b were invited to complete a second wave about 14 days later, and 644 participants did so (for an attrition rate of 27%).[7] The two studies are nearly identical in terms of design, so we begin by discussing the common elements and then describe where the two experiments differ.

After completing some basic questions, respondents were randomly assigned to one of two conditions that had different instructions for answering the political knowledge questions. In the *discourage condition*, respondents saw this set of instructions:

> "Now we have a set of questions concerning various political issues. We want to see how much information about them gets out to the public from television, newspapers, and the like. 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 language has been shown to be an effective method for deterring cheating because people generally comply with requests from an interviewer (Clifford and Jerit 2016). In the *allow condition*, respondents saw a slightly different version:

---

[6] The two studies (hereafter Study 1a and Study 1b) were administered separately and each was approved by the Committee for the Protection of Human Subjects at the University of Houston. We took precautions to ensure that people from the earlier experiment did not participate in the later one.

[7] Panel attrition was unrelated to treatment assignment (Discourage: 27.7%, Allow: 27.3%; $\chi^2(1) = 0.02$, $p = .891$).

> "Now we have a set of questions concerning various political issues. We want to see how much information about them gets out to the public from television, newspapers, and the like. It is alright with us if you use the internet to double check your answer or look for the correct response if you do not already know it."

The purpose of the language in the allow condition was to make explicit an assumption that many respondents appear to have—namely, that it is permissible to search for answers on factual questions.

Past work shows that students search for answers even when explicitly instructed not to. For this reason and to maximize statistical power, we contrast the *Discourage* condition with the *Allow* condition (i.e., there is no control condition). In a later section, our simulations show that the validity of knowledge scales is harmed at rates of search behavior that have been observed in previous studies. The simulations allow us to determine the coefficient strength of knowledge at levels of search that would have been uncovered if a control condition had been included.

After the instructions respondents answered a series of political knowledge questions. Study 1a had six items pertaining to public officials, partisan control of the Senate and long-standing partisan symbols ($\alpha = .59$). Study 1b included a ten-item battery ($\alpha = .68$) that covered traditional civics themes (e.g., the identity of Chief Justice) and policy-specific topics (e.g., the unemployment rate). Study 1b's questions also varied in terms of the recency of the fact, with some items pertaining to current developments (e.g., new countries added to the Trump administration's travel ban) and others relating to older, established facts (e.g., the length of a U.S. Senate term).[8] The relatively long scale used in Study 1b helps to rule out ceiling effects as

[8] A "Don't Know" option was not included on any of the knowledge items (Miller and Orr 2008). Following the recommendations of Mondak (2001) and others, we employ closed-ended knowledge questions (also see Nadeau and Neimi 1995; Strabac and Aalberg 2011).

the explanation for changes in validity and thus provides a more conservative test of our hypotheses.

Compliance with the instructions was assessed in two ways. First, after completing the knowledge questions, respondents were asked whether they looked up any answers. Second, a difficult “catch” question was inserted in random order among the political knowledge items (Bullock et al. 2015; Motta, Callaghan, and Smith 2016). The catch question asked about the year of an obscure court case (United States vs. Segui), and respondents who provided the correct year were assumed to have looked up the answer. The response distribution to both items indicates that the instruction sets had the intended effect on search behavior. In the analyses below, a person is coded as engaging in outside search if they answered the self-report affirmatively or provided the correct answer to the catch question (also see note 9).

Prior to the knowledge items, we asked a series of questions that would be used as criterion variables for investigating the convergent and discriminant validity of knowledge batteries in each condition. To assess convergent validity, we created a two-item interest scale based on questions that asked about a person’s interest in politics and attention to political news (Study 1a: α =.81, Study 1b: α =.76). To explore discriminant validity, we estimated latent survey effort as a function of several common indicators of satisficing, such as instructed responses, Instructional Manipulation Checks (IMCs), factual manipulation checks (Kane and Barabas 2018), a count of straight-lining in grids, and time spent on pre-treatment survey pages (Berinsky, Margolis, and Sances 2013; Hillygus, Jackson, and Young 2014; Kane and Barabas 2018; Lopez and Hillygus 2018). These items (all measured pretreatment) are designed to capture the respondent’s willingness to read and follow instructions within the survey, and they were scaled together to produce a fine-grained estimate of survey effort (for details see

Appendix, which is included as Supplemental Materials at http://prq.sagepub.com). Crucially, political interest and survey effort are only weakly related to one another (*r* = .13 in Study 1a and *r* = .04 in Study 1b), supporting our argument that these two variables represent distinct constructs.

Study 1b had additional questions that permitted us to examine the explanatory power of knowledge scales when outside search is allowed. First, a series of open-ended questions allowed us to examine the number of considerations respondents could list. In Wave 1, we asked about two salient political issues (mass shootings and high medical costs), and in Wave 2 we asked about likes and dislikes of President Trump, and reasons for and against voter ID laws (the latter followed an article on this topic). A coder tallied the number of considerations raised by respondents which was summed to create a measure of political awareness (Clifford and Jerit 2016). Second, political engagement was measured with three questions (Wave 1) asking whether respondents were registered to vote, whether they voted in the 2016 presidential election, and whether they had engaged in any of four campaign activities (α = .50; attend a meeting, put up a sign, work for a candidate, donate money). Third, we asked respondents positions on 11 political issues in Wave 1 and 11 issues in Wave 2. We used these questions to measure ideological constraint (e.g., Mason 2018), which has previously been linked with political sophistication (e.g., Ansolabehere, Rodden, and Snyder 2008; Federico 2004). Specifically, we recoded the issue attitudes to range from liberal to conservative, averaged the items, and then folded the scale at the midpoint. Thus, people receiving the minimum score have either provided non-substantive responses to all of the questions (i.e., “neither favor nor oppose”), chosen a mix of liberal and conservative positions, or some combination of both response patterns. Those receiving the maximum score have reported strongly held attitudes from a consistent ideological viewpoint.

Finally, in Wave 2, respondents read an article about a recent court decision regarding voter registration laws (based on a *New York Times* story on the topic). Following the article, respondents were asked several questions that gauged their comprehension and recall of key details from the story. Given the length of the article (approximately 550 words) and complexity of the topic, we expected people with higher levels of political knowledge to better understand and retain information from the article (Fiske, Lau, and Smith 1990).

Studies 1a & 1b: Experimental Results

Effects on Search Behavior and Knowledge Scores

The manipulation of knowledge instructions had a sizable effect on search behavior, as shown in the left-hand panels of Figure 1. In Study 1a, 20% of respondents in the discourage condition searched for answers, while 69% of respondents in the allow condition did so ($\chi^2(1) = 279.89$, $p < .001$).[9] Consistent with that pattern, respondents in the allow condition also spent more time answering the knowledge questions (the median respondent averaged 20 seconds per question versus 13 seconds; $t(1168) = 11.08$, $p < .001$). There is a similar pattern for Study 1b. Only 22% of respondents in the discourage condition sought outside assistance, while 72% of

[9] The two measures are strongly related ($\alpha = .77$) and the results are similar regardless of which is used (Self-report: 63% vs. 13%; Catch: 57% vs. 10%). Outside search still occurred in the discourage condition, which is not unusual in student samples (Clifford and Jerit 2016).

respondents in the discourage condition did so ($\chi^2(1) = 223.42$, $p < .001$).[10] Respondents in the allow condition took longer to answer the questions than people in the discourage condition: 22 seconds versus 13 seconds ($t(885) = 10.54$, $p < .0001$).

The knowledge scores are presented in the right-hand panels of Figure 1, and it is clear that merely allowing people to search for answers has a sizeable effect on the observed levels of political knowledge. In Study 1a, scores on the six-item scale were lower in the discourage condition (M = 3.6) than in the allow condition (M = 4.8; $t(1160) = 15.01$, $p < .0001$). Using an instrumental variables approach, we can also calculate the causal effect of outside search on knowledge scores (or the local average treatment effect).[11] This calculation indicates that engaging in outside search increased knowledge scores by 2.4 questions. In Study 1b, respondents answered 4.6 out of 10 questions correctly on average in the discourage condition, while respondents in the allow condition answered 6.5 out of 10 questions correctly ($t(885) = 12.15$, $p < .0001$).[12] An instrumental variables analysis shows that engaging in outside search increased knowledge scores by 3.7 questions.

---

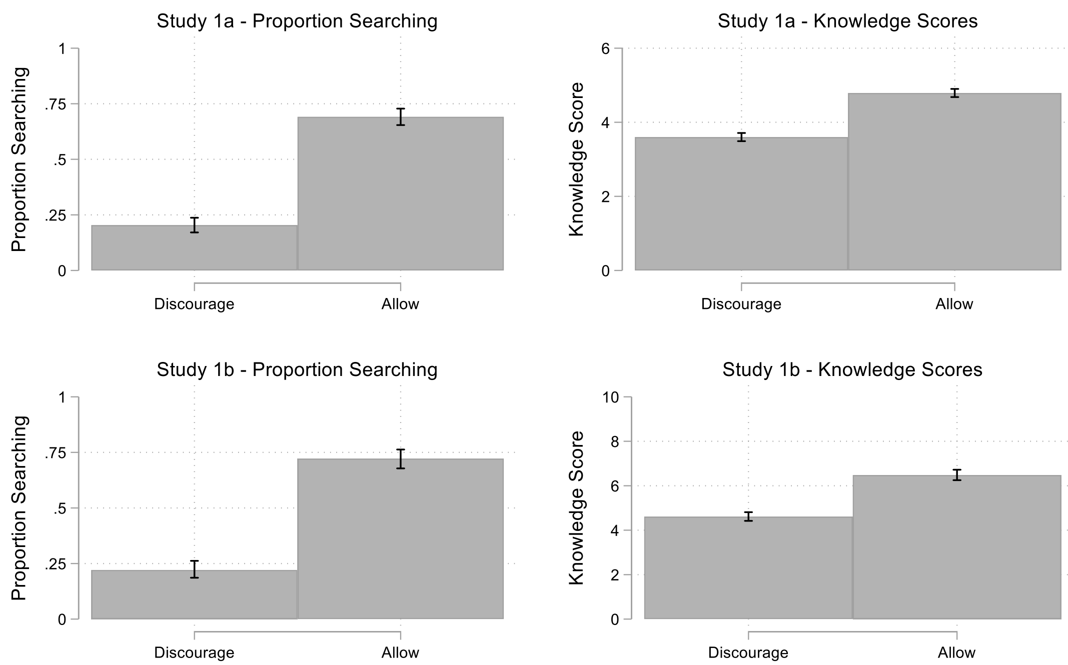
[10] Once again, the two measures are strongly related ($\alpha = .80$) and the results are similar regardless of which measure is used (Self-report: 67% vs. 16%; Catch: 63% vs. 12%).

[11] Specifically, knowledge is the dependent variables, outside search is the independent variable, and random assignment to condition is the instrumental variable. This approach calculates the average gain in knowledge scores from searching among compliers.

[12] There is a noticeable difference across conditions in the likelihood of ceiling effects. In Study 1a, for example, 44% of respondents received the maximum score on the knowledge scale in the allow condition, whereas only 9% of respondents did so in the discourage condition. In Study 1b,

Across the 16 individual knowledge items across the two studies, average treatment effects ranged from 2 to 38 percentage points. A naive interpretation is that certain questions are more prone to search, which would have implications for scale design. However, the percentage of respondents correctly answering a question in the discourage condition is strongly negatively related to the size of the treatment effect for that item (r = -.67, p = .005, n = 16; see Appendix for details). In other words, people are more likely to search when they don't know the answer.

Figure 1. Search Rates and Knowledge Scores by Experimental Condition



Convergent and Discriminant Validity

Having shown that the manipulation had the intended effect, we turn to the convergent and discriminant validity of knowledge scales across conditions. Political interest is strongly

---

where the scale consisted of more items, ceiling effects were minimal: 10% of respondents received the maximum score in the allow condition and 2% did so in the discourage condition.

linked with political knowledge from both a conceptual and empirical standpoint (Delli Carpini and Keeter 1996; Prior 2010), so we use this measure as our test of convergent validity (i.e., there should be a strong positive relationship between interest and knowledge). Our measure of survey effort will be used in a test of discriminant validity, and in that test, a *weak or null* relationship between survey effort and political knowledge indicates higher discriminant validity.

To test these expectations, we utilize data from Study 1a and the first wave of Study 1b. For each study, we use an OLS model to predict political knowledge as a function of political interest, survey effort, treatment assignment, and interactions between treatment assignment and each of the other two covariates.[13] All variables are standardized for analysis. Figure 2 displays the coefficients for political interest (top row) and survey effort (bottom row) in the discourage (left panels) and allow conditions (middle panels). The right-most panels display the coefficients for the interaction terms. Within each panel, effects are shown for Study 1a, Study 1b, and a model pooling both studies. Full model details are provided in Table A2 in the Appendix.

As expected, political interest is a positive and significant predictor of political knowledge in the discourage condition in both studies (Study 1a: $b = .28$, $p < .001$; Study 1b: $b = .29$, $p < .001$). However, in both studies, the effect of interest is smaller in the allow condition (Study 1a: $b = .13$, $p < .001$; Study 1b: $b = .19$, $p < .001$). The right-hand panel of Figure 2 displays the interaction terms, showing that the effect of political interest consistently is larger in the discourage condition than in the allow condition (Study 1a: $p = .004$; Study 1b: $p = .079$;

[13] We also used IRT models to estimate latent knowledge scores separately in each experimental condition (see Gooch 2015). However, latent scores were extremely highly correlated with additive scores (rs > .96) and produced the same substantive findings.

Pooled: $p = .001$). These findings also show up in bivariate relationships, as survey effort is only weakly related to political interest (Pooled: $r = .09$).[14]

In our test of discriminant validity, we expect a weak or null relationship between survey effort and political knowledge. Starting with the discourage condition, we find that survey effort is not significantly related to political knowledge in Study 1a ($b = .03$, $p = .386$), Study 1b ($b = .02$, $p = .570$), or in the pooled data ($b = .03$, $p = .306$). This pattern changes when respondents are allowed to search. Survey effort is positively and significantly related to knowledge in the allow condition in both Study 1a ($b = .36$, $p < .001$) and Study 1b ($b = .27$, $p < .001$). Moreover, the effects of survey effort are significantly larger in the allow condition than in the discourage condition (ps < .001), suggesting that knowledge scores become confounded with survey effort when search is allowed. These patterns also appear in bivariate relationships.[15] The contrast in the effect of survey effort across experimental conditions is particularly stark. In contrast to scholars who argue that traditional measures of knowledge are flawed because people exert different levels of effort while searching their memories (Prior and Lupia 2008, 170), Studies 1a and 1b demonstrate that effort is confounded with knowledge only when respondents are allowed to engage in outside search.[16]

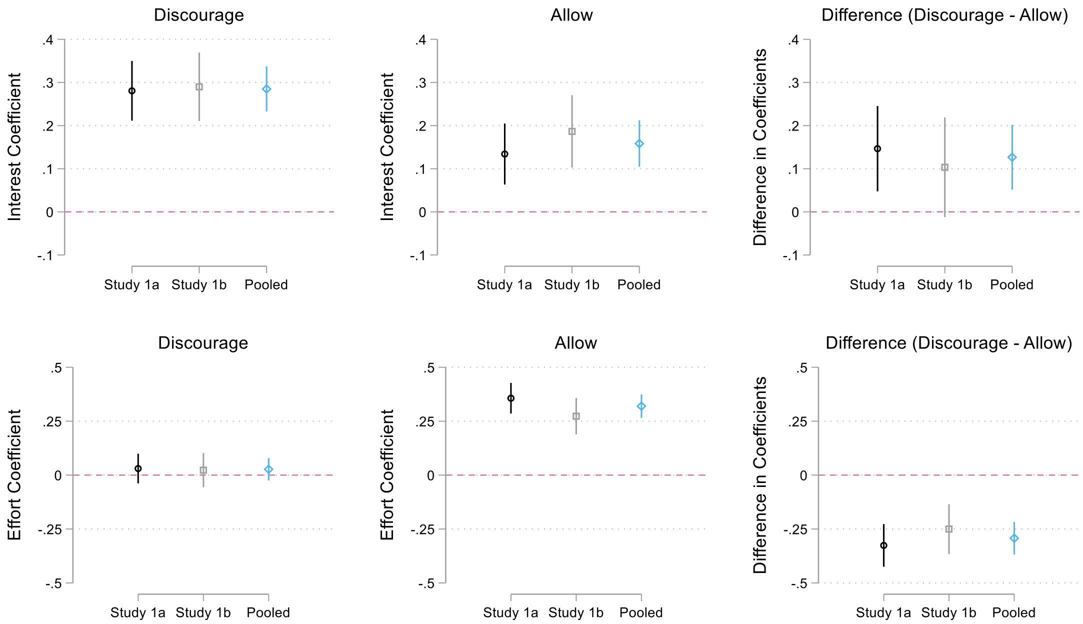
---

[14] In the pooled dataset, the bivariate correlation between knowledge and political interest is $r = .33$ in the discourage condition and $r = .20$ in the allow condition.

[15] In the pooled dataset, the bivariate correlation between knowledge and survey effort is $r = .06$ in the discourage condition and $r = .34$ in the allow condition.

[16] It is unclear whether the same pattern would occur in a control condition that does not specify whether search is permissible. However, past research suggests that some respondents interpret a

Figure 2. Convergent and Discriminant Validity of Political Knowledge



Note: Plot displays coefficients from an OLS model predicting knowledge scores. Right-hand panels display interaction terms. Bars are 95% confidence intervals. See Table A2 in the appendix for full model details.

Overall, Studies 1a and 1b demonstrate that when outside search is discouraged, knowledge reflects a person's interest in politics and is unrelated to survey effort. When search is permitted a person's level of political knowledge is highly dependent on the amount of effort he or she puts into the questionnaire, and less strongly related to political interest. Allowing outside search does not merely weaken the relationship between knowledge and criterion outcomes: it introduces a confounding relationship between knowledge and effort.

Predictive Validity

Using the additional items from Study 1b, we conduct tests of predictive validity. For each outcome, we estimate an OLS model predicting the criterion variable as a function of the

---

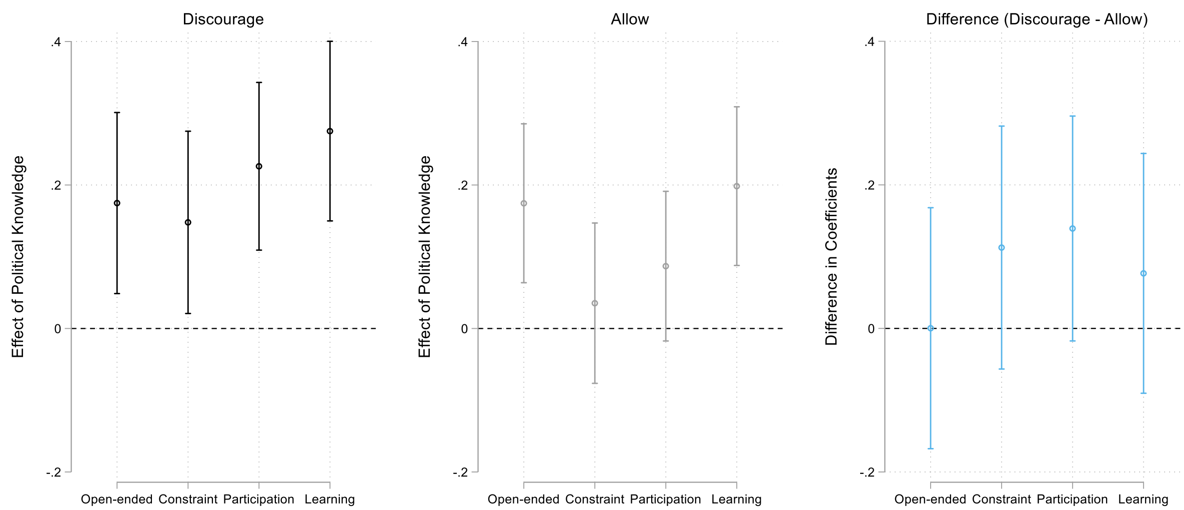
control condition as allowing search, and our results imply this will introduce a confound for these individuals.

respondent's political knowledge score, the experimental condition assigned, and an interaction between the two. Some of our outcomes included measures in both waves (e.g., open-ended comments, ideological constraint). For these outcomes, we focus on measures that are combined across waves, and show the results separately in the Appendix. We present the effects graphically below in Figure 3. Full model results are shown in Table A3 in the Appendix.

The left-hand panel of Figure 3 shows the results when search is discouraged. Starting on the left, political knowledge is a significant predictor of more elaborated open-ended responses (b = 0.17, p = .007). Moving to the right, political knowledge is a significant predictor of more constrained issue attitudes across the two waves (b = .15, p = .023). Political knowledge is also a significant predictor of political engagement (b = .23, p < .001), such as voter turnout and campaign participation. And finally, knowledge predicts better comprehension of the news article on voter ID laws (b = .28, p < .001). Overall, when outside search is discouraged, political knowledge is a consistent predictor of theoretically-linked attitudes and behaviors.

The middle panel presents analogous results when information search is allowed. Political knowledge is still a significant predictor of the complexity of open-ended responses (b = .17, p = .002) and news comprehension (b = .20, p < .001), but not ideological constraint (b = .04, p = .536), or political engagement (b = .09, p = .103). Thus, political knowledge loses much of its predictive validity when information search is allowed. The difference between the effects of political knowledge across conditions is shown in the right-hand panel of Figure 3, and here we see that the two sets of coefficients are not statistically distinguishable from each other (p = .997, p = .192, p = .082, p = .368, respectively). Although the difference between the two sets of coefficients is statistically uncertain, the pattern is clear: political knowledge is a consistent predictor of theoretically-linked outcomes when search is discouraged (average b = .21), but these effects are less clear when respondents are permitted to search (average *b* = .13). Search engine use allows people to appear politically knowledgeable without having the benefits (constraint, political participation, etc.) that accrue from sustained intellectual and cognitive engagement with public affairs.

Figure 3. Predictive Validity of Political Knowledge by Experimental Condition



Note: Plot displays knowledge coefficients from a series of OLS models predicting each criterion variable. Bars are 95% confidence intervals. See Table A3 in the Appendix for full model details.

Across the two experiments, there was strong evidence that knowledge scales have worse convergent and discriminant validity when outside search is permitted, and suggestive evidence that the predictive validity declines. In our next study, we observe people as they naturally complete online survey questionnaires and examine the relationship between outside search and scale validity.

Study 2: Observational Data and Measures

Subjects for Study 2 were recruited through Survey Sampling International (SSI) to participate in a four-wave panel study fielded during the 2016 US general election.[17] SSI participants are part of a non-random online sample, but are recruited from multiple diverse sources and participate in many different kinds of surveys. Wave 1 took place July 1-18 and included an initial sample of 3,552 US resident adults. Subsequent waves took place September 10-16, October 20-29, and November 7-10. Each wave had a planned attrition rate; all Wave 1 participants were recontacted for subsequent waves, but each wave was closed after reaching the desired participation number. Thus, Wave 2 includes 2,024 participants, Wave 3 1,234, and Wave 4 1,730.[18]

Political knowledge was measured with an index of four multiple-choice questions in the Wave 1 survey ($\alpha = .58$). The four items asked respondents to identify the job or political office held by Paul Ryan, the political party in control of the House of Representatives, the length of a U.S. Senate term, and who nominates judges to the Federal Courts (again, with no "don't know"

[17] Data collection was conducted and funded by the Center for the Study of Political Psychology at the University of Minnesota, and approved by the University of Minnesota Institutional Review Board, #1605E88143.

[18] Respondents were offered a higher incentive for participating in Wave 4 than for participating in prior waves. Respondents were recontacted for each wave even if they only completed Wave 1; however, no new respondents were added to the sample after Wave 1. In the analyses below, responses are weighted to approximate the demographics of the US population. Unweighted analyses produce substantively similar results.

option). Participants were told to answer knowledge questions "to the best of your ability" and were instructed to *not* look up answers online (Motta, Callaghan, and Smith 2016; Vezzoni and Ladini 2017). Search engine use was measured with a difficult catch question, measured immediately after the knowledge items.[19] We assume that anyone providing the correct response defied the instructions regarding outside search (i.e., we consider these respondents "cheaters"; all other respondents are "non-cheaters").

We assess the convergent validity of political knowledge between cheaters and non-cheaters using a three-item political interest scale. The three items (how interested the respondent is in politics, how much they follow campaigns, how much they care who wins the presidential election) were measured at Wave 1 and form a reliable scale ($\alpha = .83$). To test predictive validity, we measured political engagement and ideological constraint. Engagement was measured using a scale of turnout intention (measured at Waves 2 and 3) and reported turnout (measured at Wave 4). Turnout intention and reported turnout were placed on a 0-1 scale and then averaged to create a more reliable over-time measure of political engagement ($\alpha = .76$).[20] Ideological constraint was measured at Wave 1 by asking participants' positions on 12 political issues. Issue attitudes were recoded to range from liberal to conservative, averaged ($\alpha = .81$), and the scale was folded at the

---

[19] The question asked "In what year did the United States Supreme Court decide the case Geer v. Connecticut?"

[20] Due to the planned-attrition design of the survey, only 624 participants completed all four waves of the panel. Participants are included in the scale if they answered at least two of the three turnout measures.

midpoint. Finally, we measure survey effort by creating a latent measure based on two indicators of satisficing: the incidence of straight-lining and survey duration.[21]

The analyses differ from those in Study 1 in two ways. First, due to the observational nature of the data, we control for several factors (all measured in Wave 1) which could potentially confound the relationship between political knowledge and our criterion variables. Second, given that the data in Study 2 approximate a nationally representative sample when weighted, we examine the individual-level factors that are associated with cheating. In addition to asking about demographic characteristics, several personality traits were measured in the survey: openness to experience (desire for information and engagement), agreeableness (warmth and sympathy for others), conscientiousness (dependability and dutifulness), and need for closure (desire for unambiguous information) (Gosling, Rentfrow, and Swann 2003; Webster and Kruglanski 1994).

Study 2: Observational Results

Eleven percent of respondents in Study 2 sought outside assistance on the catch question, despite having been instructed not to look up answers. On average, people who cheated on the catch question had a significantly higher score on the knowledge scale than non-cheaters (3.04

[21] Wave 1 included 18 response grids with four items or more. Straight-lining is measured as the count of these grids that had no variance. Survey duration measures the amount of time respondents took between opening and submitting Wave 1 (not including time spent on the knowledge battery). Survey duration is coded into quintiles to combine with straight-lining and to create the latent measure of survey effort.

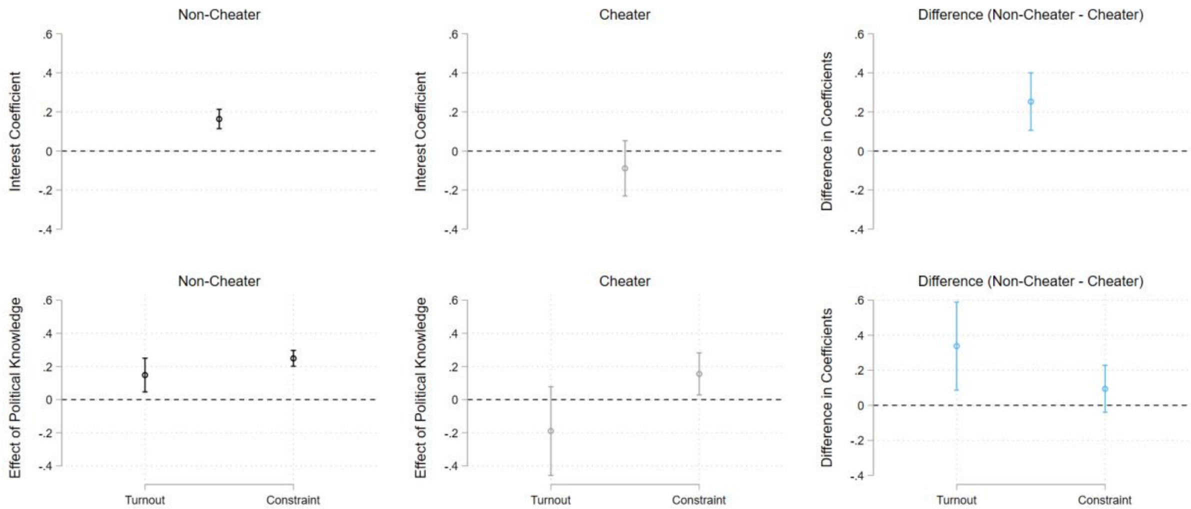
vs. 2.60; $t(3527) = -6.90$, $p < .001$). Participants who cheated also spent about twice as long on the knowledge battery than non-cheaters: three minutes compared to 90 seconds ($t(3517) = -3.32$, $p < .001$). Once again, we also see an increase in ceiling effects, which is exacerbated by the shorter knowledge scale. Nearly half (46%) of cheaters answered all four questions correctly, while only 28% of non-cheaters did so. Cheating effects on individual questions ranged from 8 to 22 percentage points, with the effect of cheating largest on the most difficult questions.

Taking advantage of our national sample, we sought to explore which respondents engaged in outside search. We estimated a logistic regression predicting cheating as a function of several Big Five personality traits, need for closure, and basic demographic variables (full model results shown in the Appendix). Consistent with the expectation that those who are most comfortable with technology would be more likely to cheat, younger respondents were significantly more likely to look up answers ($p < .001$). We also found that respondents higher in Openness were more likely to cheat ($p = .023$). Overall, these results illustrate that search engine use is not a random behavior. The fact that outside search is concentrated among certain types of respondents underlines our claim that measures of knowledge become confounded with other factors when respondents look up answers.

Turning to our main tests of validity, we expected a weaker association between political knowledge and criterion outcomes among cheaters than non-cheaters. We test convergent validity using an OLS model to predict political knowledge as a function of the interaction between political interest and cheating behavior, while controlling for survey effort, and the demographic and personality variables discussed above. We test predictive validity using OLS models to predict turnout and ideological constraint as a function of the interaction between political knowledge and cheating behavior, as well as the controls. Key outcomes and predictors (knowledge, interest, turnout, and constraint) are standardized for ease of comparison. The results are reported in Figure 4 (full model details are shown in Tables A6 and A7 in the Appendix).

In the top row of Figure 4, political interest is a positive and significant predictor of political knowledge among non-cheaters ($b = .16$, $p < .001$) but is not a significant predictor of political knowledge among cheaters ($b = -.09$, $p = .217$). This difference between cheaters and non-cheaters is large and statistically significant ($p = .001$; top right-hand panel). When it comes to predictive validity (bottom row of Figure 4), political knowledge is also a weaker predictor of turnout among cheaters than non-cheaters. As expected, among non-cheaters higher political knowledge predicts higher scores on the turnout scale ($b = .15$, $p = .004$). Among cheaters, however, political knowledge does not significantly predict turnout ($b = -.19$, $p = .165$). As shown in the bottom right-hand panel of Figure 4, this difference between cheaters and non-cheaters is large and statistically significant ($p = .008$). Finally, political knowledge is a somewhat stronger predictor of ideological constraint among non-cheaters ($b = .25$, $p < .001$) than cheaters ($b = .16$, $p = .017$), although this difference is not statistically significant ($p = .170$). Consistent with Studies 1a and 1b, political knowledge is a strong predictor of theoretically relevant outcomes among respondents who did not look up answers. However, political knowledge loses its ability to predict turnout among those who looked up answers. Overall, the findings from Study 2 are similar to those of Studies 1a and 1b despite the different methodological approach.

Figure 4. Convergent and Predictive Validity of Political Knowledge by Cheating Behavior



Note: Upper row displays interest coefficients from OLS model predicting knowledge scores. Bottom row displays knowledge coefficients from a series of OLS models predicting each criterion variable. Right-hand panels display interaction terms. Bars are 95% confidence intervals. See Tables A6 and A7 in the appendix for full model details.

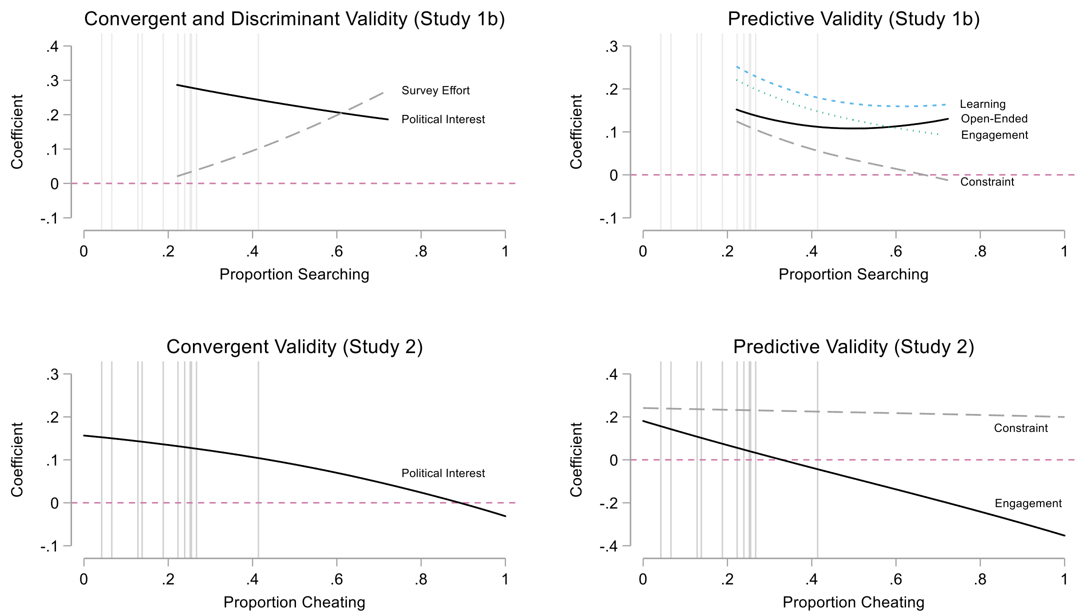
How Concerned Should We Be About Search Engine Use?

Studies 1 and 2 have shown that when respondents search for the answers to political knowledge questions, the resulting scales suffer from weakened convergent, discriminant, and predictive validity of knowledge measures. Yet, search rates are highly variable across studies, with published estimates ranging from 4% to 41% of the sample (Clifford and Jerit 2016). This variation makes it difficult to know when search behavior is a problem. Here we explore the issue through a series of simulations. Drawing on Study 1b, we randomly resample respondents from each treatment condition at different proportions and re-estimate naïve versions of the models reported above. That is, we exclude indicators of experimental condition (and interaction terms) and test the validity of the measure among the full sample. We simulate different search rates by shifting the proportion of respondents drawn from the discourage condition from 100% down to 0% while shifting the proportion of respondents drawn from the allow condition from

0% to 100%. At each percentile, we resampled with replacement, recorded the model results, and repeated this process 1,000 times. Figure 5 plots a LOWESS curve through the mean coefficient at each level of cheating. Because the search rate was 22% in the discourage condition and 72% in the allow condition, we can only estimate our models within this range of cheating.

The top left panel shows the effects of political interest and survey effort on knowledge scores as the rate of outside searching increases. Vertical lines represent published cheating rates from previous studies that did not attempt to control cheating, and thus one may interpret these rates as the prevalence of cheating at “baseline” levels. The far left side of the graph represents a model estimated only from respondents in the discourage condition, corresponding with a search rate of 22%. Here, interest is a strong predictor, while survey effort is not. However, as the frequency of outside search increases, these patterns reverse and interest becomes a weaker predictor, while survey effort rapidly becomes a stronger predictor. Naturally, these patterns would likely be more dramatic if we could estimate these coefficients at a search rate of zero. The top right panel shows the coefficients for political knowledge from each of our predictive validity models. As outside search increases, the coefficient on knowledge decreases steadily in all models with the exception of the open-ended comments. The drop in validity is particularly steep at lower levels of search (e.g., between .2 and .4), suggesting that search engine use may be significantly undermining the predictive validity of our knowledge measure even in the discourage condition.

Figure 5. Simulated Validity Across Cheating Rates



Note: Each panel displays the estimated criterion validity of political knowledge. In the top row, the proportion searching is varied by resampling at different rates from the experimental condition. In the bottom row, the proportion cheating is varied by resampling at different rates from the cheating variable.

We took the same approach with Study 2. Because Study 2 is observational, we resample from cheaters and non-cheaters at varying proportions while holding sample size constant, and then estimate each model with the full set of control variables described above (excluding the indicator of cheating and associated interaction terms). The observational nature of the study means we must make an assumption about the similarity of cheaters and non-cheaters, conditional on the control variables. However, our observational and experimental findings converged quite well, and this simulation allows us to estimate validity across the full range of cheating rates. The bottom left panel displays the coefficients for political interest. While political interest positively predicts political knowledge at low rates of cheating, the size of the coefficient drops as cheating rates increase, and reaches zero when approximately three-quarters of the sample is cheating. The bottom right panel displays the coefficients from our predictive validity models. Political knowledge is a strong predictor of political engagement, but this effect steadily drops as cheating rates increase. In fact, at a cheating rate of approximately 32%, political knowledge is unrelated to engagement—a rate that is lower than some rates of search behavior that have been observed in studies that did not attempt to control cheating. Turning to ideological constraint, knowledge remains a positive predictor across cheating rates. Taken together, these simulations demonstrate that the convergent, discriminant, and predictive validity of political knowledge measures consistently declines as search rates increase.

From a measurement standpoint, the decrease in validity is concerning (Mondak 2001), but there are broader implications given how frequently measures of knowledge are used in empirical research. For example, one of the most robust patterns in the study of political knowledge is the difference in observed levels of knowledge across men and women, with the latter having lower levels of observed knowledge. It is natural to wonder whether search engine use affects this empirical regularity, and if so, how. Past research has shown that when information is made available to respondents, the gender gap in knowledge diminishes and is sometimes eliminated (Jerit and Barabas 2017). In addition, women are more willing and attentive survey respondents (e.g., Berinsky, Margolis, and Sances 2013), and thus may be more likely to search for information when allowed. Both patterns imply that we would replicate the gender gap in knowledge when search is discouraged, but observe a smaller gap when it is allowed.

We explore this question with the pooled data from Study 1a and 1b. When search is discouraged, there are no differences in rates of cheating across women and men (Women: 23%, Men: 19%; $\chi^2(1) = 2.25$, $p = .134$). However, when outside search is allowed, women are significantly more likely to look up answers (Women: 74%, Men: 66%; $\chi^2(1) = 7.08$, $p = .008$). This is due, in part, to higher survey effort among women ($t(2{,}055) = 3.95$, $p < .001$). Turning to knowledge scores, when search engine use is discouraged, we replicate the gender gap in knowledge with women scoring 0.18 standard deviations lower than men (t(1,034) = 3.26, p = .001). However, when search is allowed, this reliable difference evaporates, with women now scoring a non-significant 0.03 standard deviations higher than men (t(1,011) = 0.47, p = .641). Moreover, a regression model shows a significant interaction between gender and the experimental condition (p = .012) demonstrating that the gender gap is significantly smaller when information search is allowed. This pattern of results underscores our claim that knowledge scales measure something different when search is allowed versus when it is discouraged.

A skeptic might point out that the motivation and ability to look up answers to factual questions is a politically relevant skill. We agree. However, as shown here, this skill is *distinct* from traditional conceptions of political knowledge. Therefore, scholars interested in measuring a person's skills at searching the internet for political information would be better served by developing independent measures of search motivation and ability. Importantly, our findings highlight an obstacle to this line of research: information search is substantially confounded with the effort a respondent is willing to put into completing a survey—a factor that is only weakly related to interest in politics. Recall that in Studies 1a and 1b, the relationship between interest and effort was only r = .09 (pooled data).

Conclusion

Political knowledge is a central concept in political science. As surveys are increasingly administered online, however, respondents are able to search the web for answers, potentially altering what scholars are measuring with factual knowledge questions. A recent body of research demonstrates that outside search occurs and can affect the estimated levels of political knowledge among the public (Burnett 2016; Clifford and Jerit 2014, 2016; Motta, Callaghan, and Smith 2016; Shulman and Boster 2014), but that literature has yet to establish how search behavior affects the validity of knowledge measures. Across experimental and observational studies, we find a consistent pattern of results—namely, that search engine use reduces the validity of political knowledge measures and undermines the ability to replicate canonical findings in the public opinion literature.

Of course, the extent to which the validity of political knowledge scales is degraded will be a function of the prevalence of cheating within the sample. But this should provide little comfort to researchers. While some of the highest cheating rates have been observed in student samples, search rates in the ANES 2018 Pilot Study range from 15% to 25% (depending on questionnaire format). Thus, no online data source is immune from this problem. Moreover, the evidence shows that search rates can vary substantially *within* populations (e.g., rates on MTurk have ranged from 5% to 25%), making it difficult to know when search engine use will reach problematic rates in a study. Search behavior is more common among younger respondents (see Table A5), however, and is likely to become prevalent as respondents become more comfortable with and reliant upon technology.

The increasing ease with which respondents can look up answers poses a threat to the validity of political knowledge measures, but has particularly troubling implications for understanding the temporal dynamics of knowledge. For example, outside search could lead political scientists to prematurely conclude that the gender gap in knowledge has disappeared. Or, due to the weakening predictive power of political knowledge, researchers might conclude that knowledge is no longer a pre-requisite for an engaged, ideological citizenry in a polarized era. More generally, the likely growing rate of outside search may create a spurious time trend that could mislead researchers in a variety of ways. Understanding whether and how the "currency of citizenship" maintains its relevance, and how that resource is distributed, will require scholars to ensure the continued validity of the measure.

Considering these challenges, our first recommendation for researchers is to actively discourage outside search. Simple instructions to refrain from looking up answers can reduce search engine use (Motta, Callaghan, and Smith 2016; Vezzoni and Ladini 2017), but do not eliminate it, as shown in Study 2. A more effective tactic is to ask respondents to commit to not looking up answers, but even this approach is not perfect (Clifford and Jerit 2016). Thus, our second recommendation is to diagnose cheating through self-reports, difficult catch questions, or both. Inclusion of such measures allows researchers to identify whether cheating poses a threat to the validity of their data, as well as contribute to a better understanding of when cheating occurs (also see Deisenhofer and Musch 2017).[22] The question of how to handle cheaters is thornier, but we borrow advice from the literature on satisficing (Berinsky, Margolis, and Sances 2013). Dropping respondents from the analysis (e.g., those who engage in outside search) may harm the representativeness of the sample and should be discouraged, especially if knowledge is measured post-treatment in an experimental design (Montgomery, Nyhan, and Torres 2016). Another intuitive solution is to control for cheating in statistical models. Yet our own supplementary analyses show that controlling for cheating did little to reduce the biases we reported above.

Given that some respondents will look up answers even when instructed not to, a promising line of future research is to develop instrumentation that is immune to this problem.

[22] We recommend that the catch question be placed last among the knowledge questions. There is some evidence (Study 1b) that respondents who received the catch question first were more likely to look up answers, but this effect occurred only when search was allowed.

Longer scales will help by reducing ceiling effects, however, this change will do nothing to prevent the confounds introduced by outside search.[23] Visual political knowledge questions (Munzert and Selb 2015) and text-based approaches (Kraft n.d.) represent alternative ways for assessing knowledge, but more research is needed to determine if these formats circumvent the problems with online search. In light of the fact that cheating levels vary considerably across surveys from the same population, it would be valuable to investigate whether aspects of survey design affect the prevalence of cheating. For example, a person may be motivated to cheat after answering items about vote choice to avoid looking like an uninformed voter. Similarly, respondents may be inclined to cheat following questions about group identity, such as partisanship, in order to make their group look well-informed. Either finding would have implications for question order in survey design.

The shift towards conducting surveys online has brought about a number of advantages, such as lower costs, faster data collection, and decreased social desirability bias (e.g., Kreuter, Presser, and Tourangeau 2008). However, the online administration of surveys also comes with costs, such as higher levels of satisficing (e.g., Heerwegh 2009). Our research adds to the list of challenges faced by scholars conducting research online, demonstrating that the ability to search for information can undermine the validity of recall-based measures used by researchers throughout the social sciences.

[23] Another tempting solution is to include a time limit on factual knowledge questions. Evidence on the effectiveness of timers is equivocal (e.g. Jensen and Thomsen 2014), perhaps because of the difficulty of choosing a time limit that deters online lookup without interfering with the response process of "honest" respondents.

References

Abrajano, Marisa. 2015. "Reexamining the 'Racial Gap' in Political Knowledge." *The Journal of Politics* 77(1): 44–54.

Ahler, Douglas J., and Stephen N. Goggin. 2017. *Assessing Political Knowledge: Problems and Solutions in Online Surveys*.

Althaus, Scott L. 2003. *Collective Preferences in Democratic Politics: Opinion Surveys and the Will of the People*. Cambridge: Cambridge University Press.

Alwin, Duane F. 1991. "Family of Origin and Cohort Differences in Verbal Ability." *American Sociological Review* 56(5): 625.

Ansolabehere, Stephen, Jonathan Rodden, and James M. Snyder. 2008. "The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting." *American Political Science Review* 102(02): 215–32.

Berinsky, Adam J., Michele F. Margolis, and Michael W. Sances. 2013. "Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys." *American Journal of Political Science*: n/a-n/a.

Bialek, Michal, and Gordon Pennycook. 2017. "The Cognitive Reflection Test Is Robust to Multiple Exposures." *Behavior Research Methods*: 1–7.

Bullock, John G., Alan S. Gerber, Seth J. Hill, and Gregory A. Huber. 2015. "Partisan Bias in Factual Beliefs about the Economy." *Quarterly Journal of Political Science* 10: 519–78.

Burnett, Craig M. 2016. "Exploring the Difference in Participants Factual Knowledge between Online and In-Person Survey Modes." *Research & Politics* 3(2): 120–31.

Clifford, Scott, and Jennifer Jerit. 2014. "Is There a Cost to Convenience? An Experimental Comparison of Data Quality in Laboratory and Online Studies." *Journal of Experimental*

Political Science: 1–12.

———. 2016. “Cheating on Political Knowledge Questions in Online Surveys.” *Public Opinion Quarterly* 80(4): 858–87.

Delli Carpini, Michael X., and Scott Keeter. 1996. *What Americans Know about Politics and Why It Matters*. Yale University Press.

Federico, Christopher M. 2004. “Predicting Attitude Extremity: The Interactive Effects of Schema Development and the Need to Evaluate and Their Mediation by Evaluative Integration.” *Personality and Social Psychology Bulletin* 30(10): 1281–94.

Fiske, Susan T., Richard R. Lau, and Richard A. Smith. 1990. “On the Varieties and Utilities of Political Expertise.” *Social Cognition* 8(1): 31–48.

Gibson, James L., and Gregory A. Caldeira. 2015. “Knowing the Supreme Court? A Reconsideration of Public Ignorance of the High Court.” *https://doi.org/10.1017/S0022381609090379*.

Gooch, Andrew. 2015. “Measurements of Cognitive Skill by Survey Mode: Marginal Differences and Scaling Similarities.” *Research & Politics* 2(3): 205316801559068.

Gosling, Samuel D, Peter J Rentfrow, and William B Swann. 2003. “A Very Brief Measure of the Big-Five Personality Domains.” *Journal of Research in Personality* 37(6): 504–28.

Heerwegh, D. 2009. “Mode Differences Between Face-to-Face and Web Surveys: An Experimental Investigation of Data Quality and Social Desirability Effects.” *International Journal of Public Opinion Research* 21(1): 111–21.

Hillygus, D. Sunshine, Natalie Jackson, and Mckenzie Young. 2014. “Professional Respondents in Non-Probability Online Panels.” In *Online Panel Research - A Data Quality Perspective,* eds. Mario Callegaro et al. Wiley, 219–36.

Jacoby, William G. 2006. “Value Choices and American Public Opinion.” *American Journal of Political Science* 50(3): 706–23.

Jensen, Carsten, and Jens Peter Frølund Thomsen. 2014. “Self-Reported Cheating in Web Surveys on Political Knowledge.” *Quality & Quantity* 48(6): 3343–54.

Jerit, Jennifer, and Jason Barabas. 2017. “Revisiting the Gender Gap in Political Knowledge.” *Political Behavior* 39(4): 817–38.

Kahan, Dan M. 2016. “‘Ordinary Science Intelligence’: A Science-Comprehension Measure for Study of Risk and Science Communication, with Notes on Evolution and Climate Change.” *Journal of Risk Research*: 1–22.

Kam, Cindy D. 2005. “Who Toes the Party Line? Cues, Values, and Individual Differences.” *Political Behavior* 27(2): 163–82.

Kam, Cindy D., and Marc J. Trussler. 2016. “At the Nexus of Observational and Experimental Research: Theory, Specification, and Analysis of Experiments with Heterogeneous Treatment Effects.” *Political Behavior*: 1–27.

Kane, John V., and Jason Barabas. 2018. “No Harm in Checking: Using Factual Manipulation Checks to Assess Attentiveness in Experiments.” *American Journal of Political Science*.

Kinder, Donald R. 2006. “Belief Systems Today.” *Critical Review* 18(1–3): 197–216.

Kinder, Donald R., and Nathan P. Kalmoe. 2017. *Neither Liberal nor Conservative : Ideological Innocence in the American Public*. Chicago: University of Chicago Press.

Kreuter, F., S. Presser, and R. Tourangeau. 2008. “Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity.” *Public Opinion Quarterly* 72(5): 847–65.

Lau, Richard R., and David P. Redlawsk. 2001. “Advantages and Disadvantages of Cognitive

Heuristics in Political Decision Making." *American Journal of Political Science* 45(4): 951.

Lizotte, Mary-Kate, and Andrew H. Sidman. 2009. "Explaining the Gender Gap in Political Knowledge." *Politics & Gender* 5(02): 127.

Lopez, Jesse, and D. Sunshine Hillygus. 2018. "Why So Serious?: Survey Trolls and Misinformation."

Luskin, Robert C., and John G. Bullock. 2011. "'Don't Know' Means 'Don't Know': DK Responses and the Public's Level of Political Knowledge." *The Journal of Politics* 73(2): 547–57.

Mason, Lilliana. 2018. "Ideologues without Issues: The Polarizing Consequences of Ideological Identities." *Public Opinion Quarterly* 82(S1): 280–301.

Miller, M. K., and S. K. Orr. 2008. "Experimenting with a 'Third Way' in Political Knowledge Estimation." *Public Opinion Quarterly* 72(4): 768–80.

Mondak, Jeffery J., and Mary R. Anderson. 2004. "The Knowledge Gap: A Reexamination of Gender-Based Differences in Political Knowledge." *The Journal of Politics* 66(2): 492–512.

Mondak, Jeffrey J. 2001. "Developing Valid Knowledge Scales." *American Journal of Political Science* 45(1): 224.

Montgomery, Jacob M, Brendan Nyhan, and Michelle Torres. 2016. "How Conditioning on Post-Treatment Variables Can Ruin Your Experiment and What to Do about It."

Motta, Matthew P., Timothy H. Callaghan, and Brianna Smith. 2016. "Looking for Answers: Identifying Search Behavior and Improving Knowledge-Based Data Quality in Online Surveys." *International Journal of Public Opinion Research* 29(4): 575–603.

Prior, Markus, and Arthur Lupia. 2008. "Money, Time, and Political Knowledge: Distinguishing Quick Recall and Political Learning Skills." *American Journal of Political Science* 52(1):

169–83.

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(3): 177–89.

Vezzoni, Cristiano, and Riccardo Ladini. 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(03): 251–65.

Webster, Donna M., and Arie W. Kruglanski. 1994. “Individual Differences in Need for Cognitive Closure.” *Journal of Personality and Social Psychology* 67(6): 1049–62.

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