How Internet Search Undermines the Validity of Political Knowledge Measures

Brianna Smith

Scott Clifford

Jennifer Jerit

Abstract

Political knowledge is central to understanding the breadth and depth of citizens' engagement with politics. Yet, as opinion polls are increasingly conducted online, survey respondents' ability to search the web may undermine the validity of factual knowledge measures. A recent body of literature has demonstrated that this search behavior is common, even when respondents are explicitly instructed not to search for answers. However, we know little about how this behavior affects the validity of political knowledge measures. We investigate this question across a series of experimental and observational studies. The analyses provide consistent evidence that cheating degrades the convergent, discriminant, and predictive validity of political knowledge measures. Search behavior seems to be a function of the effort a respondent is willing to put into the survey, rather than their political awareness. Our findings imply that researchers conducting online surveys need to take steps to discourage and diagnose cheating behavior.

Presented at the annual meeting of the American Political Science Association, Boston, 2018.

Political knowledge is considered the "currency of citizenship" because it helps people process new information and link their values and interests to their attitudes. One of the most common ways to operationalize political knowledge 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 the measurement of this concept, 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.

Although there is some debate regarding the prevalence of "cheating" (e.g., Ansolabehere and Schaffner 2014; Gooch and Vavreck 2016), this behavior has been uncovered in a variety of subject populations with rates reaching as high as 41% (Clifford and Jerit 2016). Moreover, search behavior tends to inflate aggregate levels of knowledge (Burnett 2016; Clifford and Jerit 2014; Motta, Callaghan, and Smith 2017; Shulman and Boster 2014), even in high quality surveys such as the American National Election Study (Kraft n.d.).[1] Existing research has focused on the detection and prevention of cheating (Berinsky, Huber, and Lenz 2012; Clifford and Jerit 2016; Motta, Callaghan, and Smith 2017). Remarkably, however, the key issue—

[1] In this context, cheating may not be motivated by respondent dishonesty. We employ the term because it is more frequently used in the literature than phrases such as "search-engine use."

namely, the *validity* of knowledge measures when outside search is permitted—has gone unexamined. If knowledge scales reproduce canonical findings when internet search is allowed, then concerns about cheating diminish. Yet there is scant evidence on this point.

We contribute to the literature with three studies exploring the measurement properties of knowledge scales collected in online settings. In the first and second studies, we examine the validity of knowledge measures in an experiment where people are either discouraged from seeking outside assistance on factual questions or told that it is acceptable to do so. As a result, we can draw conclusions about the causal impact of cheating on the validity of knowledge measures. In the third study, we examine cheating as it naturally occurs in a national online sample of adults and conduct a similar analysis of the measurement properties of knowledge scales. Our findings suggest that the ability to recall information is strongly tied to theoretically-linked concepts such as political interest, while the ability to find information is related to the effort a person devotes to answering survey questions. Consequently, when respondents can seek outside assistance, knowledge scales have weaker associations with criterion outcomes such as political engagement and ideological constraint. We recommend that scholars who are interested in measuring political knowledge take efforts to minimize and diagnose cheating behavior. Importantly, this advice can be applied in other contexts such as the assessment of science knowledge (Kahan 2016), verbal knowledge (Alwin 1991), and reasoning ability (Bialek and Pennycook 2017).[2]

---

[2] For example, the General Social Survey uses a ten-item measure of verbal intelligence and, at present, the administrators keep the specific questions confidential in order to avoid invalidation of the items via cheating (see Cor et al. 2012).

What Knowledge Scales Measure, With and Without Cheating

There "is compelling evidence 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 the literature on "Don't Know" responses. As Mondak (2001) and others (Lizotte and Sidman

[3] Kam and Trussler (2016: 791) make a parallel argument regarding the ability of researchers to manipulate this latent 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."

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, competitiveness, 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. Cheating 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.

But therein lies the rub: if search behavior is not solely a function of a person’s engagement with politics, then cheating will degrade the measurement properties of knowledge scales by adding bias and noise. In the analyses that follow, we explore this possibility in terms of convergent, discriminant and predictive validity.[4] As a test of convergent validity, we examine the relationship between interest and political knowledge. The literature has shown that a person’s level of interest in and attention to politics is the strongest predictor of his or her level

[4] Convergent validity pertains to whether a measure is related to conceptually-related variables. Discriminant validity assesses whether measures that should *not* be related to one another in fact show no relationship. Finally, predictive validity concerns the extent to which a measure is causally related to outcomes that are specified by the existing literature.

of knowledge (Delli Carpini and Keeter 1996). Thus, we examine whether cheating weakens the relationship between knowledge and political interest. As a test of discriminant validity, we investigate whether cheating causes political knowledge scores to become confounded with the effort a respondent is willing to put into the survey. Finally, we examine the predictive power of knowledge scales in areas where the existing literature leads us to expect positive relationships between political awareness and specific outcomes, such as political participation, attitude constraint, complexity of open-ended responses, and news comprehension. We determine if cheating compromises the predictive power of political knowledge scales in explaining those outcomes. Taken together, our analyses provide some of the first systematic evidence regarding the effect of search behavior on the validity of knowledge scales.

Empirical Evidence

We explore how cheating affects the measurement properties of knowledge scales with three empirical studies. Studies 1 and 2 investigate whether cheating reduces the validity of knowledge measures in experiments where people were either discouraged from looking up the answers to factual knowledge questions or told they were allowed to do so. In Study 3, we examine cheating behavior in a naturally occurring survey environment (i.e., an online survey completed at the respondent's discretion) and analyze the relationship between political knowledge and various criterion measures among cheaters and non-cheaters.

Study 1: Experimental Data and Measures

The purpose of Study 1 was to examine the convergent and discriminant validity of knowledge scales when search is allowed. In the spring of 2017 we embedded an experiment in a survey of 1,170 undergraduates from a large public university in the southern United States.[5] After completing sections for unrelated studies, respondents were randomly assigned to one of two conditions that featured different instructions for answering the political knowledge questions. In the *discourage condition*, respondents were told that it was important to the researchers that they do not "use outside sources like the Internet to search for the correct answer," and asked whether they would commit to answering the questions without help. 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 were told that it was acceptable "to use the Internet to double check their answer or look for the correct response" if they did not already know it. Respondents then were asked six political knowledge questions. Four items pertained to public officials, including one on Trump's recent nominee for the Supreme Court. The remaining two questions asked about partisan control of the Senate and long-standing partisan symbols.[6]

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 among the battery of political knowledge questions in random order (Motta, Callaghan, and Smith 2017). The catch question asked about the year of an obscure court case, and respondents who provided the correct year were assumed to have looked

---

[5] The study consisted of an approximately 20-minute online survey. It was approved by the Committee for the Protection of Human Subjects at <university>.

[6] The study also included four additional knowledge questions that are prone to partisan bias. These items were included for the purpose of another study and are analyzed elsewhere.

up the answer. The response distribution to both items indicates that our instruction sets had the intended effect on search behavior (details reported below).

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 ($\alpha = .81$). To explore discriminant validity, we estimated latent survey effort as a function of several common indicators of satisficing, including an instructed response, an Instructional Manipulation Check (IMC), a count of straight-lining in grids, time spent on the survey, and self-reported survey effort (Berinsky, Margolis, and Sances 2013; Hillygus, Jackson, and Young 2014; Lopez and Hillygus 2018). These items 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 (see Appendix for details).

Study 1: Experimental Results

As expected, the instructions had a sizable effect on cheating rates. In the allow condition, 69% of respondents cheated according to at least one of the measures. In the discourage condition, the corresponding figure was only 20% ($\chi^2(1) = 279.89$, $p < .001$).[7] Much as one might expect, respondents in the allow condition spent more time answering the political

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

knowledge questions ($t(1168) = 11.08$, $p < .001$).[8] The median respondent in the allow condition spent an average of 20 seconds on each questions, while the median respondent in the discourage condition spent an average of 13 seconds. Knowledge scores also were higher in the allow condition (M = 4.8) than in the discourage condition (M = 3.6; $t(1160) = 15.01$, $p < .0001$). Finally, there is a noticeable difference across conditions in the likelihood of ceiling effects: while 44% of respondents received the maximum score (6) on the knowledge scale in the allow condition, only 9% of respondents did so in the discourage condition.

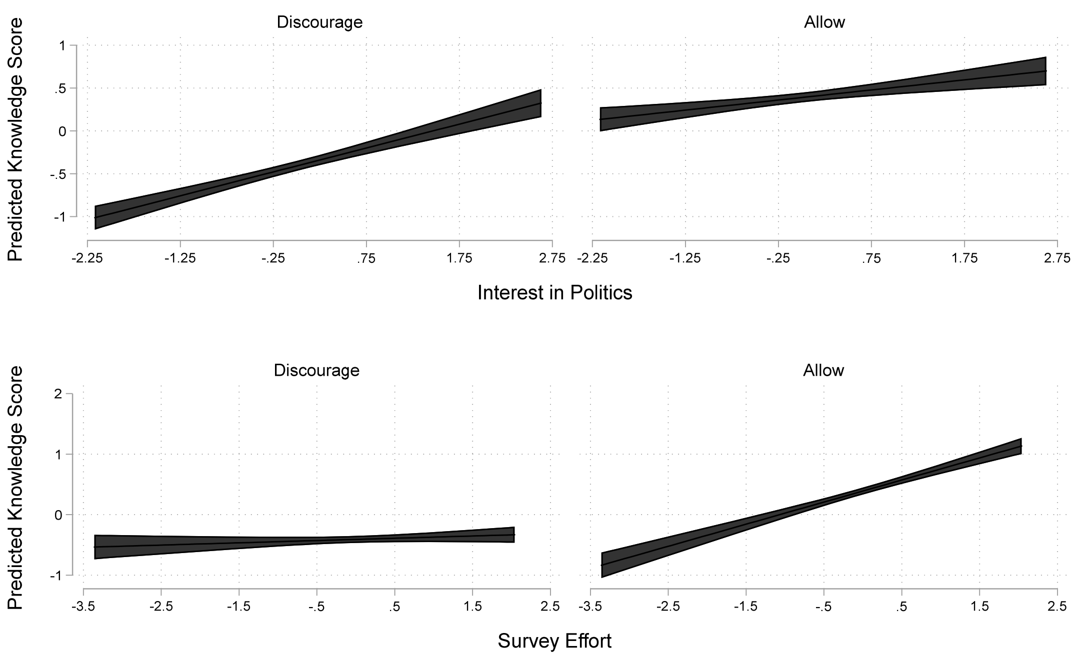
Having shown that the manipulation had the intended effect, we turn to the convergent and discriminant validity of knowledge scales across the two conditions. Political interest and attention to the news is strongly linked with political knowledge from both a conceptual and empirical standpoint (Delli 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. As others have observed, survey respondents may be unwilling to put in the effort to fully recall and report political facts they have stored in long-term memory, creating a confound between survey effort and measures of political knowledge (Prior and Lupia 2008, 170). Thus, a *weak or null* relationship between survey effort and political knowledge indicates higher discriminant validity.

To test these expectations, 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. All variables are standardized for

[8] The difference of means test is conducted on a logged measure of time due to the skewed nature of the data.

analysis. Figure 1 displays the estimated relationships between political knowledge and political interest (top row) and survey effort (bottom row), with full model details provided in Table A2 in the Appendix. As shown in the top row of Figure 1, political interest is a positive, significant predictor of political knowledge scores in both the discourage (b = .28, p < .001) and allow (b = .12, p = .001) conditions. However, interest is more strongly related to knowledge scores when information search is discouraged (p = .002). A different pattern emerges in the bottom panels. Survey effort is unrelated to political knowledge in the discourage condition (b = .04, p = .287), but it is strongly related to knowledge in the allow condition (b = .36, p < .001). The difference between the two coefficients is statistically significant (p < .001).[9]

Figure 1. Validity of Political Knowledge by Experimental Condition



[9] The bivariate correlation between knowledge and political interest was r = .32 in the discourage condition and r = .20 in the allow condition. The correlation between knowledge and survey effort was r = .09 in the discourage condition and r = .41 in the allow condition.

Overall, Study 1 demonstrated that when cheating is discouraged, knowledge reflects a person's interest in politics and is unrelated to survey effort. When cheating is permitted (allow condition) 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. Thus, measures of knowledge display better convergent and discriminant validity when cheating is discouraged. Given the significant differences in mean levels of knowledge across conditions, knowledge scales will be less accurate, in a descriptive sense, when outside search is permitted (e.g., see Luskin and Bullock 2011). Furthermore, the greater likelihood of ceiling effects in the allow condition suggests that cheating weakens the validity of the scale (see Ahler and Goggin 2017). In the next study, we replicate these findings with a different sample and explore the predictive validity of political knowledge scores.

Study 2: Experimental Data and Measures

In Study 2, student subjects were recruited from required introductory courses at a large public university in the south. The study took place in the fall of 2017 and it consisted of a two-wave online panel survey, separated by about 14 days. Eight hundred and eighty seven students completed the first wave and 644 students completed the second wave of the study, for a panel attrition rate of 27%.[10]

---

[10] Panel attrition was unrelated to treatment assignment (Discourage: 27.7%, Allow: 27.3%%; $\chi^2(1) = 0.02$, $p = .891$). We took precautions to ensure that participants from the first study did not participate in the second study. This study was approved by the Committee for the Protection of Human Subjects at <university>.

The dependent variable consists of a ten-item political knowledge measure in the Wave 1 survey (α = .68). Following the recommendations of Barabas et al. (2014) we included questions that covered traditional civics themes (e.g., the identity of Chief Justice) and policy-specific topics (e.g., the unemployment rate). The 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). The experimental treatment consisted of the same manipulation from Study 1 (discourage versus allow), and compliance was measured similarly (self-report and catch question).

As with Study 1, convergent and divergent validity were assessed with scales measuring interest and survey effort. The former was based on two-items asking about a person's interest in politics and attention to political news (α = .76), and the latter was estimated as a latent variable using several common indicators of satisficing (see Appendix for details).[11]

For our tests of predictive validity, we measured constructs linked to political knowledge by past literature. 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

[11] To avoid post-treatment bias, we only include timers from pages that were prior to our measurement of political knowledge (Montgomery, Nyhan, and Torres 2016). As expected, latent survey effort is only weakly related to political interest in Wave 1 ($r = .10$).

create a measure of political awareness (Clifford and Jerit 2016). 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 ($\alpha = .50$; attend a meeting, put up a sign, work for a candidate, donate money). To gauge respondents' ideological constraint, we asked their positions on 11 political issues in Wave 1 and 11 issues in Wave 2. We used these questions measure ideological constraint (e.g., Mason 2018), which has previously been linked with political sophistication (e.g., 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, respondents 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. Respondents receiving the maximum score have reported strongly held attitudes from a consistent ideological viewpoint.

Finally, in Wave 2, we also asked respondents to read an article covering a recent court decision regarding voter registration laws. The article was based on an August 2017 *New York Times* story about Supreme Court action on a North Carolina voting law. 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 respondents with higher levels of political knowledge to better understand and retain information from the article (Fiske, Lau, and Smith 1990).

Study 2: Experimental Results

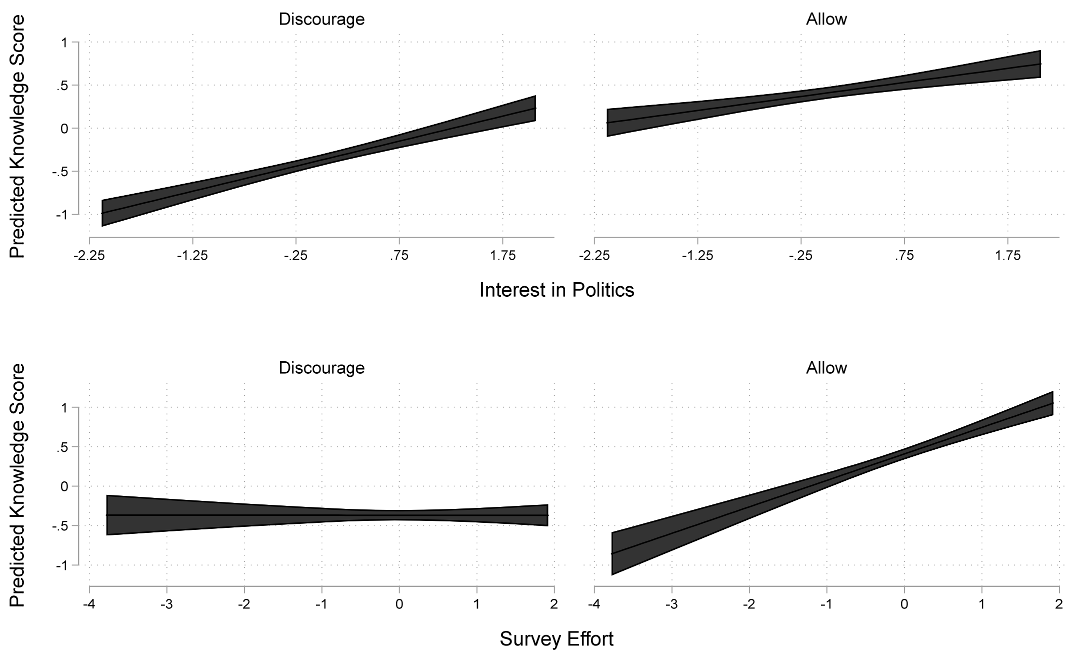
According to both manipulation checks, treatment assignment had a substantial effect on cheating behavior. In the allow condition, 72% of respondents cheated according to at least one of the measures, while 22% did so in the discourage condition ($\chi^2(1) = 223.42$, $p < .001$).[12] Respondents in the allow condition also took longer to answer the political knowledge questions than people in the discourage condition: 22 seconds versus 13 seconds ($t(885) = 10.54$, $p < .0001$). Consistent with the patterns from Study 1, the instruction manipulation affected political knowledge scores. In the allow condition, respondents answered 7.8 out of 10 questions correctly, on average, while respondents in the discourage condition, answered 5.5 out of 10 questions correctly ($t(885) = 12.15$, $p < .0001$). Despite using a relatively long knowledge battery, there is some evidence of ceiling effects: 10% of respondents received the maximum scores in the allow condition, while only 2% of respondents did so in the discourage condition.

Following the approach from Study 1, we investigated convergent and discriminant validity with an OLS model in which political knowledge was the dependent variable. The independent variables were political interest, survey effort, treatment assignment, and interactions between treatment assignment and each of the other two covariates. Figure 1 displays the estimated relationships between political knowledge and political interest (top row) and survey effort (bottom row). Full model details are shown in Table A2 in the Appendix. Once again, political interest is a positive, significant predictor of political knowledge in both the discourage ($b = .29$, $p < .001$) and allow ($b = .16$, $p < .001$) conditions, however, interest is more

[12] Once again, the two measures of cheating 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%).

strongly related to knowledge when information search is discouraged ($p = .029$). In the bottom panels, survey effort is unrelated to political knowledge in the discourage condition ($b = -.0004$, $p = .992$), but it is strongly related to knowledge in the allow condition ($b = .36$, $p < .001$). The difference between the two coefficients is statistically significant ($p < .001$). Consistent with our first study, measures of knowledge display better convergent and discriminant validity when cheating is discouraged.[13]

Figure 2. Validity of Political Knowledge by Experimental Condition



In the next series of analyses, we conduct tests of predictive validity. For each outcome, we estimate an OLS model predicting the criterion variable as a function of the respondent's political knowledge score, the experimental condition assigned, and an interaction between the

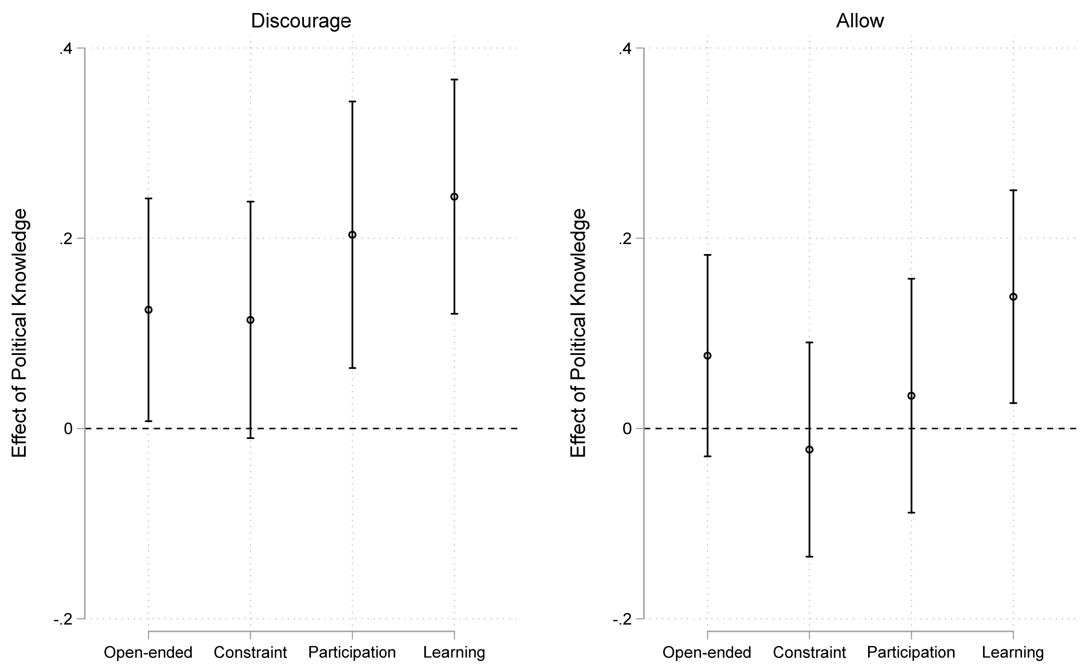
[13] The bivariate correlation between knowledge and political interest was $r = .34$ in the discourage condition and $r = .20$ in the allow condition. The correlation between knowledge and survey effort was $r = .03$ in the discourage condition and $r = .35$ in the allow condition.

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. Given that political knowledge is confounded with survey effort in the allow condition, we control for survey effort in all of the predictive validity tests below. Additionally, for any test that spans both waves of the survey, we control for survey effort in each wave. 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 cheating is discouraged. Starting on the left, political knowledge is a significant predictor of more elaborated open-ended responses ($b = 0.12$, $p = .037$). Moving to the right, political knowledge is a marginally significant predictor of more constrained issue attitudes across the two waves ($b = .11$, $p = .072$). Political knowledge is also a significant predictor of political engagement ($b = .20$, $p = .005$), such as voter turnout and campaign participation. And finally, it predicts better comprehension of the news article on voter ID laws ($b = .24$, $p < .001$). Overall, when cheating is discouraged, political knowledge is a consistent predictor of theoretically linked attitudes and behaviors.

The story changes, however, when information search is allowed, as shown in the right-hand panel of Figure 3. Political knowledge is no longer a significant predictor of the complexity of open-ended responses ($b = .08$, $p = .157$), issue attitude constraint ($b = -.02$, $p = .700$), or political engagement ($b = .03$, $p = .583$). Only in the case of news comprehension does political knowledge retain any explanatory power ($b = .14$, $p = .015$). Thus, political knowledge loses much of its predictive validity when information search is allowed. The coefficients on political knowledge are largely not statistically distinguishable from each other across conditions ($p = .548$, $p = .110$, $p = .074$, $p = .213$, respectively). However, the pattern is clear: political knowledge is a consistent predictor of theoretically linked outcomes, but only when cheating is discouraged. Cheating allows people to appear politically knowledgeable without having the benefits (constraint, political engagement, etc.) that accrue from sustained attention to politics, and that the latent disposition knowledge scales are intended to measure.

Figure 3. Predictive Validity of Political Knowledge by Experimental Condition



Across Studies 1 and 2, there was strong evidence that knowledge scales have worse convergent and divergent validity when outside search is permitted, and some evidence that the predictive validity declines. As a result of random assignment to conditions that either discouraged or allowed search (and evidence that respondents heeded those instructions), Studies 1 and 2 demonstrate the causal effect of search behavior on the validity of knowledge scales. In our third study, we observe people as they naturally complete online survey questionnaires and examine the relationship between cheating and scale validity.

Study 3: Observational Data and Measures

In Study 3, we examined the impact of naturally occurring internet search behavior. Subjects were recruited through Survey Sampling International (SSI) to participate in a four-wave panel study fielded during the 2016 US general election.[14] SSI participants are part of a non-random online sample, but are recruited from multiple diverse sources and do not exclusively participate in political 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.[15]

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. These questions are drawn from those recommended by Delli Carpini and Keeter (1996), and similar items are routinely used in the American National Election Studies (ANES) and in published political science research(Huddy, Mason, and Aarøe 2015; Jessee 2009)(Huddy, Mason, and Aarøe 2015; Jessee

[14] Data collection was conducted and funded by [redacted], and conducted under IRB #[redacted].

[15] Wave 4 was a post-election study and respondents were offered a higher incentive than the other waves. Respondents were recontacted even if they only completed Wave 1. However, no new respondents were added to the sample after Wave 1.

2009)(Huddy, Mason, and Aarøe 2015; Jessee 2009)(Huddy, Mason, and Aarøe 2015; Jessee 2009). Participants were instructed to answer knowledge questions "to the best of your ability," but were also asked to not look up answers online (Motta, Callaghan, and Smith 2017; Vezzoni and Ladini 2017). Cheating behavior was measured with a difficult catch question, measured immediately after the knowledge items.

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 turnout intention (measured at Waves 2 and 3) and reported turnout (measured at Wave 4). 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 ($\alpha = .83$), and the scale was folded at the midpoint.

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. First, we control for survey effort by creating a latent measure based on two indicators of satisficing: the incidence of straight-lining and survey duration.[16] Second, we

[16] 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

control for a variety of factors that are associated with cheating in our data (see Table A6). In particular, cheaters are significantly younger, more educated, and less likely to be non-Hispanic white than non-cheaters (all *p*s < .10). Four personality traits were also 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). Both openness to experience and agreeableness are related to cheating behavior (all *p*s < .10). Given that past work has tied these traits to political engagement (Mondak 2010), we control for them in the analyses below.[17]

Study 3: Observational Results

Eleven percent of respondents in Study 3 cheated on the catch question, despite having been instructed not to look up answers. Cheating on the catch question corresponded with a 13 percentage point increase in knowledge scores compared to those who did not cheat. 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

---

knowledge battery). We then code survey duration into quintiles to combine with straight-lining and create the latent measure of survey effort.

[17] We weight responses to approximate the demographics of the general U.S. population in the multivariate analyses below. Survey weights were constructed by raking on respondents' age, race, ethnicity, income, gender, and education. Unweighted analyses produce substantively similar results.

in ceiling effects, which is exacerbated by the shorter knowledge scale. While 28% of non-cheaters answered all four questions correctly, 46% of cheaters did so.

Our expectation is that there will be a weaker association between political knowledge and criterion outcomes among cheaters than non-cheaters. We test these expectations using OLS models to predict political interest, turnout intention, and issue constraint as a function of the interaction between political knowledge and cheating behavior, while controlling for survey effort and the demographics and personality variables discussed above. Reported turnout is predicted with an analogous binary logistic regression. The results are reported in Figure 3 (full model details are shown in Table A5 in the Appendix).

Focusing first on our test of convergent validity, political knowledge is positively and significantly associated with political interest among non-cheaters ($b = .14$, $p < .001$) but is not significantly associated with political interest among cheaters ($b = -.04$, $p = .394$). This difference between cheaters and non-cheaters is significant ($p < .001$).

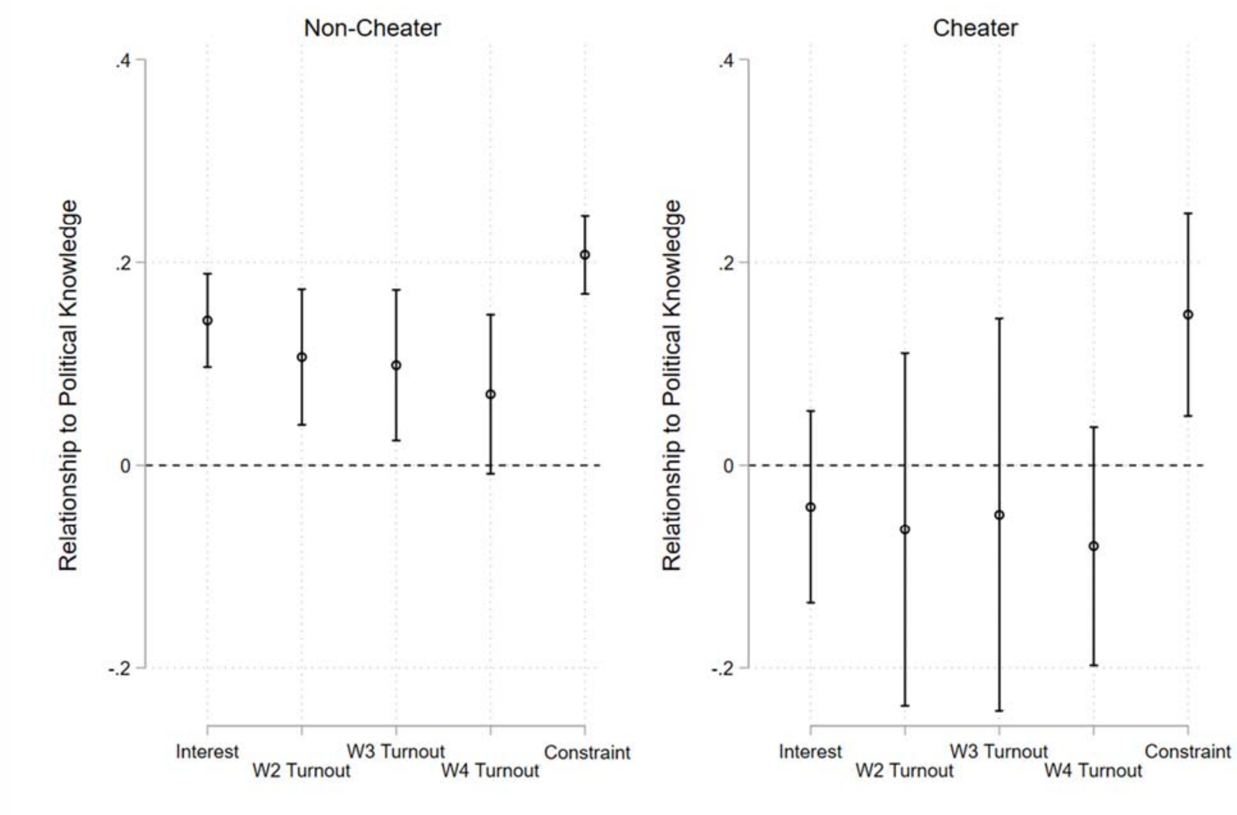
Turning to predictive validity, political knowledge is also a weaker predictor of turnout among cheaters than non-cheaters. Among non-cheaters, higher political knowledge predicts greater turnout intention at both Wave 2 ($b = .11$, $p = .002$) and Wave 3 ($b = .10$, $p = .009$), as well as reported turnout at Wave 4 (marginal effect = .07, $p = .070$).[18] Among cheaters, however, political knowledge does not significantly predict turnout at Wave 2 ($b = -.06$, $p = .475$), Wave 3

[18] The marginal effect of political knowledge on reported turnout is derived from a binary logistic regression model, in order to better compare to other outcomes. The estimated coefficient predicting reported turnout from political knowledge is .88 ($p = .071$) for non-cheaters, and -1.2 ($p = .261$) for cheaters.

(*b* = -.05, *p* =.621), or Wave 4 (marginal effect = -.08, *p* = .205). The difference between the coefficients for cheaters and non-cheaters is marginally significant at Wave 2 (*p* = .057) and Wave 4 (*p* = .071), though not at Wave 3 (*p* = .138). Finally, political knowledge is a somewhat stronger predictor of ideological constraint among non-cheaters (*b* = .21, *p* < .001) than cheaters (*b* = .15, *p* = .004), although this difference is not statistically significant (*p* = .275).

The results are summarized below in Figure 4. As shown in the left-hand panel, political knowledge is a strong predictor of theoretically relevant outcomes among respondents who did not look up answers. However, the right-hand panel shows that political knowledge loses nearly all of its predictive power among those who looked up answers. Overall, the findings from Study 3 are similar to those of Studies 1 and 2 despite the different methodological approach.

Figure 4. Validity of Political Knowledge by Cheating Behavior



How Concerned Should We Be About Cheating?

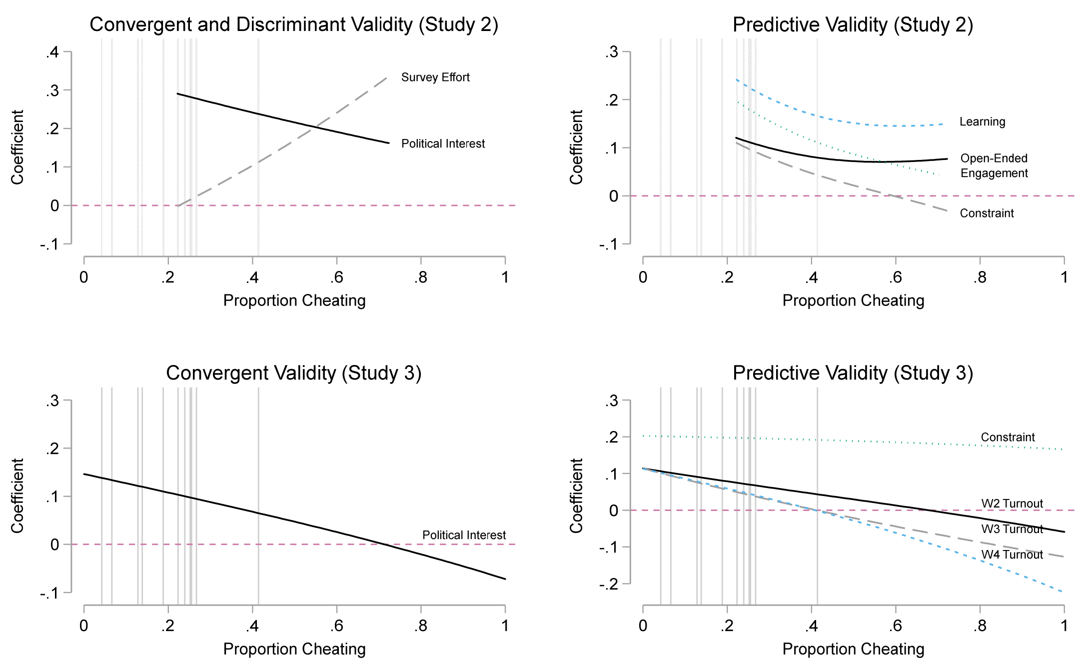
Studies 1 – 3 have shown that when respondents search for the answers to political knowledge questions, the resulting scales suffer from weakened convergent, discriminant, and predictive validity of knowledge measures. However, cheating rates are highly variable across studies, with published estimates ranging from 4% to 41% of the sample (Clifford and Jerit 2016). This variation makes it difficult to know when cheating is a problem. Here we explore the issue through a series of simulations. Drawing on Study 2, we randomly resample respondents from each treatment condition at different proportions and re-estimate the models reported above.[19] Specifically, we shift 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 cheating 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 cheating rate increases. Vertical lines represent published cheating rates from previous studies that did not attempt to control cheating (i.e., the vertical lines represent naturally occurring levels). The far left side of the graph represents a model estimated only from respondents in the discourage condition, corresponding with a cheating rate of 22%. Here,

[19] Given the similarity of the findings in Studies 1 and 2, the results of the simulation are similar regardless of which study we use.

interest is a strong predictor, while survey effort is not. However, as cheating rates rise, these patterns reverse and interest becomes a weaker predictor, while survey effort rapidly becomes a stronger predictor. Naturally, these patterns would be more dramatic if we could estimate these coefficients at a cheating rate of zero. The top right panel shows the coefficients for political knowledge from each of our predictive validity models. As cheating rates increase, the coefficient on knowledge from all four models steadily decrease. The drop in validity is particularly steep at lower levels of cheating (e.g., between .2 and .4), suggesting that cheating may be significantly undermining the predictive validity of our knowledge measure even in the discourage condition.

Figure 5. Simulated Validity Across Cheating Rates



We took the same approach with Study 3. Because Study 3 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. 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 is positively associated with 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. For all three measures of turnout, political knowledge is a positive predictor at low levels of cheating, but these effects again decline as cheating rates increase. In fact, for one measure of turnout, the coefficient on political knowledge reaches zero at a cheating rate of about 40%—a rate that has been observed in studies that did not attempt to control cheating. The only exception to this general pattern is constraint, for which knowledge remains a positive predictor across cheating rates, in contrast to Study 2. Overall, these simulations demonstrate that the convergent, discriminant, and predictive validity of political knowledge measures consistently declines as cheating rates increase. In some cases, political knowledge is completely unrelated with the criterion variable within ranges of cheating that have been observed in past studies.

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 changing what scholars are measuring with factual knowledge questions. A recent body of research demonstrates that cheating is common and can affect the estimated levels of political knowledge among the public (Burnett 2016; Clifford and Jerit 2014, 2016; Motta, Callaghan, and Smith 2017; Shulman and Boster 2014), but that literature has yet to establish how cheating affects the validity of these measures. Across a series of experimental and observational studies, we find a consistent pattern of results—namely, that cheating reduces the validity of political knowledge measures and undermines the ability to replicate canonical findings in the public opinion literature. When respondents look up answers, knowledge questions reflect the effort respondents are willing to put into the survey rather than their latent engagement with the politics.

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. In addition to differences in cheating rates across study populations (e.g., MTurk vs. students), cheating rates vary substantially *within* samples (cheating rates on MTurk have ranged from 5% to 25%). This variation makes it hard to know when cheating will reach problematic rates in a study.[20] Cheating is more common among younger respondents (see Table A6 in the Appendix), however, it is likely to become prevalent as all respondents become more comfortable with technology. Thus, even high quality national samples may experience increased rates of cheating in the years to come.

In light of these challenges, our first recommendation for researchers is to actively discourage cheating. Simple instructions to refrain from looking up answers can reduce cheating (Motta, Callaghan, and Smith 2017; Vezzoni and Ladini 2017), but do not eliminate it, as shown in our third study. A more effective tactic is to ask respondents to commit to not looking up

---

[20] The relationship between convenience samples (e.g., students, MTurk) and the target population is unknowable (Mullinix et al. 2016), which means that the baseline rate of cheating in such samples could change over time.

answers, but even this approach is not perfect (Clifford and Jerit 2016).[21] 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.[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 analysis may harm the representativeness of the sample and so should be discouraged if representativeness is a critical concern.[23] As an alternative, researchers should also present additional models that interact cheating with political knowledge and estimate effects among each subgroup (as in Study 3). This approach will provide insight into how sensitive the results are to cheating behavior.

A promising line of future research is to develop questions that are immune to cheating. For example, the answers to visual knowledge questions may be difficult to look up (Prior 2014). Alternative methods for gauging sophistication, such as text-analysis of open-ended questions

[21] Placing a time limit on each question has been a popular solution, but it is difficult to choose a time limit that is short enough to prevent cheating but does not interfere with the regular response process (Clifford and Jerit 2016).

[22] We recommend that the catch question be placed last among the knowledge questions. There is some evidence (Study 2) that respondents who received the catch question first were more likely to cheat, but this effect occurred only when search was allowed.

[23] Of course, cheaters should not be dropped from analysis if knowledge is measured post-treatment in an experimental design (Montgomery, Nyhan, and Torres 2016).

(Kraft n.d.), is another possibility. Given that levels of cheating vary considerably across surveys from the same population, it would be valuable to investigate whether aspects of survey design affect the prevalence of cheating.

A skeptic might object 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 very weakly related to interest in politics.[24]

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.

[24] Recall that in Study 2, which included a detailed measure of survey effort, the relationship between interest and effort was only $r = .10$.

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, and Brian F. Schaffner. 2014. "Does Survey Mode Still Matter? Findings from a 2010 Multi-Mode Comparison." *Political Analysis* 22(03): 285–303.

Baldassarri, Delia, and Andrew Gelman. 2008. "Partisans without Constraint: Political Polarization and Trends in American Public Opinion." *AJS; American journal of sociology* 114(2): 408–46.

Barabas, Jason, Jennifer Jerit, William Pollock, and Carlisle Rainey. 2014. "The Question(s) of Political Knowledge." *American Political Science Review* 108(04): 840–55.

Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.Com's Mechanical Turk." *Political Analysis* 20(3): 351–68.

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.

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.

Cor, M. Ken, Edward Haertel, Jon A. Krosnick, and Neil Malhotra. 2012. "Improving Ability Measurement in Surveys by Following the Principles of IRT: The Wordsum Vocabulary Test in the General Social Survey." *Social Science Research* 41(5): 1003–16.

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, and Lynn Vavreck. 2016. "How Face-to-Face Interviews and Cognitive Skill Affect Item Non-Response: A Randomized Experiment Assigning Mode of Interview."

Political Science Research and Methods 36(1): 1–20.

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.

Huddy, Leonie, Lilliana Mason, and Lene Aarøe. 2015. "Expressive Partisanship: Campaign Involvement, Political Emotion, and Partisan Identity." *American Political Science Review* 109(01): 1–17.

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

Jessee, Stephen A. 2009. "Spatial Voting in the 2004 Presidential Election." *American Political Science Review* 103(01): 59–81.

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.

Kalmoe, Nathan P. 2018. *The Weakness of Issues: How Multiple Measures Mislead on Public Opinion*.

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.

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.

Kraft, Patrick. *Let's Talk Politics: A Naive Approach for Measuring Political Sophistication*.

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.

Lelkes, Yphtach. 2016. "Mass Polarization: Manifestations and Measurements." *Public Opinion Quarterly* 80(S1): 392–410.

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.

———. 2010. *Personality and the Foundations of Political Behavior*. New York: Cambridge University Press.

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*: edw027.

Mullinix, Kevin J., Thomas J. Leeper, James N. Druckman, and Jeremy Freese. 2016. “The Generalizability of Survey Experiments.” *Journal of Experimental Political Science* 2(02): 109–38.

Prior, Markus. 2014. “Visual Political Knowledge: A Different Road to Competence?” *The Journal of Politics* 76(1): 41–57.

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.

Appendix for "How Internet Search Undermines the Validity of Political Knowledge Measures"

Table A1. Study 1 and Study 2 Demographics

	Study 1	Study 2	
		Wave 1	Wave 2
Partisan Identity			
Democrat	66%	67%	68%
Republican	24%	22%	21%
Independent/Other	10%	10%	11%
Ideology			
Liberal	54%	51%	50%
Conservative	18%	17%	16%
Moderate	29%	32%	33%
Race and Ethnicity			
White	23%	27%	26%
Black	12%	10%	9%
Asian	31%	33%	34%
Hispanic	28%	26%	26%
Other	6%	5%	5%
Born in the U.S.	78%	79%	80%
Male	45%	43%	41%
N	1170	887	644

Table A2. Convergent and Discriminant Validity (Study 1 & 2)

	Study 1		Study 2	
Discourage	-0.80	***	-0.78	***
	(.05)		(.06)	
Political Interest	0.12	**	0.16	***
	(.04)		(.04)	
Survey Effort	0.36	***	0.34	***
	(.04)		(.04)	
Discourage x Political Interest	0.16	**	0.13	*
	(.05)		(.06)	
Discourage x Survey Effort	-0.33	***	-0.34	***
	(.05)		(.06)	
Constant	0.39	***	0.41	***
	(.04)		(.04)	
N	1162		887	
R^2	0.28		0.26	

Note: + p < .10, * p < .05, ** p < .01, *** p < .001.

Table A3. Predictive Validity (Study 2)

	Open-Ended		Ideological Constraint		Participation	Learning
	Wave 1	Waves 1-2	Wave 1	Waves 1-2	Wave 1	Wave 2
Political Knowledge	0.08 +	0.08	0.02	-0.02	0.08	0.14 *
	(.05)	(.05)	(.05)	(.06)	(.05)	(.06)
Commitment	0.11	0.13	0.03	0.04	0.07	0.16 +
	(.07)	(.08)	(.07)	(.08)	(.08)	(.08)
Knowledge x Commitment	0.02	0.05	0.14 +	0.14	0.15 +	0.11
	(.07)	(.08)	(.07)	(.09)	(.08)	(.08)
Survey Effort - W1	0.25 ***	0.08 +	0.18 ***	0.04	0.03	0.08 +
	(.03)	(.04)	(.03)	(.05)	(.04)	(.05)
Survey Effort - W2	-	0.34 ***	-	0.21 ***	-	0.16 ***
		(.04)		(.05)		(.05)
	-	-				
Constant	0.05	0.06	0.01	0.00	0.08	-0.07
	(.05)	(.06)	(.05)	(.06)	(.06)	(.06)
N	886	642	887	642	702	636
R^2	0.08	0.17	0.05	0.06	0.03	0.09
Marginal Effects of Knowledge						
Commitment:	0.10 *	0.12 *	0.16 **	0.11 +	0.22 ***	0.24 ***
	(.05)	0.06	(.05)	0.06	(.06)	(.06)
Encouragement:	0.08 +	0.08	0.02	-0.02	0.08	0.14 *
	(.05)	0.05	(.05)	0.06	(.05)	(.06)

Note: + p < .10, * p < .05, ** p < .01, *** p < .001.

Table A4. Study 3 Demographics

	Wave 1	Wave 2	Wave 3	Wave 4
Partisan Identity				
Democrat	42%	43%	44%	41%
Republican	32%	32%	41%	33%
Independent/Other	26%	26%	15%	26%
Ideology				
Liberal	31%	31%	27%	29%
Conservative	37%	37%	43%	39%
Moderate	32%	33%	31%	33%
Race and Ethnicity				
White	77%	78%	81%	79%
Black	14%	15%	13%	14%
Asian	3%	3%	3%	4%
Native American	1%	1%	1%	1%
Other	5%	3%	2%	3%
Hispanic	17%	14%	7%	13%
Male	38%	38%	34%	37%
Median Education	Associate's Degree	Associate's Degree	Associate's Degree	Associate's Degree
Median Income	$50,000-$74,999	$50,000-$74,999	$50,000-$74,999	$50,000-$74,999
Average Age	51	54	57	55
N	3552	2024	1234	1730

Table A5. Convergent and Predictive Validity (Study 3)

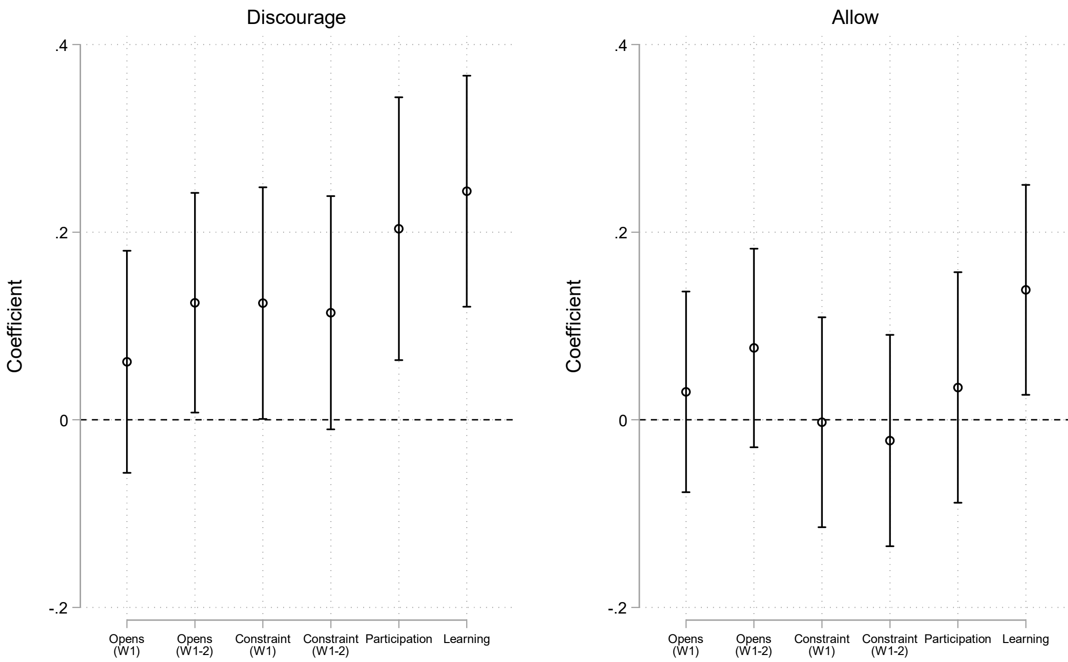
Outcome	Political Interest	W2 Turnout	W3 Turnout	W4 Turnout	Constraint
Model	OLS	OLS	OLS	Logit	OLS
Political Knowledge	0.14 ***	0.11 **	0.10 **	0.88 +	0.21 ***
	(.02)	(.03)	(.04)	(.49)	(.02)
Cheater	0.16 ***	0.13 *	0.08	1.69 +	0.04
	(.04)	(.06)	(.07)	(.98)	(.04)
Knowledge x Cheater	-0.18 ***	-0.17 +	-0.15	-2.12 +	-0.06
	(.05)	(.09)	(.10)	(1.17)	(.05)
Survey Effort	0.12 ***	0.11 *	0.12 *	1.10 *	0.06
	(.03)	(.05)	(.05)	(.48)	(.02) **
Age	0.00 **	0.00 ***	0.00 **	0.03 ***	0.00
	(.00)	(.00)	(.00)	(.01)	(.00)
Non-Hispanic White	-0.01	0.02	-0.02	-0.13	-0.06 ***
	(.01)	(.02)	(.03)	(.27)	(.01)
Education	0.12 ***	0.12 **	0.07 +	1.20 *	0.03
	(.02)	(.04)	(.04)	(.54)	(.02)
Openness to Experience	0.04	0.08	0.04	0.59	0.11 ***
	(.04)	(.05)	(.06)	(.65)	(.03)
Agreeableness	0.19 ***	0.04	0.01	-0.34	0.15 ***
	(.04)	(.06)	(.06)	(.63)	(.03)
Constant	0.28 ***	0.42 ***	0.59 ***	-1.22 *	0.01
	(.04)	(.06)	(.06)	(.62)	(.03)
N	3447	1927	1190	1672	3443
R^2	0.13	0.10	0.08	-	0.14
Marginal Effects of Knowledge					
Non-Cheater	0.14 ***	0.11 **	0.10 **	0.07 +	0.21 ***
	(.02)	(.03)	(.04)	(.04)	(.02)
Cheater:	-0.04	-0.06	-0.05	-0.08	0.15 ***
	(.05)	(.09)	(.10)	(.06)	(.05)

Note: + p < .10, * p < .05, ** p < .01, *** p < .001.

Table A6. Study 3 Comparison of Cheaters and Non-Cheaters

	Cheaters	Non-Cheaters	*p*-value
	Mean	Mean	*t*-test
Age	46.43	51.94	<.001
	(.77)	(.29)	
Openness to Experience	0.69	0.64	<.001
	(.01)	(.00)	
Agreeableness	0.72	0.70	0.06
	(.01)	(.00)	
Conscientiousness	0.71	0.70	0.12
	(.01)	(.00)	
Need for Closure	0.47	0.46	0.37
	(.01)	(.00)	
	Percent	Percent	*t*-test
Male	41%	38%	0.20
	(.02)	(.01)	
Non-Hispanic White	55%	66%	<.001
	(.03)	(.01)	
	Median	Median	$\chi 2$
Education	Bachelor's Degree	Associate's Degree	0.06
Income	$50,000-$74,999	$50,000-$74,999	0.49

Figure A1. Predictive Validity of Political Knowledge by Experimental Condition



Latent Measures of Survey Effort (Studies 1 and 2)

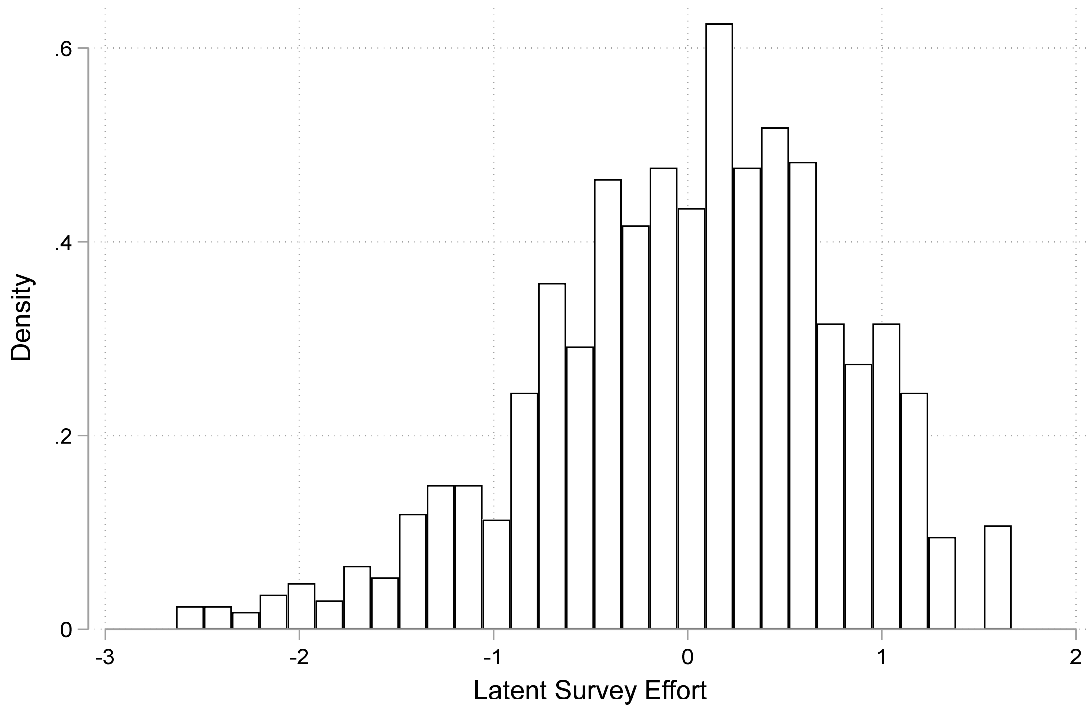
In Study 1, latent survey effort was estimated using six indicators of satisficing. The first item was a standard manipulation check consisting of three facts that appeared in a brief story for the purpose of another study. The second item asked respondents (at end of survey) to report how carefully they had answered questions on a five-point scale. The third item is a count of the number of grids that a respondent straight-lined (provided the same response option to every item). The scale is reversed so that a "5" indicates that the respondent did *not* straight-line at all. The fourth item is the total time spent on the survey. Time spent on the knowledge questions is subtracted from the total time to avoid any influence of treatment assignment, and then divided into quintiles. The sixth item consists of a question placed in a grid instructing respondents to select a particular response option. The final item is an instructional manipulation check. The six items were scaled together using a hybrid item response model that combines a graded response model for the four ordinal items and a two-parameter logistic model for the two dichotomous items. The results of the IRT model are shown below. All items have good discrimination parameters, indicating they contribute to estimation of the same latent variable.

Table A7. Latent Estimates of Survey Effort (Study 1)

	Discrimination		Difficulty Parameters									
	Parameter		>=1		>=2		>=3		>=4		=5	
Graded Response Model												
Manipulation Checks (0-3)	1.19	***	-3.15	***	-1.42	***	0.08					
	(.11)		(.25)		(.12)		(.06)					
Self-Reported Effort (0-4)	1.02	***	-5.69	***	-3.65	***	-0.55	***	2.14	***		
	(.10)		(.62)		(.33)		(.08)		(.19)			
Straight-lining (0-5)	1.40	***	-4.98	***	-4.37	***	-3.47	***	-2.91	***	-1.42	***
	(.16)		(.16)		(.44)		(.31)		(.24)		(.12)	
Time on Survey Quintiles (0-4)	0.92	***	-1.71	***	-0.44	***	0.58	***	1.78	***		
	(.09)		(.16)		(.08)		(.09)		(.16)			
Two-Parameter Logistic												
Grid Instructed Response (0-1)	2.49	***	-1.79	***								
	(.39)		(.12)									
IMC (0-1)	1.32	***	0.75	***								
	(.16)		(.08)									
N	1170											

Note: + p < .10, * p < .05, ** p < .01, *** p < .001.

A histogram of latent survey effort is shown below.



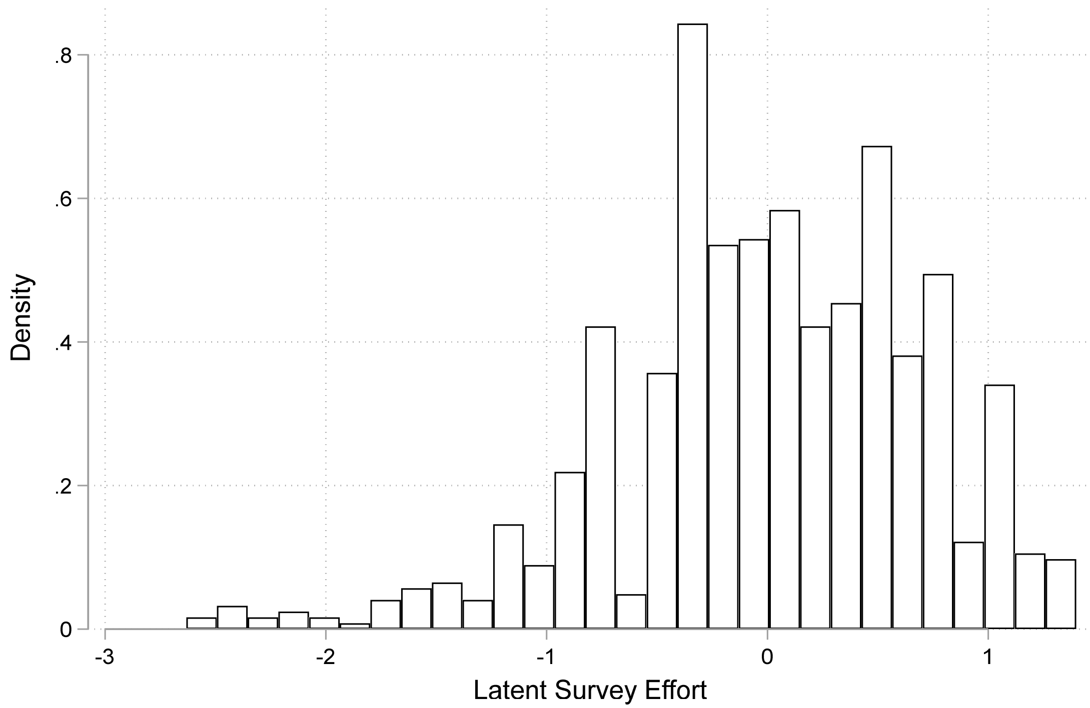
The first wave of Study 2 used a very similar approach with three key differences. First, no substantive manipulation checks were available, so this item was not included. Second, open-ended responses were available, so an item was created that counted the number of open-ended questions that a respondent left completely blank (reverse-scored). Third, a more fine-grained measure of time was available, so time was only used prior to the political knowledge battery to avoid any influence of experimental condition. The results of the IRT model are shown in Table A8 below. Again, all items have acceptable discrimination parameters, indicating they are all contributing to estimation of the same latent variable.

Table A8. Latent Estimates of Survey Effort (Study 2, Wave 1)

	Discrimination		Difficulty Parameters							
	Parameter		>=1		>=2		>=3		>=4	
Graded Response Model										
Self-Reported Effort (0-4)	0.82	***	-7.39	***	-4.72	***	-0.70	***	2.65	***
	(.13)		(1.25)		(.68)		(.13)		(.36)	
Straight-lining (0-4)	1.26	***	-4.96	***	-4.26	***	-3.23	***	-1.97	***
	(.23)		(.78)		(.62)		(.44)		(.25)	
Time on Survey Quintiles (0-4)	0.69	***	-2.15	***	-0.59	***	0.71	***	2.23	***
	(.11)		(.32)		(.14)		(.14)		(.33)	
Open-Ended Blank (0-2)	0.51	*	-7.49	*	-5.87	*				
	(.21)		(2.95)		(2.28)					
Two-Parameter Logistic										
Grid Instructed Response (0-1)	1.83	***	-2.22	***						
	(.36)		(.24)							
IMC (0-1)	1.39	***	0.35	***						
	(.27)		(.08)							
N	887									

Note: + p < .10, * p < .05, ** p < .01, *** p < .001.

A histogram of latent survey effort is shown below.

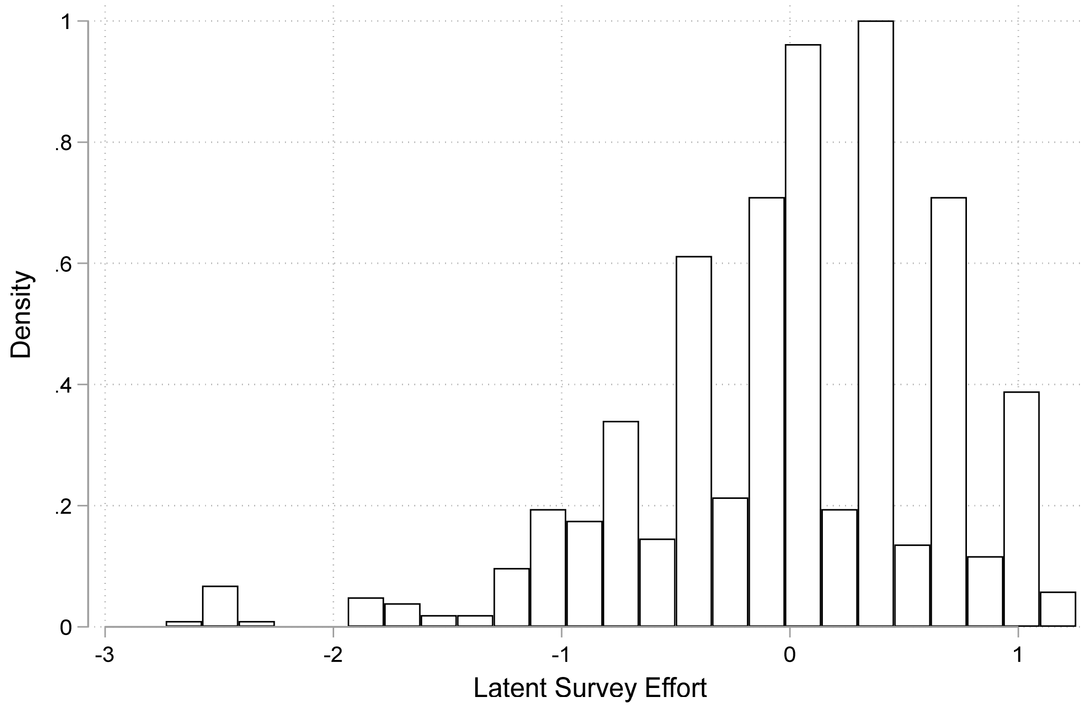


The second wave of Study 2 used a very similar approach as to the first wave. IRT model results are shown in Table A9, below.

Table A9. Latent Estimates of Survey Effort (Study 2, Wave 2)

	Discrimination Parameter		Difficulty Parameters >=1		>=2		>=3		>=4	
Graded Response Model										
Self-Reported Effort (0-4)	0.53	***	-7.79	***	-4.84	***	0.48	*	5.23	***
	(.13)		(1.92)		(1.15)		(.19)		(1.25)	
Straight-lining (0-4)	1.58	***	-3.87	***	-2.72	***	-1.63	***		
	(.34)		(.60)		(.38)		(.21)			
Time on Survey Quintiles (0-4)	1.12	***	-1.50	***	-0.41	***	0.48	***	1.53	***
	(.23)		(.24)		(.11)		(.11)		(.24)	
Two-Parameter Logistic										
Grid Instructed Response (0-1)	2.40	**	-2.20	***						
	(.70)		(.27)							
IMC (0-1)	0.60	***	-2.07	***						
	(.17)		(.54)							
Skipped Open-Ended (0-1)	0.59	**	-3.56	**						
	(.19)		(1.03)							
N	645									

Note: + p < .10, * p < .05, ** p < .01, *** p < .001.



Density
1
.8
.6
.4
.2
0
-3
-2
-1
0
1
Latent Survey Effort