SURVEY-BASED MEASURES OF POLITICAL KNOWLEDGE:

CONSIDERATIONS & BEST PRACTICES

Jennifer Jerit
Department of Political Science
Stony Brook University

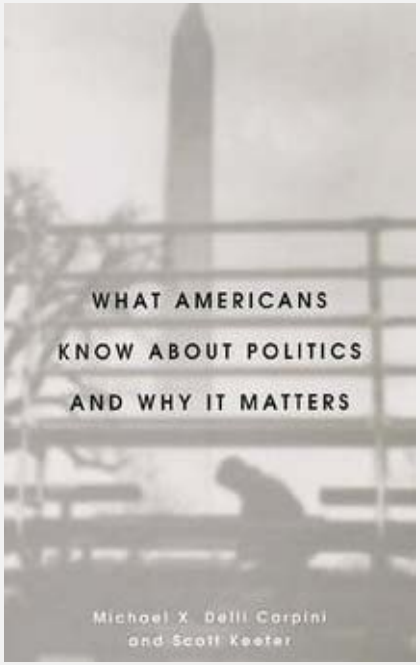
OUTLINE OF TALK

- Why political knowledge?
- Considerations regarding measurement
- Online look-up (a *new* measurement challenge)

"Political knowledge has attained the status as a cornerstone construct in research on political behavior."

Jeffery Mondak
American Journal of Political Science (2001)

DEFINITION (DELLI CARPINI AND KEETER 1996)



- Political knowledge is "the range of factual information about politics stored in long term memory" (1996, 10).
- Can measure concept with a small number of survey questions that can be scaled together.

DELLI CARPINI & KEETER'S (1996) CANONICAL FIVE-ITEM INDEX

- Party control of the U.S. House
- Veto override percent
- Party ideological location
- Judicial review
- Identifying the Vice President of the United States

British Election Study (BES) has similar set of items asking about leaders from UK and other countries.

RECALL-BASED KNOWLEDGE VERSUS OTHER CONCEPTUALIZATIONS

- Presumption: knowledge can (and should) be measured with survey questions

- Other viewpoints (*briefly*)
  - Recall-based knowledge versus other conceptions
    - Learning skills (Prior and Lupia 2009)
    - "Off-loading" to the Internet

THE CASE FOR RECALL-BASED MEASURES OF POLITICAL KNOWLEDGE

The New York Times | https://nyti.ms/2rArYVn

SundayReview

You Still Need Your Brain

Gray Matter

By DANIEL T. WILLINGHAM MAY 19, 2017

Most adults recall memorizing the names of rivers or the Pythagorean theorem in school and wondering, "When am I ever gonna use this stuff?" Kids today have a high-profile spokesman. Jonathan Rochelle, the director of Google's education apps group, said last year at an industry conference that he "cannot answer" why his children should learn the quadratic equation. He wonders why they cannot "ask Google." If Mr. Rochelle cannot answer his children, I can.

Google is good at finding information, but the brain beats it in two essential ways. Champions of Google underestimate how much the meaning of words and sentences changes with context. Consider vocabulary. Every teacher knows that a sixth grader, armed with a thesaurus, will often submit a paper studded with words used in not-quite-correct ways, like the student who looked up "meticulous," saw it meant "very careful," and wrote "I was meticulous when I fell off the cliff."

With the right knowledge in memory, your brain deftly puts words in context. Consider "Trisha spilled her coffee." When followed by the sentence "Dan jumped up to get a rag," the brain instantly highlights one aspect of the meaning of "spill" —

THE CASE FOR RECALL-BASED MEASURES OF POLITICAL KNOWLEDGE

- Politically knowledgeable differ from the less informed in a myriad of ways relating to opinion quality:
  - Link values and interests to their attitudes
  - Retention of new information
  - Correct use ideological concepts
  - Use of evidence in political discussions

"Political awareness is best represented with data from survey batteries that measure factual political knowledge."

Jeffery Mondak
American Journal of Political Science (2001)

CONSIDERATIONS REGARDING MEASUREMENT

- Format of question
- Response options
- Scale construction

CONSIDERATIONS REGARDING MEASUREMENT

- Format of question
  - Open versus closed-ended

CONSIDERATIONS REGARDING MEASUREMENT

- Format of question
  - Open versus closed-ended
  - Verbal versus visual questions

CONSIDERATIONS REGARDING MEASUREMENT

- Format of question
- Response options
  - Include an explicit "Don't Know" option?

CONSIDERATIONS REGARDING MEASUREMENT

- Format of question
- Response options
- Scale construction
  - What kinds of questions to include?

THE PROBLEM OF "OUTSIDE SEARCH"

THE PROBLEM OF "OUTSIDE SEARCH"

- As surveys are increasingly administered online, respondents can use search engines to look up answers to questions.

THE PROBLEM OF "OUTSIDE SEARCH"

- As surveys are increasingly administered online, respondents can use search engines to look up answers to questions.

- Observed levels of knowledge tend to be higher in online surveys, suggesting that such scales are *descriptively* less accurate.

AN ILLUSTRATION FROM A MODE EXPERIMENT

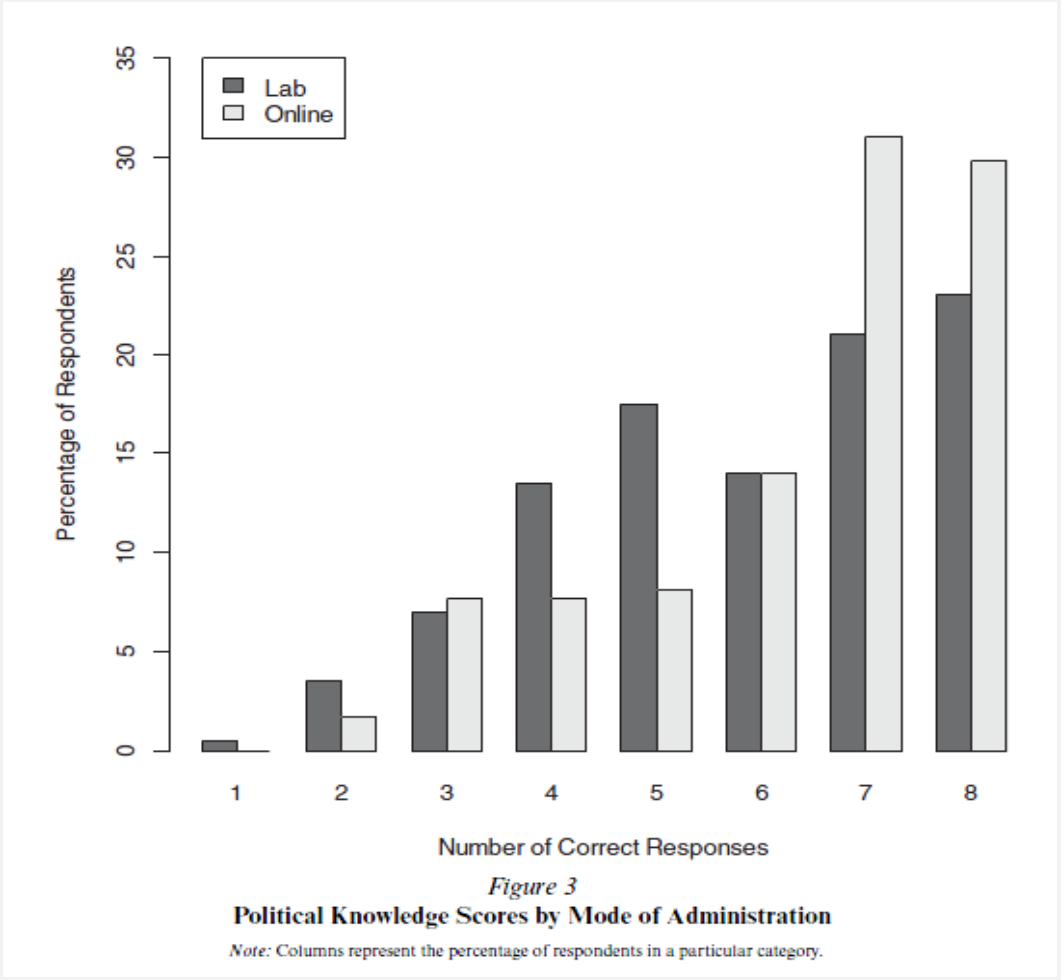


Figure 3
Political Knowledge Scores by Mode of Administration
Note: Columns represent the percentage of respondents in a particular category.

From Clifford and Jerit (2014)

Average correct by condition 6.4 vs. 5.9 ($p < .01$)

61% score of 7 or higher online
44% score of 7 or higher in lab

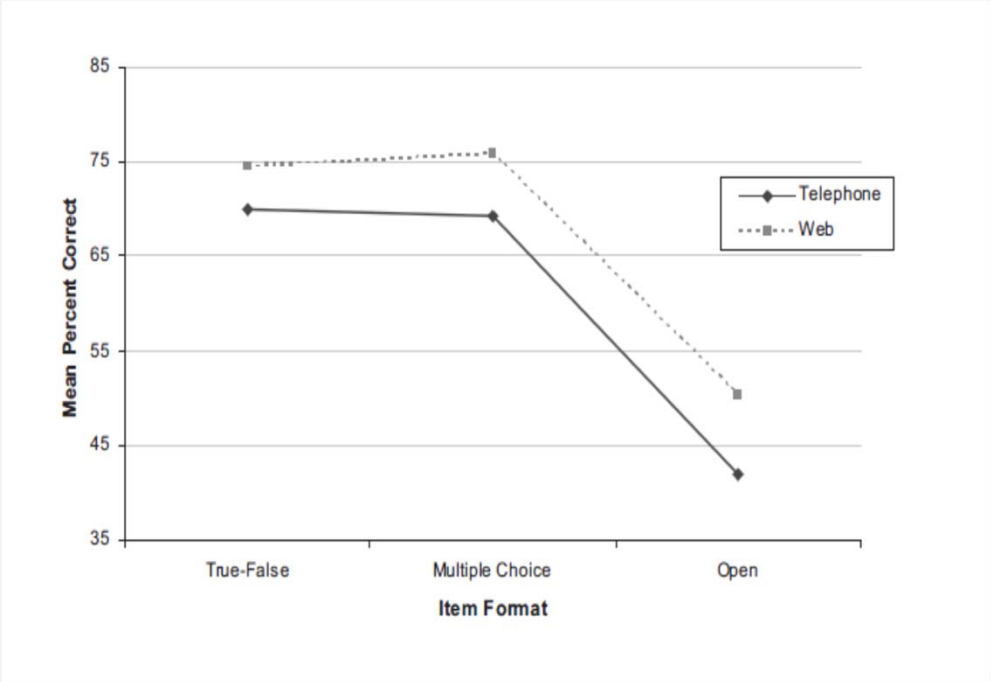
ANOTHER ILLUSTRATION

Public Opinion Quarterly, Vol. 69, No. 3, Fall 2005, pp. 370–392

AN EXPERIMENTAL COMPARISON OF WEB AND TELEPHONE SURVEYS

SCOTT FRICKER
MIRTA GALESIC
ROGER TOURANGEAU
TING YAN

Abstract We carried out an experiment that compared telephone and Web versions of a questionnaire that assessed attitudes toward science and knowledge of basic scientific facts. Members of a random digit dial (RDD) sample were initially contacted by telephone and answered a few screening questions, including one that asked whether they had Internet access. Those with Internet access were randomly assigned to complete either a Web version of the questionnaire or a computer-assisted telephone interview. There were four main findings. First, although we offered cases assigned to the Web survey a larger incentive, fewer of them completed the online questionnaire; almost all those who were assigned to the telephone condition completed the interview. The two samples of Web users nonetheless had similar demographic characteristics. Second, the Web survey produced less item nonresponse than the telephone survey. The Web questionnaire prompted respondents when they left an item blank, whereas the telephone interviewers accepted "no opinion" answers without probing them. Third, Web respondents gave less differentiated answers to batteries of attitude items than their telephone counterparts. The Web questionnaire presented these items in a grid that may have made their similarity more salient.



OUTSIDE SEARCH IN THE AMERICAN NATIONAL ELECTION STUDIES (ANES)

- 2012 and 2016 ANES offer circumstantial evidence
- 2018 ANES Pilot had two "catch" questions

OUTSIDE SEARCH IN THE AMERICAN NATIONAL ELECTION STUDIES (ANES)

- 2012 and 2016 ANES offer circumstantial evidence
- 2018 ANES Pilot had two "catch" questions

International Journal of Public Opinion Research

doi: 10.1093/ijpor/edw027

Looking for Answers: Identifying Search Behavior and Improving Knowledge-Based Data Quality in Online Surveys

Matthew P. Motta[1], Timothy H. Callaghan[2], and Brianna Smith[1]

[1]Political Science Department, University of Minnesota, Minneapolis, MN, USA; [2]Department of Health Policy and Management, Texas A&M University, College Station, TX, USA

Abstract

Internet surveys are cost-effective and convenient, but give respondents the opportunity to search the Web for answers to factual questions. Surprisingly, few have investigated the quality of political knowledge data collected in crowd-sourced Internet samples. Using a novel technique to identify individual "cheaters" across six data sets—collected on Amazon's Mechanical Turk (professionalized paid "workers") and SurveyMonkey (less-professionalized unpaid "audiences")—we find that cheating is pervasive. In addition to detecting "cheaters," we also show that their political knowledge is significantly inflated compared with non-cheaters, posing problems for the accurate measurement of knowledge. Fortunately, we also demonstrate that question-wording alterations can significantly improve data quality and show that cheating is less-common on mobile survey-taking platforms (compared with desktop/laptop computers).

OUTSIDE SEARCH IN THE AMERICAN NATIONAL ELECTION STUDIES (ANES)

- 2012 and 2016 ANES offer circumstantial evidence
- 2018 ANES Pilot had two "catch" questions
  - In what year did the Supreme Court of the United States decide Geer vs. Connecticut? (1896)
  - In what year was the Alaska Purchase Treaty signed? (1867)

International Journal of Public Opinion Research

doi: 10.1093/ijpor/edw027

Looking for Answers: Identifying Search Behavior and Improving Knowledge-Based Data Quality in Online Surveys

Matthew P. Motta[1], Timothy H. Callaghan[2], and Brianna Smith[1]

[1]Political Science Department, University of Minnesota, Minneapolis, MN, USA; [2]Department of Health Policy and Management, Texas A&M University, College Station, TX, USA

Abstract

Internet surveys are cost-effective and convenient, but give respondents the opportunity to search the Web for answers to factual questions. Surprisingly, few have investigated the quality of political knowledge data collected in crowd-sourced Internet samples. Using a novel technique to identify individual "cheaters" across six data sets—collected on Amazon's Mechanical Turk (professionalized paid "workers") and SurveyMonkey (less-professionalized unpaid "audiences")—we find that cheating is pervasive. In addition to detecting "cheaters," we also show that their political knowledge is significantly inflated compared with non-cheaters, posing problems for the accurate measurement of knowledge. Fortunately, we also demonstrate that question-wording alterations can significantly improve data quality and show that cheating is less-common on mobile survey-taking platforms (compared with desktop/laptop computers).

OUTSIDE SEARCH IN THE AMERICAN NATIONAL ELECTION STUDIES (ANES)

- 2012 and 2016 ANES offer circumstantial evidence
- 2018 ANES Pilot had two "catch" questions
  - In what year did the Supreme Court of the United States decide Geer vs. Connecticut? (1896)
    - 18% of respondents "caught" on this item
  - In what year was the Alaska Purchase Treaty? (1867)
    - 23% of respondents "caught" on this item

25% percent of respondents looked up the answer to at least one of the catch questions and 16% looked up the answer to both

International Journal of Public Opinion Research

doi: 10.1093/ijpor/edw027

Looking for Answers: Identifying Search Behavior and Improving Knowledge-Based Data Quality in Online Surveys

Matthew P. Motta[1], Timothy H. Callaghan[2], and Brianna Smith[1]

[1]Political Science Department, University of Minnesota, Minneapolis, MN, USA; [2]Department of Health Policy and Management, Texas A&M University, College Station, TX, USA

Abstract

Internet surveys are cost-effective and convenient, but give respondents the opportunity to search the Web for answers to factual questions. Surprisingly, few have investigated the quality of political knowledge data collected in crowd-sourced Internet samples. Using a novel technique to identify individual "cheaters" across six data sets—collected on Amazon's Mechanical Turk (professionalized paid "workers") and SurveyMonkey (less-professionalized unpaid "audiences")—we find that cheating is pervasive. In addition to detecting "cheaters," we also show that their political knowledge is significantly inflated compared with non-cheaters, posing problems for the accurate measurement of knowledge. Fortunately, we also demonstrate that question-wording alterations can significantly improve data quality and show that cheating is less-common on mobile survey-taking platforms (compared with desktop/laptop computers).

EFFECTS OF OUTSIDE SEARCH

Comparing Cheaters and Non-Cheaters in 2018 ANES*

	Non-Cheaters (n=917)	Cheaters (n=346)
Mean Score 4 pt Knowledge Scale	1.55	2.71
% Max Score on Knowledge Scale	13.26%	25.08%
Time (seconds) to answer:		
Open-end John Roberts	19.89	42.1
Open-end Angela Merkel	18.26	28.91
Open-end Length Senate Term	13.1	29.56
Closed-end Government Spending	15.02	24.19
Geer (Catch 1)	15.01	44.54
Alaska (Catch 2)	12.39	38.04

Note: Cheaters are Rs who answered at least one catch question correctly. Differences in score and timing variables are statisitcally significant.

*Respondents in control condition of [rand_pk] experiment.

THE PROBLEM OF "OUTSIDE SEARCH"

- As surveys are increasingly administered online, respondents can use search engines to look up answers to questions.
- Observed levels of knowledge tend to be higher in online surveys, suggesting that such scales are *descriptively* less accurate.
- Does outside search affect the *validity* of knowledge scales?

OUR ARGUMENT

- Zaller (1992): Knowledge items are used to measure a person's latent "intellectual and cognitive engagement with public affairs."

OUR ARGUMENT

- Zaller (1992): Knowledge items are used to measure a person's latent "intellectual and cognitive engagement with public affairs."

- Search behavior may improve a person's score but it does not necessarily relate to the latent ability the scale is designed to measure.

OUR ARGUMENT

- Zaller (1992): Knowledge items are used to measure a person's latent "intellectual and cognitive engagement with public affairs."
- Search behavior may improve a person's score but it does not necessarily relate to the latent ability the scale is designed to measure.
- Search behavior reflects effort/attentiveness (and possibly other traits).

OUR ARGUMENT

- Zaller (1992): Knowledge items are used to measure a person's latent "intellectual and cognitive engagement with public affairs."
- Search behavior may improve a person's score but it does not necessarily relate to the latent ability the scale is designed to measure.
- Search behavior reflects effort/attentiveness (and possibly other traits).
- When outside search takes place, knowledge scales will be inferior in terms of convergent, discriminant, and predictive validity.

EMPIRICAL STUDIES

- **Study 1**
  - Experiment with student sample (N=1,170) in Spring 2017
  - Compare convergent and discriminant validity across conditions where search is discouraged vs. allowed
- **Study 2**
  - Experiment with student sample (N= 887) in Fall 2017
  - Compare convergent, discriminant, and predictive validity across conditions where search is discouraged vs. allowed
- **Study 3**
  - National sample (SSI); 4-wave panel in Fall of 2016
  - Observational (measure search behavior as it naturally occurs in an online survey)
  - Examine convergent and predictive validity among cheaters and non-cheaters

STUDIES 1 & 2
MEASURES

MANIPULATING SEARCH BEHAVIOR

- **Discourage condition**

  "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?"

- **Allow condition**

  "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."

OTHER MEASUREMENT DETAILS

- Measure online look up with self-report and catch question

OTHER MEASUREMENT DETAILS

- Measure online look up with self-report and catch question
- Political interest (convergent validity) is a multi-item scale (α = .76 to .83)

OTHER MEASUREMENT DETAILS

- Measure online look up with self-report and catch question
- Political interest (convergent validity) is a multi-item scale (α = .76 to .83)
- Latent survey effort (discriminant validity) estimated as a function of common indicators of satisficing (e.g., IMC, straight-lining, time spent, instructed response items, factual manipulation checks)

OTHER MEASUREMENT DETAILS

- Measure online look up with self-report and catch question
- Political interest (convergent validity) is a multi-item scale (α = .76 to .83)
- Latent survey effort (discriminant validity) estimated as a function of common indicators of satisficing (e.g., IMC, straight-lining, time spent, instructed response items, factual manipulation checks)
- Criterion outcomes (predictive validity) consist of open-ended elaboration, political engagement, issue constraint, recall from article.

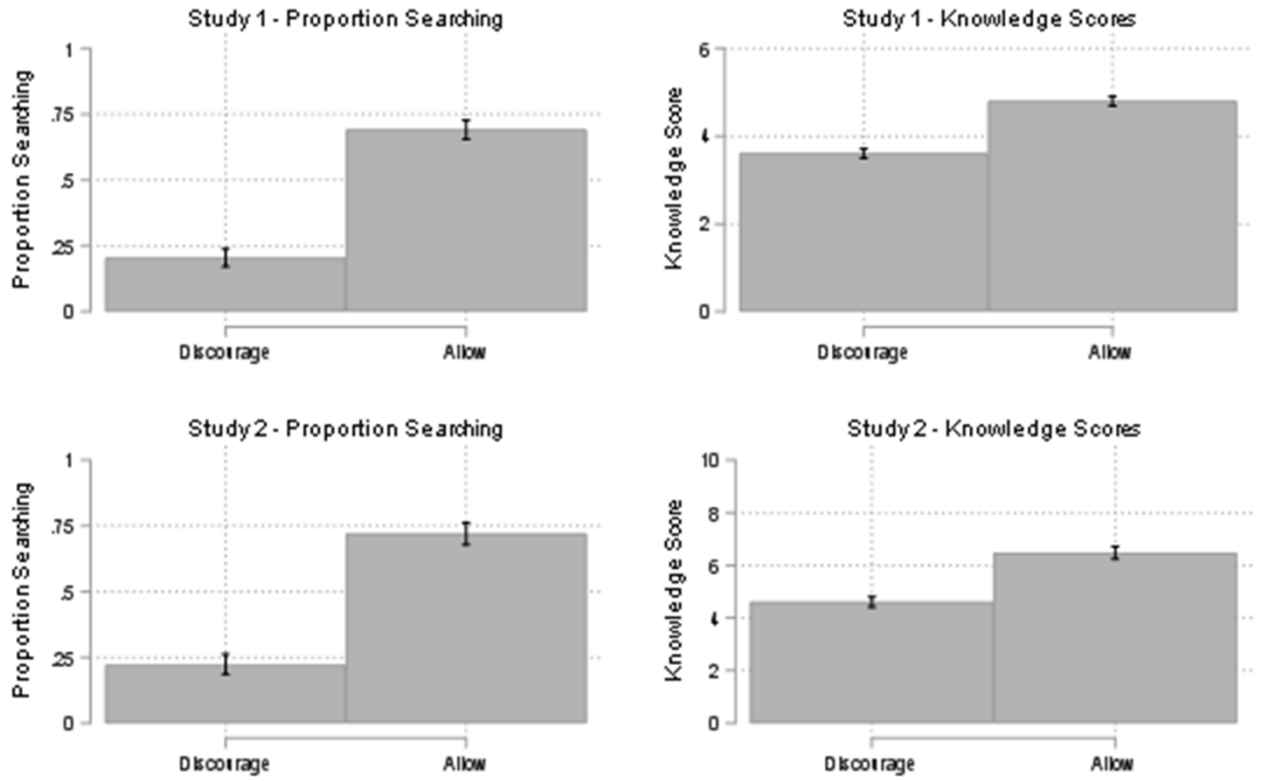
OTHER MEASUREMENT DETAILS

- Measure online look up with self-report and catch question
- Political interest (convergent validity) is a multi-item scale (α = .76 to .83)
- Latent survey effort (discriminant validity) estimated as a function of common indicators of satisficing (e.g., IMC, straight-lining, time spent, instructed response items, factual manipulation checks)
- Criterion outcomes (predictive validity) consist of open-ended elaboration, political engagement, issue constraint, recall from article.
- Knowledge measured with 6 items in Study 1 and 10 items in Study 2.

STUDIES 1 & 2
RESULTS

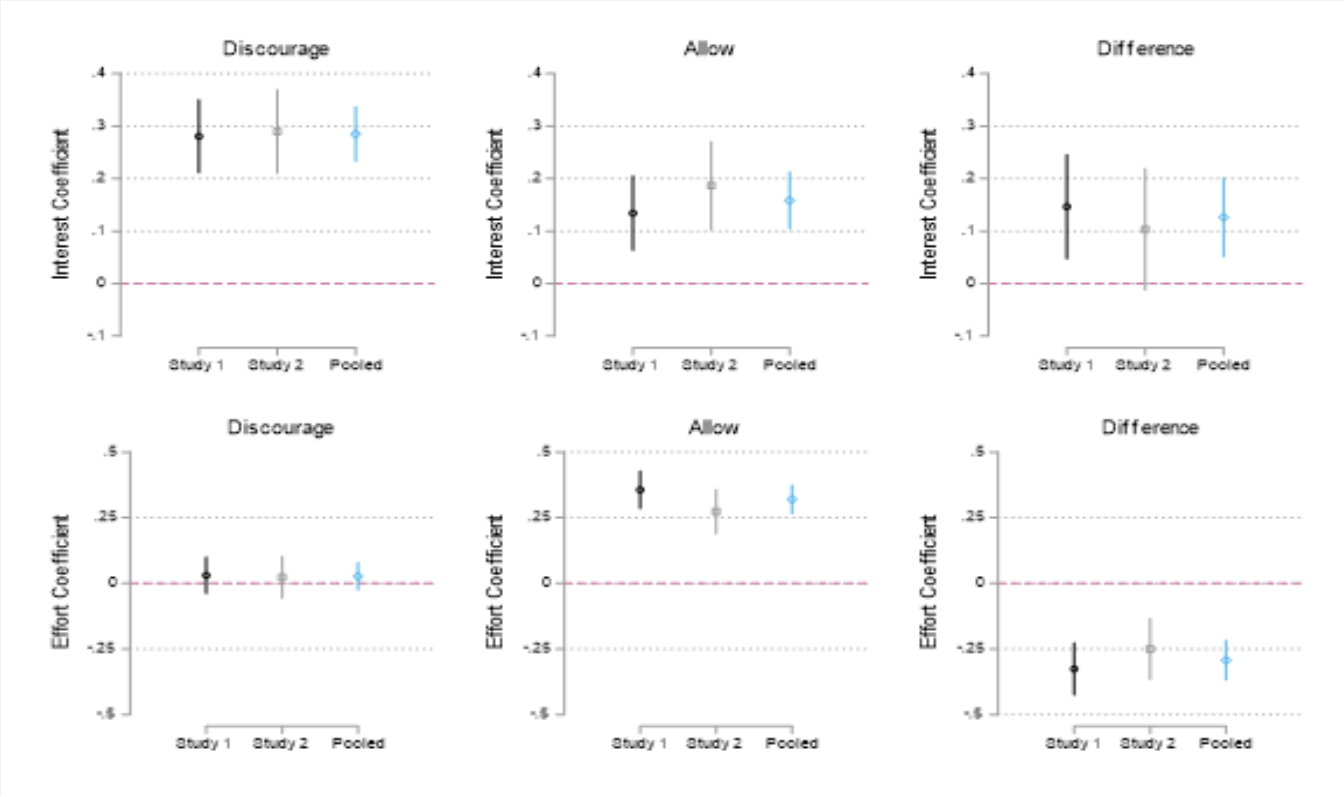
MANIPULATION CHECK

Figure 1. Search Rates and Knowledge Scores by Experimental Condition

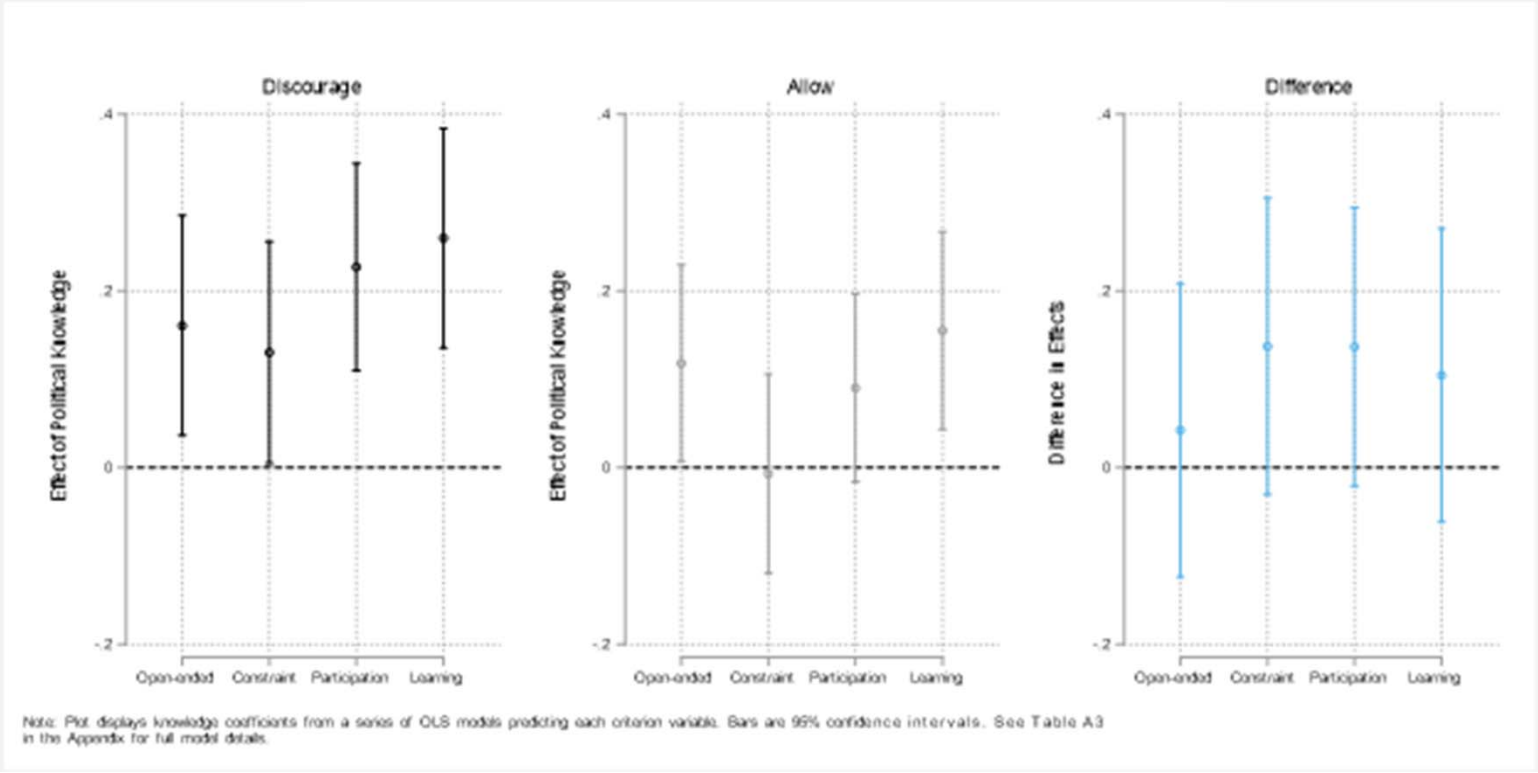


STUDIES 1 & 2 RESULTS
CONVERGENT AND DISCRIMINANT VALIDITY

STUDIES 1 & 2 RESULTS
CONVERGENT AND DISCRIMINANT VALIDITY



STUDY 2 RESULTS
PREDICTIVE VALIDITY

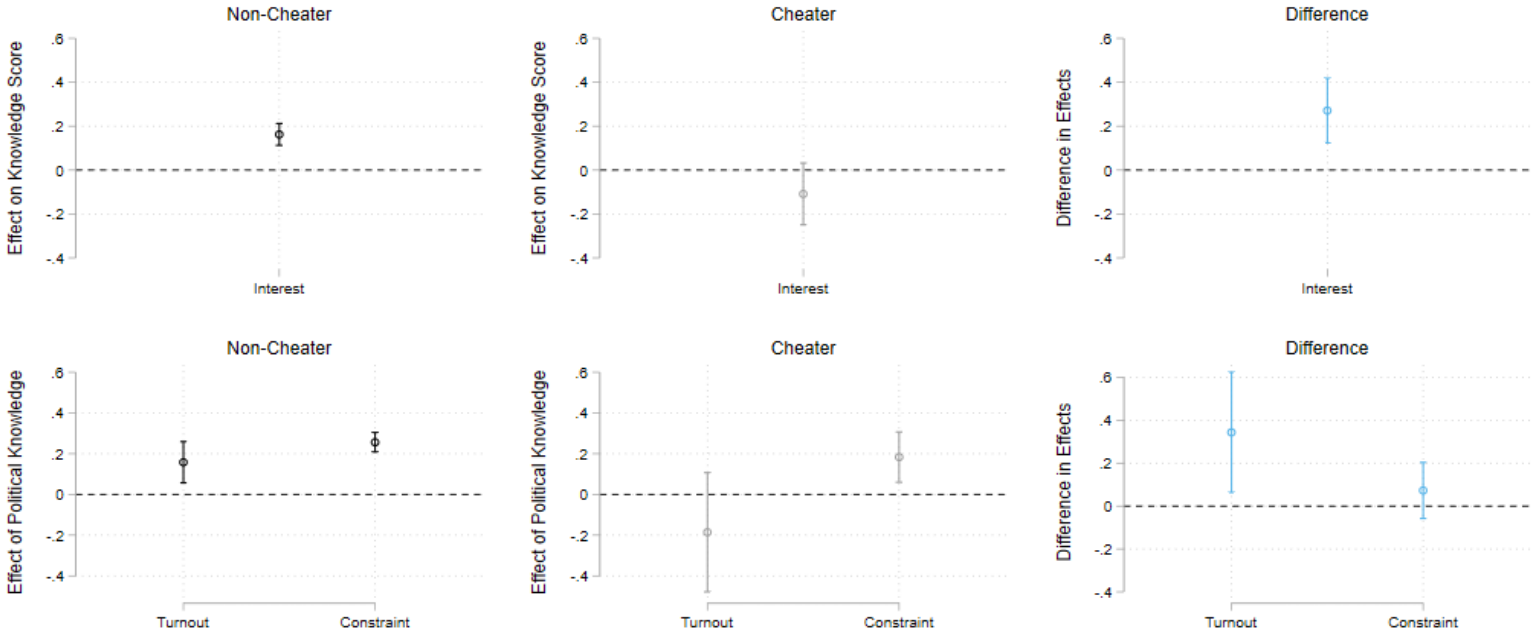


Note: Plot displays knowledge coefficients from a series of OLS models predicting each criterion variable. Bars are 95% confidence intervals. See Table A3 in the Appendix for full model details.

STUDY 3 MEASURES

- Measured cheating as it naturally occurs in an online survey.
- Respondents were instructed *not* to look up answers.
- Cheaters (identified with catch question; 11% of sample) spent twice as long on knowledge battery as non-cheaters and were more likely to get a perfect score on the 4-item scale (46% vs. 28%).

STUDY 3 RESULTS
CONVERGENT AND PREDICTIVE VALIDITY



Note: Upper row displays interest coefficients from OLS model predicting knowledge scores. Bottow 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 A5 and A6 in the appendix for full model details.

SO WHAT?

- When respondents look up answers, knowledge questions reflect the *effort* people are willing to put into the survey rather than their *latent* intellectual engagement with politics.

SO WHAT?

- When respondents look up answers, knowledge questions reflect the *effort* people are willing to put into the survey rather than their *latent* intellectual engagement with politics.

- It is hard to predict when outside search will be problematic because the behavior varies across samples *and within the same sample over time*.

SO WHAT?

- When respondents look up answers, knowledge questions reflect the *effort* people are willing to put into the survey rather than their *latent* intellectual engagement with politics.
- It is hard to predict when outside search will be problematic because the behavior varies across samples *and within the same sample over time*.
- Likely to increase as people become more comfortable with technology.

SO WHAT?

- When respondents look up answers, knowledge questions reflect the *effort* people are willing to put into the survey rather than their *latent* intellectual engagement with politics.
- It is hard to predict when outside search will be problematic because the behavior varies across samples *and within the same sample over time*.
- Likely to increase as people become more comfortable with technology.
- **Recommendation**: discourage outside search and diagnose this behavior when conducting online surveys.

REMAINING QUESTIONS/CHALLENGES

- Existing measures of online look up are based on *indirect* methods (self-report, catch question). At present, no *direct* measure has been employed widely.

REMAINING QUESTIONS/CHALLENGES

- Existing measures of online look up are based on *indirect* methods (self-report, catch question). At present, no *direct* measure has been employed widely.

- Understanding search behavior: which respondents engages in online look up and for what reasons?

Thank you!

APPENDIX MATERIALS

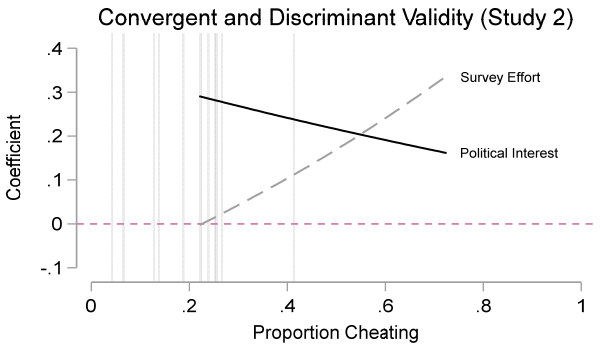
STUDY 2 KNOWLEDGE ITEMS

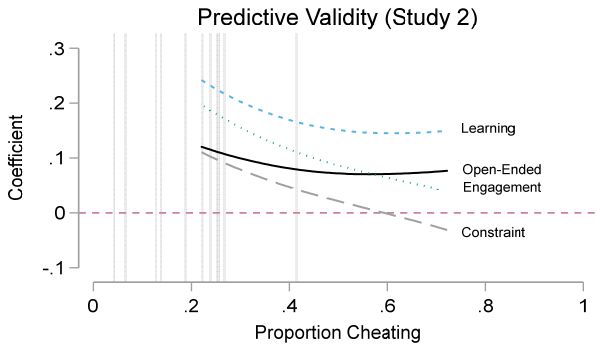
- 10 item scale (α= .68); all closed ended questions
- Length of senator's term; who is Chief Justice; who is PM of Great Britain; who is current Secy of State; which foreign country holds most U.S. debt; what is current unemployment rate; identify which country was recently added to travel ban list; why Secy of HHS (Tom Price) was recently in news; what does Trump's tax reform plan do to income brackets; how many members of U.S. Supreme Court are women

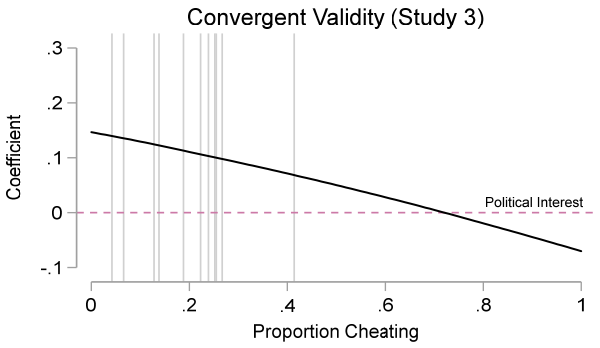
STUDY 3 KNOWLEDGE ITEMS

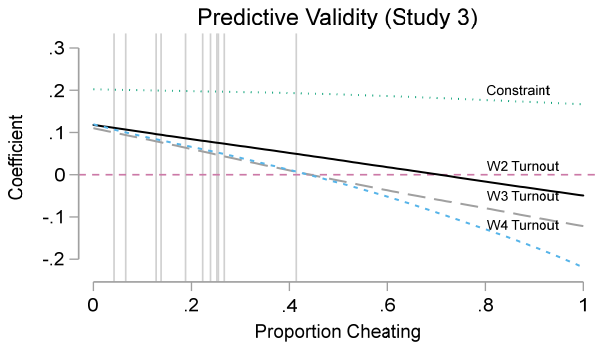
- 4 item scale (α= .58); all closed ended questions
- What is job/office held by Paul Ryan; which party controls U.S. House of Representatives; what is the length of U.S. Senator's term; who nominates judges to Federal Courts

SIMULATION EVIDENCE









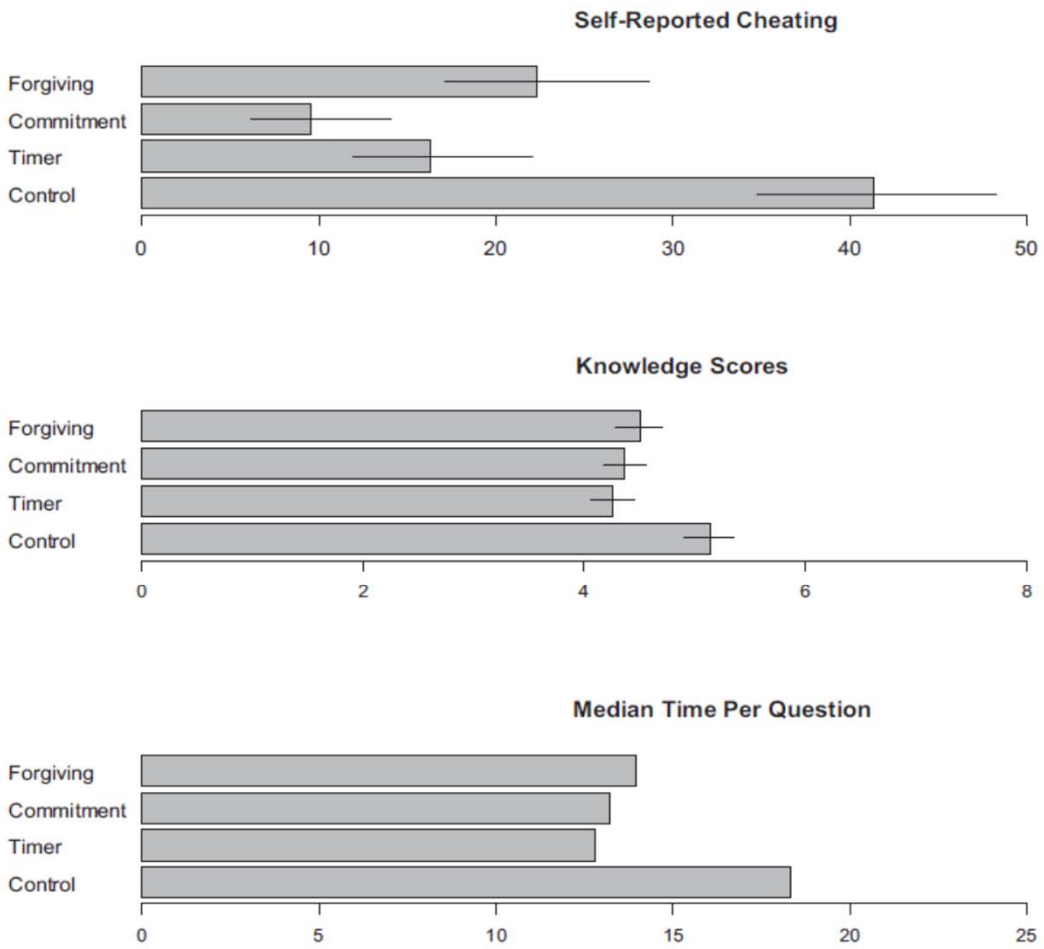


Figure 2. A Test of Three Methods to Reduce Cheating. Bar length represents the percentage of subjects self-reporting cheating (top), knowledge scores, measured as number of questions correct out of eight items (middle), and the median response time per question (bottom). Data comes from Student 2 sample ($N = 845$).

Correlates of Cheating in 2018 ANES Pilot Study

	Coefficient	S.E.	
Follows politics	.59	.29	*
Education	.74	.26	**
White	-.32	.16	*
Income	.42	.32	
Income Not Reported	.32	.21	
Age (< 35 years old)	.56	.21	**
Age (35 to 64 years old)	.65	.17	**
Trump Voter	.53	.15	**
Instructions Not to Cheat	-.64	.14	**
N	2500		

Note: Entries are logit coefficients and standard errors.
All variables standardized to 0-1 scale.
** $p < .01$ * $p \leq .05$ # $p < .10$

EXAMPLE OF VISUAL POLITICAL KNOWLEDGE QUESTION (PRIOR 2014)

Verbal-Only Condition

Visual-Only Condition

Verbal-and-Visual Condition

Note: This figure shows screenshots for the three different experimental conditions in Study 2 for one political knowledge question in the "Which office…" design. For a list of all questions, see Appendix Table A1 in the supporting information.