Thou shalt not cheat: how to reduce internet use in web surveys on political knowledge

CRISTIANO VEZZONI[1]* AND RICCARDO LADINI[2]

[1]*Department of Sociology and Social Research, University of Trento, Trento, Italy*
[2]*Doctoral School of Social Sciences, University of Trento, Trento, Italy*

By means of a split-ballot survey experiment, we study whether a normative instruction not to use the internet when answering political knowledge questions reduces cheating in web surveys. The knowledge questions refer to basic facts about the European Union and the data come from the Italian National Election Study web panel carried out in Italy before the 2014 European Election. Our analysis shows that a simple normative instruction significantly reduces cheating. We also show that reducing cheating is important to achieve a correct assessment of reliability of knowledge scales, while a decrease of cheating leaves unaltered the knowledge gap between lower and higher educated respondents. These results invite caution when including political knowledge questions in an online survey. Our advice is to include a normative instruction not to search the internet to reduce cheating and obtain more genuine answers. More generally, we conclude by stressing the need to consider the implications of online data collection when building questionnaires for public opinion research.

Keywords: cheating in web surveys; survey experiments; political knowledge; European Union; reliability and validity

Introduction

The rise of web surveys has raised concerns of potential non-random biases related to the measurement of political knowledge. In particular, attention has been paid to a phenomenon labelled as 'cheating in web surveys' (Jensen and Thomsen, 2014), when web survey respondents give a correct answer after having searched the internet for it. Thanks to search engines like Google, respondents can in fact find the right answer to a knowledge question in just a few seconds. In light of these considerations, researchers have started to explore strategies to overcome the problem of on-line cheating. The solutions proposed have been various in nature, with some researchers simply suggesting to ask respondents not to look up the answers on the internet.

In this study, we follow this hint, controlling the efficacy of a normative prompt dissuading web searches to reduce cheating when answering political knowledge questions. We do so by running a survey experiment manipulating the introduction to a battery of knowledge questions on basic facts about the European Union (EU).

* E-mail: Cristiano.vezzoni@unitn.it

After an introduction on political knowledge questions in web surveys, we describe the experimental design. We then present data, measures, and methods, we illustrate the results and we close the article with a discussion on the implications of our findings for the measurement of political knowledge in web surveys and, more generally, for online research on public opinion.

Political knowledge questions in web surveys

Political knowledge is commonly defined as 'a measure of a citizen's ability to provide correct answers to a specific set of fact-based questions' (Boudreau and Lupia, 2011: 171). Thus, political knowledge refers to the individual level of factual knowledge on political issues. Political knowledge is crucial in the study of political behaviour and public opinion, as Jensen and Thomsen (2014) highlight in their literature review: 193 studies published in four top journals of that strand of research (*Political Behavior*, *Political Psychology*, *Political Communication*, and *Public Opinion Quarterly*) over the time span 2002–12 have examined it as either a dependent or independent variable. Generally, political knowledge is not regarded as a concept *per se*, but is used as an indicator of political sophistication (Luskin, 1987) and the related concepts of political expertize, awareness, and involvement (Delli Carpini and Keeter, 1993).

Batteries of factual items are the most employed tool to measure the concept, although some studies underline how little individual motivation there is to answer those kinds of question correctly, with the associated risk of underestimating respondents' political knowledge (Prior and Lupia, 2008).

Further difficulties to address knowledge questions have emerged with the increased use of web surveys in public opinion research.[1] The well-known advantages of this data collection mode (Callegaro *et al*., 2015, 18–25) come with a cost, especially in terms of control over the interview environment. This shortcoming of online questionnaires is particularly relevant in the case of questions that require an answer that can be found on the internet.

Previous research has empirically tested the cheating behaviour of respondents in web surveys. In an experimental study which randomly allocates individuals either to a computer lab or to an online administration of the same questionnaire, Clifford and Jerit (2014) find that the rate of right answers to political knowledge items is significantly higher in the online condition, suggesting the presence of cheating when control over the respondent decreases. Similarly, Jensen and Thomsen find that a substantial 22% of the respondents to a web survey admit to having used the internet to answer questions on political knowledge.[2]

---

[1] In the above mentioned literature review, Jensen and Thomsen report that 23% of the studies examined are web based.

[2] The text of the question is the following: 'The Internet has made it much easier for ordinary people to get access to information about important questions. Many use the Internet on a regular basis. Therefore,

As to the strategies to overcome the problem of on-line cheating, Vavreck (2012) states that the main issue to be addressed is the low level of environmental control when administering on-line questionnaires. The presence of a controlled environment, however, is by definition in contrast with the web surveys design. Thus, Vavreck suggests testing the presence of cheating by downloading the browser histories of respondents. On one side, this strategy represents an effective way to detect cheaters but, on the other side, it is difficult to implement and does not work as an antidote to deceptive behaviour. Other scholars recommend introducing a time limit when answering political knowledge questions (Iyengar *et al.*, 2010; Strabac and Aalberg, 2011). Although partially successful, that strategy cannot prevent cheating, considering that in most cases online search for correct answers takes fewer seconds than any reasonable time limit could hope to prevent (Jensen and Thomsen, 2014). Recently, Munzert and Selb (2015) have tested the use of visual questions as an antidote against cheating. Nonetheless, their findings do not prove the efficacy of visual instruments in reducing cheating.

In the face of such difficulties, Shulman and Boster propose simpler and more straightforward advice: that some 'suggestions to discourage online cheating [should] include adding information not to look up answers online' (2014: 187). The strength of such a recommendation finds its foundation in the idea that an individual will behave consistently with the implicit commitment taken when exposed to the normative message, reducing potentially problematic behaviour (Cialdini, 1984: 51–57; Schultz *et al.*, 2007). Following this suggestion, by means of a split-ballot web survey experiment our study analyses the effect of the introduction of a normative prompt dissuading web searches when answering political knowledge questions.

Experimental design and hypotheses

The respondents to a web survey ($N = 3243$) are assigned either to a treatment group or to a control group through a simple randomization. The experiment manipulates the preamble to a battery of three political knowledge items. Next to a neutral introduction to the political knowledge questions, the treated group receives a normative instruction that invites respondents not to search the internet for the correct answers. The text of the two parts is as follows:

Neutral introduction: Finally, we will propose some questions on political knowledge.
Normative instruction: We ask you to answer without searching the internet.

The precise formulations of the experimental conditions are as follows:
Control   (Neutral introduction)
Treatment   (Neutral introduction) + (Normative instruction)

we would like to know if you used the Internet when answering one or more of the previous four questions?' (Jensen and Thomsen, 2014: 3348).

The first aim of the experiment is to test whether such a simple normative instruction is successful in reducing cheating behaviour. We thus expect that in the treatment group the proportion of correct answers to the knowledge questions is lower, as the number of people giving correct answers is not inflated by cheaters. Assuming that neither treatment decreases knowledge nor its absence increases it, the difference between the two groups could be fully attributed to the deterrent effect against cheating of the normative prompt. When treated, fewer respondents search the internet and consequently fewer respondents give a correct answer without knowing it.

Second, we control the impact of cheating on the reliability and validity of an additive knowledge scale. Although a good measurement should be both reliable and valid (Carmines and Zeller, 1979), we hypothesize a paradoxical outcome with increasing reliability coupled with a degradation of validity when cheating is more widespread (control group). In fact, more cheating produces a higher proportion of respondents with full scores (all correct answers), which artificially inflates inter-item correlations between the knowledge items. However, for cheaters, those high scores do not indicate higher knowledge, jeopardizing the validity of the scale.

Finally, we look at whether cheating alters the relation between knowledge and socio-demographic dimensions, focussing on education. Previous research based on self-reported cheating found that lower educated respondents show a higher probability (of reporting) of having searched the internet to answer knowledge questions in web surveys (Jensen and Thomsen, 2014). Consequently, it has been argued that in web surveys the educational gap in political knowledge could be underestimated, as lower educated respondents compensate for their lack of knowledge by booting up the internet. Our expectation is therefore that the distance in knowledge performance between educational groups is larger (and more genuine) in the treatment condition, where cheating is reduced by the effect of the normative prompt.

Data, measures, and methods

Data come from the third wave of the Italian National Election Study (ITANES) on-line panel, which spans over the Italian electoral cycle 2013–15. The survey was carried out immediately before the 2014 European Parliament (EP) elections, which took place on the 25 May 2014.

The sample is made of 3243 individuals, selected from an opt-in community group of a private research company (SWG). It is a non-probabilistic sample aiming to reproduce the quotas for gender, age, and territorial distribution of the Italian population.

Although the sample non-representativeness can be an issue when aiming at producing inferences to the general population, in this paper we focus on a specific cognitive mechanism and we overcome the sample weaknesses by a randomized experimental design.

The questionnaire includes a three-item battery on knowledge of EU (political) matters, as in De Vreese and Boomgaarden (2006). The reason to test EU knowledge rather than general (or national) political knowledge is connected to the framework of the survey, aimed at studying electoral behaviour in EP elections. The questions were as follows:

Item A: How many countries are members of the EU? *(Correct answer: 28)*
Item B: Who is the European People's Party candidate for the presidency of the European commission? *Multiple choice, 5 options (Correct answer: Juncker)*
Item C: Who is the Party of European Socialists candidate for the presidency of the European commission? *Multiple choice, 5 options (Correct answer: Schulz)*

The three items differ both in answer mode and content (Barabas *et al.*, 2014). As far as content is concerned, while item A tests general knowledge, also defined as textbook knowledge (Jennings, 1996), items B and C address surveillance knowledge, relative to current events. Looking at the answer mode, the question on the number of countries belonging to EU (item A) is open ended with a short numeric answer, while the other two questions (items B and C) are multiple choice. This implies that respondents have higher chances of guessing the correct answer for those items (Shulman and Boster, 2014). The option 'don't know' was allowed only for item A (filling '9999', as indicated in the text of the question). For items B and C, respondents were forced to give an answer and 'don't know' was not allowed (Mondak and Davis, 2001).[3]

In the coming analysis, EU political knowledge is measured considering the answers to each item separately and by an additive scale reporting for each respondent the sum of correct answers to the three knowledge questions.

The effect of treatment will be tested through the comparison of the aggregate proportions of correct answers in the two experimental groups.

In the assessment of the impact of cheating on the quality of the knowledge scale measure we will perform reliability analysis and we will consider construct validity.

As for reliability, Pearson's correlations[4] among the three knowledge items and their internal consistency (Cronbach's α) will be calculated for the two groups. If cheating biases the performance in a non-random way, thereby increasing the chances of giving correct answers to all the items, we expect that pairwise correlations and Cronbach's α will be higher in the control group.

As for construct validity,[5] we will compute Pearson's correlations between the knowledge scale and two theoretically related variables (Atkin *et al.*, 1976; Liu and Eveland, 2005; Eveland and Hively, 2009): interest in the election outcome

---

[3] Full description of the items and distributions are available in the online Appendix.

[4] In 2×2 tables, the Pearson's correlation coefficient coincides with the φ coefficient for two binary variables.

[5] Construct validity is defined as 'the extent to which a particular measure relates to other measures consistent with theoretically derived hypotheses concerning the concepts (or constructs) that are being measured' (Carmines and Zeller, 1979: 23).

measured by an 11-point scale and frequency of political discussion in the last month by a 5-point scale (0: never, 4: every day). In both cases, we expect positive correlations which increase in the treatment group, less biased by cheating. Finally, the impact of education on cheating will be tested by means of a linear regression (dependent variable: additive knowledge scale), adding interaction terms between the treatment condition (normative instruction) and education (three categories: primary, secondary, tertiary), controlling for gender and age (three categories: 18–34, 35–54, 55 and older). A positive interaction coefficient indicates that the gap between educational groups increases in the treatment condition, that is when cheating is reduced.

Results

Comparing the proportion of correct answers in the two experimental groups, it turns out that the performance on knowledge items is significantly lower in the treatment group. Assuming that neither treatment decreases knowledge nor its absence increases it, the difference between the two groups could be fully attributed to the deterrent effect against cheating of the normative prompt. Table 1 shows that this holds for all the items (P-values < 0.01).[6] Thus, the simple strategy of adding a normative instruction not to look up answers on the internet turns out to work as a successful antidote to cheating.

Considering the different items, the open-ended question produces larger differences between the two experimental groups (10 percentage points, against 6 and 7 percentage points, respectively, for the items B and C). However, it is difficult to say whether this difference should be attributed either to the answer mode or to the difficulty of the item, being A also the most difficult item.[7]

The last row of Table 1 shows the proportions of respondents who answered correctly to all the three questions on EU political knowledge. Again the difference between the two experimental groups is highly significant (P-value < 0.01) and in the expected direction. This outcome brings us directly to the next step of our analysis that is the evaluation of the impact of cheating on the quality of the knowledge measurement.

As stated in the previous section, cheating can affect correlations between the different knowledge items and consequently influence the reliability of the scale built using those items. Our findings are in line with these expectations. Pairwise correlations between knowledge items are always higher in the control group

---

[6] Regardless of the experimental condition the three knowledge questions meet the criterion introduced by Delli Carpini and Keeter (1993: 1187), who suggest avoiding items either too easy (correct answers >90%) or too difficult (correct answers <10%). A complete distribution of the answers to the three knowledge questions, including the distribution of wrong answers, is supplied in Appendices A1 and A2.

[7] An experimental study considering knowledge questions with different levels of complexity did not find an association between cheating activities and difficulties of the items (Clifford and Jerit, 2016).

Table 1. Proportions of correct answers on political knowledge items in the two groups

| | % Correct answers | | | |
	Control	Treatment	χ^2 (1)	P-value
Item A: number of EU countries	35.0	24.5	42.8	0.000
Item B: EPP candidate	50.3	44.2	12.0	0.001
Item C: PES candidate	66.8	59.7	17.8	0.000
All three items correct	22.5	13.7	41.6	0.000
N	1631	1612		

EU = European Union; EPP = European People's Party; PES = Party of European Socialists.

Table 2. Item correlations for knowledge questions, Cronbach's α of the additive knowledge scale and Z-test on differences between the two experimental groups

Item combinations	Questions	Control (N = 1631)	Treatment (N = 1612)	Two-tailed Z-test[a]	Significance (two-tailed)
A–B	EU countries – EPP candidate	0.30	0.23	1.99	0.047
A–C	EU countries – PES candidate	0.26	0.16	2.92	0.003
B–C	EPP candidate – PES candidate	0.39	0.37	0.93	0.354
				Two-tailed W-test[b]	
A–B–C	Cronbach's α	0.58	0.50	0.84	0.000

EU = European Union; EPP = European People's Party; PES = Party of European Socialists.
[a]Differences between correlation coefficients tested using Fisher r-to-Z transformation.
[b]Differences between two Cronbach α's tested using the Feldt test (Feldt, 1969).

(see Table 2), although the difference is not statistically significant for the item combination B–C. The same applies for the internal consistency of the knowledge scale (Cronbach's α, last row of Table 2). These findings bring us to an immediate consideration: when a battery of political knowledge items is included in a web survey, the reliability analysis can lead to misleading conclusions about the quality of measurement, since cheating behaviours artificially increase the internal consistency of the scale.

The degradation of the validity of the measurement when cheating is more diffuse should represent the other side of the coin. However, the results in terms of construct validity are not as clear as the ones on reliability. Although political knowledge is significantly correlated in the expected theoretical direction with the two indicators used to test construct validity (interest in the election outcome and frequency of political discussion), differences between the two experimental groups
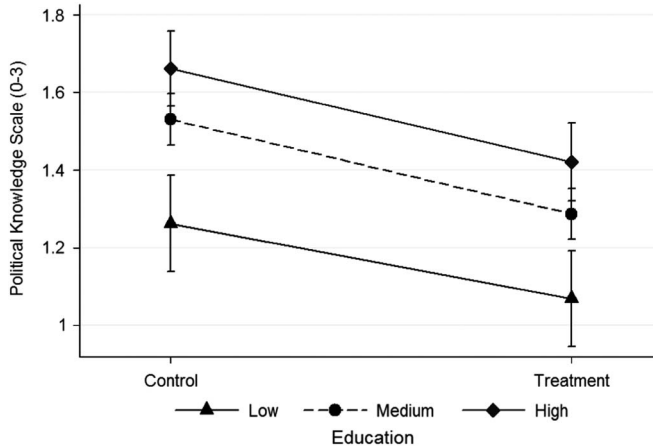
Figure 1 Predicted means of knowledge scores by experimental condition and level of education from the previous regression analysis (95% confidence intervals).

are not significant (Table A3 in the Appendix). Therefore, there is no empirical evidence supporting the idea of a decreasing validity when cheating is more widespread.

The last step of our analysis concerns the study of the impact of cheating on the relation between knowledge and education. To do so, we put our knowledge scale within a regression framework where education is interacted with the experimental condition (reference category = control group),[8] controlling for gender and age.

If cheating is more widespread among lower educated people, cheating could compensate for lack of knowledge in that group and accordingly reduce the effect of education on the additive knowledge scale. Conversely, a reduction in cheating (which actually happens in the treatment condition) should lead to an increase in the knowledge gap between education groups. If this is the case, we should find a positive interaction effect between education and treatment and increased distances in the knowledge scores for the different educational groups in the treatment condition.

No empirical evidence supporting these expectations emerges from the analysis. Although the main effects for education, as well as for the other socio-demographic variables, go in the expected direction (more education enhances knowledge, men know more than women, and older people know more than the younger), the interaction terms are not significantly different from zero, indicating that the effect of education does not increase in the treatment group. This outcome becomes clearer if we look at Figure 1: the three parallel lines indicate that the distance in

---

[8] A similar regression approach has been proposed for the study of social desirability in the framework of list experiments, where it is not known how each respondent reacts to a sensitive item presented in a longer list of control items. See for example, Janus (2010).

performance on the knowledge scale between groups of respondents characterized by different levels of education remains the same in both the treatment and control group. That is, a reduction in cheating in the treatment group does not increase the knowledge gap between lower and higher educated respondents. Full regression results are presented in the Appendix A4.

Discussion and conclusions

Measuring factual (political) knowledge in web surveys is a difficult business, as respondents who do not know the correct answer to a question can easily find it by searching the internet. What is defined as 'cheating in web-surveys' (Jensen and Thomsen, 2014) is therefore harmful for our knowledge measurement, potentially affected by non-random biases.

Our study explored the effectiveness of introducing simple normative instructions to reduce cheating in web surveys. We did so by implementing a split-ballot experiment on a sample of Italian citizens coming from a web-based electoral survey carried out before the EP election 2014 ($N > 3000$). While the interviewees belonging to the control group received only a neutral introduction to a battery of knowledge questions on basic European facts, those belonging to the treatment group were also presented with a normative prompt inviting them to answer the questions without searching the internet. The results are straightforward as the group of treated respondents consistently showed lower percentages of correct answers (to a maximum of 10 percentage points difference on a question concerning the number of countries in the EU). Realistically assuming that the administration of the normative instruction did not decrease the knowledge of respondents, the worse results in the treated group can only be attributed to a decrease in cheating following the invitation not to search the internet. This outcome highlights two important elements:

- cheating in web surveys is a widespread phenomenon;
- simple normative instructions work effectively to reduce cheating.

Prevention of cheating is relevant given that such behaviour affects the properties of our measures, potentially resulting in misleading evaluations. We indeed found that the correlations between knowledge items as well as the reliability of the additive knowledge scale artificially increases when cheating is more diffuse. We did not find any evidence, otherwise, of a degradation of the validity of our measure due to cheating. The reason for this negative finding can be linked to the fact that the battery of items used in the experiment is not as well-established nor widely tested as the one recommended by Delli Carpini and Keeter (1996). For more compelling results, then, we should wait for further research based on more established measures of political knowledge.

Finally, we did not find support for the hypothesis that suggests that cheating is more diffuse among lower educated respondents and consequently that cheating

enhances their performances, closing the gap with more educated respondents (Jensen and Thomsen, 2014). This outcome is coherent with a reading that suggests, on one hand, that cheating is a crosscutting phenomenon, whilst, on the other, that the effectiveness of normative instructions not to cheat is seemingly effective on all educational groups. Our results signal that previous findings showing a higher prevalence of cheating among lower educated respondents could be an artefact produced by the use of self-reported measures of cheating. In fact, higher educated people could be more vulnerable to social desirability and thus report cheating less often.

Of course, our analyses are not without shortcomings. The main limitation of our study is connected to the fact that we do not have individual level measures on the actual presence of the cheating behaviour. Our conclusions are therefore drawn at the aggregate level and the reduction of cheating due to the normative instruction is deduced by the comparison of the average performances on knowledge questions in the control and treatment group. The actual amount of cheating remains unknown as well as whether a person has cheated or not. This does not jeopardize our conclusion on the effectiveness of normative instructions against cheating, but it does hinder any further analysis aimed at studying the relation between cheating and individual characteristics. Moreover, any disputes over compliance with normative instructions and the effects of social desirability cannot be conclusively settled with our data since a valid indicator of cheating at the individual level is not available. Further research will be necessary to deepen these aspects. Finally, it is important to remember that our conclusions come from an intent-to-treat analysis and it is not possible to be sure that respondents actually read and considered the normative instruction (Berinsky et al., 2014).

To overcome some of these weaknesses, a suggestion could be to couple split-ballot experiments like ours with the administration in the same interview of a question on self-reported cheating to triangulate the results (Clifford and Jerit, 2016). A further articulation of this advice could go in the direction of asking about cheating behaviour not directly but by means of a list experiment (Blair and Imai, 2012) to be able to better estimate the magnitude of the phenomenon and to assess the effect of social desirability on the reports. Following the guideline suggested by Munzert and Selb (2015), the measurement of the response latencies (i.e. the time used to answer a question) could represent another strategy to detect cheating, allowing for a more precise assessment of the impact of the normative instruction.

All things considered, we still maintain that the main result of our study is robust: a normative instruction not to search the internet for help with answering knowledge questions is an effective tool to reduce cheating in web surveys. Thus, the advice when administering knowledge questions in web surveys is always to use a cheap and non-intrusive tool, such as this, in order to obtain more genuine results.

What we have shown in this article pertains to a specific kind of question meant to measure the knowledge of respondents. The peculiarity of these questions is that they usually only have one correct answer, either to be guessed (open ended) or

chosen between a list of options (multiple choice). In the case of web surveys, where the control of the researcher over the interviewee is absent, these questions are particularly vulnerable to cheating behaviour. Nonetheless, the modality in which data are collected does not only affect knowledge questions. Each measure conceived to tap into a certain concept and included in an online questionnaire should be adapted to that data collection mode, considering that a self-administered online questionnaire largely differs from a face-to-face or telephone interview. Thus, the results invite us to pay further attention to this issue, promoting a broader use of survey experiments to enhance our understanding of the cognitive mechanisms at work beneath the activity of answering structured questionnaires (Sirken *et al.*, 1999). This could help in the process of calibration and standardization of measurement instruments in public opinion research, improving both their reliability and validity.

Conflicts of Interest

None.

Data

The replication data set is available at http://thedata.harvard.edu/dvn/dv/ipsr-risp

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1017/ipo.2016.25

References

Atkin, C.K., J. Galloway and O.B. Nayman (1976), 'News media exposure, political knowledge and campaign interest', *Journalism & Mass Communication Quarterly* **53**(2): 231–237.

Barabas, J., J. Jerit, W. Pollock and C. Rainey (2014), 'The question (s) of political knowledge', *American Political Science Review* **108**(4): 840–855.

Berinsky, A.J., M.F. Margolis and M.W. Sances (2014), 'Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys', *American Journal of Political Science* **58**(3): 739–753.

Blair, G. and K. Imai (2012), 'Statistical analysis of list experiments', *Political Analysis* **20**(1): 47–77.

Boudreau, C. and A. Lupia (2011), 'Political knowledge', in J.N. Druckman, D.P. Green, J.H. Kuklinski and A. Lupia (eds), *Cambridge Handbook of Experimental Political Science*, Cambridge: Cambridge University Press, pp. 171–183.

Callegaro, M., K. Lozar Manfreda and V. Vehovar (2015), *Web Survey Methodology*, London: Sage.

Carmines, E.G. and R.A. Zeller (1979), *Reliability and Validity Assessment*, Beverly Hills, CA: Sage publications.

Cialdini, R.C. (1984), *Influence. The Psychology of Persuasion*, New York: Quill.

Clifford, S. and J. 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**(2): 120–131.

Clifford, S. and J. Jerit (2016), 'Cheating on political knowledge questions in online surveys: An assessment of the problem and solutions', *Public Opinion Quarterly* **80**(4): 858–887.

Delli Carpini, M.X. and S. Keeter (1993), 'Measuring political knowledge: putting first things first', *American Journal of Political Science* **37**(4): 1179–1206, doi: 10.2307/2111549

Delli Carpini, M.X. and S. Keeter (1996), *What Americans Know About Politics and Why it Matters*, New Haven, CT: Yale University Press.

De Vreese, C.H. and H.G. Boomgaarden (2006), 'News, political knowledge and participation: the differential effects of news media exposure on political knowledge and participation', *Acta Politica* **41**(4): 317–341.

Eveland, W.P. Jr. and M.H. Hively (2009), 'Political discussion frequency, network size, and "heterogeneity" of discussion as predictors of political knowledge and participation', *Journal of Communication* **59**(2): 205–224.

Feldt, L.S. (1969), 'A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two tests', *Psychometrika* **34**(3): 363–373.

Iyengar, S., J. Curran, A.B. Lund, I. Salovaara-Moring, K.S. Hahn and S. Coen (2010), 'Cross-national versus individual-level differences in political information: a media systems perspective', *Journal of Elections, Public Opinion and Parties* **20**(3): 291–309.

Janus, A.L. (2010), 'The influence of social desirability pressures on expressed immigration attitudes', *Social Science Quarterly* **91**(4): 928–946.

Jennings, M.K. (1996), 'Political knowledge over time and across generations', *Public Opinion Quarterly* **60**(2): 228–252.

Jensen, C. and J.P.F. Thomsen (2014), 'Self-reported cheating in web surveys on political knowledge', *Quality & Quantity* **48**(6): 3343–3354.

Liu, Y.I. and W.P. Eveland Jr. (2005), 'Education, need for cognition, and campaign interest as moderators of news effects on political knowledge: an analysis of the knowledge gap', *Journalism & Mass Communication Quarterly* **82**(4): 910–929.

Luskin, R.C. (1987), 'Measuring political sophistication', *American Journal of Political Science* **31**(4): 856–899.

Mondak, J.J. and B.C. Davis (2001), 'Asked and answered: knowledge levels when we will not take "don't know" for an answer', *Political Behavior* **23**(3): 199–224.

Munzert, S. and P. Selb (2015), 'Measuring political knowledge in web-based surveys: an experimental validation of visual versus verbal instruments', *Social Science Computer Review*. First published online on November 17, 2015: 1–17.

Prior, M. and A. Lupia (2008), 'Money, time, and political knowledge: distinguishing quick recall and political learning skills', *American Journal of Political Science* **52**(1): 169–183.

Schultz, P.W., J.M. Nolan, R.B. Cialdini, N.J. Goldstein and V. Griskevicius (2007), 'The constructive, destructive, and reconstructive power of social norms', *Psychological Science* **18**(5): 429–434.

Shulman, H.C. and F.J. Boster (2014), 'Effect of test-taking venue and response format on political knowledge tests', *Communication Methods and Measures* **8**(3): 177–189.

Sirken, M.G., D.J. Herrmann, S. Schechter, N. Schwarz, J.M. Tanur and R. Tourangeau (eds) (1999), *Cognition and Survey Research*, New York: Wiley.

Strabac, Z. and T. Aalberg (2011), 'Measuring political knowledge in telephone and web surveys: a cross-national comparison', *Social Science Computer Review* **29**(2): 175–192.

Vavreck, L. (2012), 'The myth of cheating on self-completed surveys'. Retrieved 1 December 2016 from http://today.yougov.com/news/2012/04/17/myth-cheating-self-completed-surveys/
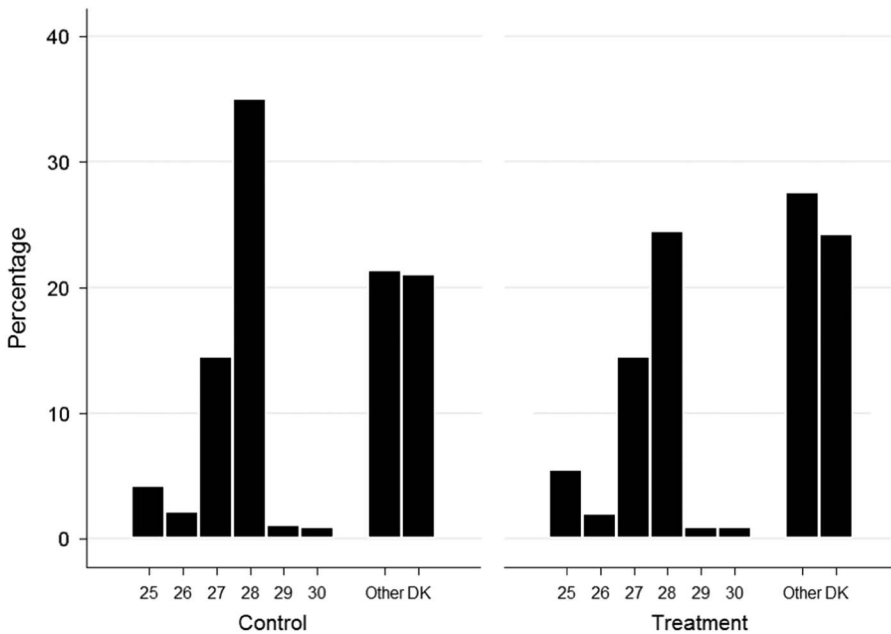
Appendix

Figure A1 Distribution of the answers to item A by experimental condition. Item A: How many countries are members of the European Union?

Table A1. Political knowledge items: question text, answer mode, and answer categories

Question formulation	Answer mode	Answer categories	Correct answer	'Don't know' option
Item A: How many countries are members of the European Union?	Short numeric	Open ended	28	9999
Item B: Who is the European People's Party candidate for the presidency of the European commission?	Multiple choice	Verhofstadt Keller Schulz Tsipras Juncker	Junker	Not allowed
Item C: Who is the Party of European Socialists candidate for the presidency of the European commission?	Multiple choice	Verhofstadt Keller Schulz Tsipras Juncker	Schulz	Not allowed

Table A2. Distributions of the answers to item B and item C by experimental condition

| | Item B | | | Item C | |
Candidates	Control	Treatment	Candidates	Control	Treatment
Verhofstadt	8	8	Verhofstadt	5	7
Keller	6	6	Keller	6	7
Schulz	27	32	Schulz	67	60
Tsipras	9	9	Tsipras	9	12
Juncker	50	44	Juncker	13	15
N	1631	1612	N	1631	1612

Item B: Who is the European People's Party candidate for the presidency of the European commission? Item C: Who is the Party of European Socialists (PES) candidate for the presidency of the European commission? Column percentages.
The options in boxes represent the correct answer.

Table A3. Correlations between knowledge scale and other related variables and two-tailed Z-tests on the differences of the correlations between the experimental groups

Variables	Control	Treatment	Z-test	N
Interest in the electoral outcome (0–10)	0.15	0.21	1.57	3060
Frequency of political discussion (0–4)	0.19	0.20	0.31	3150

Table A4. The impact of demographics on political knowledge ($N = 3242$, 1 missing value for education)

Variables	Categories	Coefficients	SE	Significance (two-tailed)
Constant		1.27	(0.07)	(0.000)
Experimental condition (ref.: control)	Treatment	−0.19	(0.09)	(0.028)
Gender (ref.: male)	Female	−0.21	(0.04)	(0.000)
Education (ref.: low)	Medium	0.27	(0.07)	(0.000)
	High	0.40	(0.08)	(0.000)
Age (ref.: 18–34)	35–54	0.06	(0.05)	(0.162)
	>54	0.20	(0.05)	(0.000)
Interaction terms				
Experiment condition × education	Treatment × medium	−0.05	(0.10)	(0.622)
	Treatment × high	−0.05	(0.11)	(0.678)
Adjusted R^2		0.04		

Unstandardized ordinary least square regression coefficients: SE and P-values in parentheses.