

Reducing overreporting of voter turnout in seven European countries – results from a survey experiment

Sylvia Kritzinger (University of Vienna, sylvia.kritzinger@univie.ac.at)

Steve Schwarzer (TNS Opinion, Brussels, steve.schwarzer@tns-opinion.com)

Eva Zeglovits (University of Vienna, eva.zeglovits@univie.ac.at)

-- First Draft --¹

Abstract

The phenomenon of overreporting presents a challenge to scholars who analyze voter turnout using surveys. There have been several attempts to reduce overreporting by introducing new ways of question wordings and a more extended list of response options. Although it is known that overreporting is sensitive to context variables, we still lack knowledge, how reducing overreporting works in different contexts.

Thus, in this paper we study how the different ways of question wordings work in different contextual settings and which impact this has on comparative survey research. To do so, we conducted a survey experiment (web survey) including the usual turnout question and two new question forms in 7 European countries in December 2011. We find that (1) it is possible to

¹ Paper was presented at the AAPOR 67th General Conference, Orlando, Florida, May 2012

reduce self-reported turnout even in online surveys using alternative question forms, and (2) that this reduction works differently in the investigated countries.

1. Introduction

Information on voter turnout is one of the central units of analysis when studying electoral behavior in liberal democracies. Most research on turnout is based on data stemming from survey research. However, the problem with survey data is that they might not reflect the actual behavior of the respondents: respondents ‘overreport’ on turnout, meaning that they report turnout but did actually not vote. Thus, we are confronted with the phenomenon that the proportion of respondents that report that they did vote is actually higher than the actual turnout in election (e.g. Traugott & Katosh, 1979). Taken this further, analyses based on this data might in the following produce biased results and conclusions.

There is broad evidence for different sources of these too high shares of reported turnout in surveys (see Holbrook & Krosnick, 2011 for an overview) ranging from address sampling or coverage errors (Granberg & Holmberg, 1991; Traugott & Katosh, 1979) to measurement error (‘overreporting’). In this paper, we will address these latter measurement errors.² Here, we have to distinguish between unintentional misreporting (e.g. memory errors) (Belli, Moore, & VanHoewyk, 2006; Belli, Traugott, Young, & McGonagle, 1999; Stockè, 2007; Stockè & Stark, 2007) and intentional misreporting (Bernstein, Chadha, & Montjoy, 2001; Silver, Anderson, & Abramson, 1986; Stockè & Stark, 2007). In the latter case, respondents report one behavior although they are conscious that they performed another behavior. Thus, intentional overreporting means that people state that they casted their ballot, but in fact they are aware that they abstained.³

² Thus, when talking about overreporting in this paper, we refer to measurement errors.

³ In contrast, underreporting has been identified as a minor problem (e.g. Traugott & Katosh 1979; Abelson, Loftus & Greenwald 1992; Belli, Traugott, Young, & McGonagle, 1999).

Why do people overreport turnout intentionally? Here, social desirability is regarded as the main driver. Non-voters would like to reduce cognitive dissonance of their behavior: they are either anxious to please the interviewers or themselves (Bernstein, et al., 2001). Given this problem, there have been several attempts to reduce the problem of overreporting induced by social desirability: new ways of wording questions and/or diversified response options were elaborated and tested. Thus by now, most electoral studies introduce the turnout question with some stimulus that aims at reducing social desirability, but its success can be doubted (Abelson, Loftus, & Greenwald, 1992). Meanwhile, Belli et al. (1999) and Belli, Moore, and VanHoewyk (2006) developed a question format presenting three different response options to report non-voting which proved to reduce overreporting substantially. More recently, another diversification of response options was successfully tested, where respondents can report that they cannot remember for sure if they voted or not (Zeglovits & Kritzing, 2011).

However, respondents also act within their environment. Thus, overreporting could also be context dependent. Indeed, there is some general evidence of country differences in overreporting: Karp and Brockington (2005) did a comparative analysis on overreporting in countries where vote choice can be validated. They found two important context characteristics that increase intentional overreporting: overall turnout and saliency of the elections. Granberg and Holmberg (1991) compared reported and validated turnout in Sweden and in the US and found that the percentage of non-voters reporting that they voted was nearly the same in these two countries.

However, to our knowledge there does not exist any comparative study which attempts to analyze whether new ways of question wordings and response options work equally well in reducing overreporting in different countries. In other words, we do not know if an approach that

successfully reduces overreporting in one country will also be successful in another country. If attempts to reduce overreporting are context-sensitive, it is particularly for comparative research projects important to know which differences have to be taken into account and which questions therefore should be used.

Thus, in this paper we study how the different ways of question wordings work in different contextual settings. In particular, we examine if and how far former results on reducing overreporting in the voter turnout question can be generalized to different settings. Thus, the paper takes an exploratory approach. The paper proceeds as follows: We start with discussing different attempts to reduce overreporting by varying question forms, by taking individual and context specific causes of overreporting into account. Next, we introduce the design of our experiment, the data collection and the question forms. After presenting our results, we discuss the conclusions of our findings.

2. Varying question formats to reduce overreporting

Attempts that aim at reducing intentional overreporting have to take the cognitive dissonance of respondents into account. If cognitive dissonance can be reduced, then, the respondent is more likely to admit non-voting. As pointed out, quite some attempts to reduce overreporting have been initiated. Most prominently, election studies include an introduction sentence that is meant to reduce social desirability of reporting turnout. For instance, the European Social Survey (round 5) uses the statement “Some people don’t vote nowadays for one reason or another“, while the American National Election Study (ANES) uses “In talking to people about elections, we often find that a lot of people were not able to vote because they were not registered, they

were sick, or they just didn't have the time". In both case, the introductory sentence is followed by a simple and direct question on whether the respondent turned out or not ("yes" versus "no"). Although, there are findings that suggest that these attempts do not reduce overreporting (Abelson, et al., 1992), they are nevertheless most commonly used.

Also Belli et al. (1999, 2006) dealt with the problem of overreporting and tried to capture both the problem of memory failures and intentional misreporting. To circumvent the first problem, they included a long explanatory statement to assist respondents in remembering the election of interest (Belli, et al., 1999).⁴ For the second problem, they presented more options to report non-voting by including three different ways of reporting non-voting (Belli, et al., 2006; Belli, et al., 1999). In addition to simply reporting that one did not vote, respondents were also given the possibility to answer (1) that they usually vote, but this time did not, and (2) that they thought about voting, but then did not vote. Instead of only one non-voting option, three were presented to the respondents. The additional two measures reduced reported turnout substantially compared to the standard question in the ANES.

The idea of diversifying response options to make it easier for respondents to admit non-voting was taken up in a single country study (Austria). Zeglovits and Kritzinger (2011) tested a new form of response options. They adopted the idea of the propensity to turn out used in pre-election studies to post-election scenarios and self-reported turnout. Respondents received four possible response options, ranging from being sure that one did not vote to being sure that one did vote, with two increments in between, where respondents could say that they were not sure, but presumably did or presumably did not vote. This approach was developed to capture valid

⁴ In the beginning of their research, they started with an even longer version of the introductory sentence (Belli, et al. 1999), but they reduced it to a shorter version in the course of their research (Belli, Moore, & VanHoewyk 2006).

turnout responses first, in surveys which are conducted *not* immediately after an election, and second, in pre-election surveys asking about turnout in the *last* election. Again, it successfully reduced reported turnout compared to the standard question and approached the official turnout rate.

So far, we have thus some evidence that changes in question wordings and a broader diversification in response options help to reduce overreporting in survey questions.⁵ However, what we do not yet know is whether these results hold across different contexts. Thus, in this paper, we explore whether they can be applied to different settings, and if so, how.

3. Context and individual specific causes of overreporting

Analyzing explanations of overreporting, one can observe that research has identified a number of individual level characteristics that are positively related to a respondent's tendency to overreport turnout.⁶ Education is known to increase overreporting, and so are political engagement, political interest, frequency of following the news, and civic duty (Bernstein, et al., 2001; Cassel, 2003; Granberg & Holmberg, 1991; Hill & Hurley, 1984; Karp & Brockington, 2005; Presser & Traugott, 1992; Silver, et al., 1986). Though, results on age are puzzling. Some scholars observe that young non-voters are more likely to acknowledge abstention than older non-voters. This means that overreporting increases with age (e.g. Granberg & Holmberg, 1991). However, other scholars report the contradicting patten, namely that young voters are more likely

⁵ Other recent attempts avoid the direct turnout question but take an indirect approach: the Item Count Techniques (Holbrook & Krosnick, 2011b; Zeglovits & Kritzinger, 2011) and the Randomized Response Techniques (Holbrook & Krosnick, 2011a). These indirect approaches have been tested with mixed success. Therefore, we will not test these techniques in a cross-country design.

⁶ In general, the characteristics that usually affect turnout also affect overreporting.

to overreport than older voters (e.g. Hill & Hurley, 1984). In contrast, Zeglovits and Kritzing (2011) did not find any age effect at all with question wordings reducing overreporting. Finally, Silver et al. (1986) describe age as having a curvilinear effect on overreporting, similar to the curvilinear effect of age on turnout. As a general conclusion, though, correlations between these characteristics and turnout might be inflated.

Unlike research at the individual level, research at the contextual level is rare if existent at all. With the exception of Karp and Brockington's study (2005) no other study describes and analyzes cross-country differences in overreporting in surveys. Moreover, to our knowledge no study deals with cross-country differences in question wording and diversifications in response options, and their effects on overreporting. Due to this gap in research, we also lack knowledge on theoretical frameworks: more precisely, we cannot deduce any particular assumption on how contextual circumstances might influence alternative question formats. For instance, alternative question formats might be *more* or might be *less* successful in reducing overreporting in circumstances where overreporting is high. Both impacts could be possible: on the one hand, if the probability of overreporting is high due to certain contextual circumstances,⁷ successfully tested attempts to reduce overreporting will work better than the standard question; on the other hand, it might be that also new question formats do not have any impact at all as social desirability behavior becomes even enhanced by particular contextual factors. Thus, based on different contextual factors cognitive dissonances might be dissolved differently.

This gap in the literature leads us to take an exploratory approach: we will simply explore whether new question formats produce a) positive differences at all, and b) whether these differences are equal across countries. From there we will draw descriptive inferences on

⁷ Saliency of the election or high overall turnout (see Karp & Brockington, 2005).

contextual factors. As contextual factors we consider overall turnout, election saliency (e.g. closeness of the race) and the time passed since the last election. Based on these contextual circumstances different question formats might perform differently. For instance, in contexts where elections took place a long time ago, respondents might be able to dissolve their cognitive dissonance better if the alternative question format reflects *likelihoods*. In context where elections took place only recently, this might not be the case.

4. Design of the experiment

a. Question formats: three different treatments

To analyze our research question we set up a survey experiment, comparing three different question formats with varying response options. We will refer to them as *treatments* in our experiment.

Treatment A used the standard question format used in most election studies, and thus formed our reference group. Here, the question wording was already set up in such a way as to make it easier for respondents to report non-voting, as it is usually done:

Treatment A: “The following question refers to the previous [federal election] in [month year]⁸. In this [federal election] a lot of people could not vote or chose not to vote for some reasons. What about you? Did you vote or not?”

In *Treatment B*, we used the approach developed by Belli et al. (1999, 2006). Our Treatment B is a translated and adapted version of this approach with a response scale including *four* possible

⁸ As the French legislative elections took place in two rounds, we referred to the second round on June 17, 2007.

answers. Although minor changes were necessary in the question wording,⁹ we were especially careful to keep the response options identical. Thus, in Treatment B we asked:

Treatment B: “The following question refers to the previous [federal election] in [month year]. In talking to people about elections, we often find that a lot of people were not able to vote because they were sick, did not have the time, or just were not interested. Which of the following statements best describes you?”

- I did not vote in the [federal election] in [month year].
- I thought about voting this time but didn’t.
- I usually vote but didn’t this time
- I am sure I voted in the [federal election] in [month year].”

Treatment C is the question format developed by Zeglovits and Kritzing (2011). We combined the stimulus of Treatment A with the four response options pointing to the likelihood of past behavior:

Treatment C: “The following question refers to the previous [federal election] in [month year]. In this election, a lot of people could not vote or chose not to vote for some reasons. Also, some time has passed since. Which of the following statements best describes you?”

- I am sure I did not vote in the [federal election] in [month year].
- I am not sure if I voted but I presumably did not.
- I am not sure if I voted but I presumably did.

⁹ E.g. with regards to voters’ registration mentioned in the original stimulus which does not apply to our setting in Europe.

- I am sure that I voted in the [federal election] in [month year].”

With this approach, we offered face-saving response options that fit the amount of time passed for people who can remember that they did not vote. Additionally, we made it easier to admit that one simply cannot remember.

b. Survey design

We conducted an online survey to compare our treatments. So far, studies on the topic of vote overreporting have only been based upon face-to-face or telephone surveys (e.g. Hill & Hurley, 1984, Silver & Anderson, 1986, Abelson, et al., 1992, Belli, et al., 2001, Cassel, 2003; Belli, et al., 2006).¹⁰ With our study we can thus contribute to the question, if turnout data gathered through web surveys reports overreporting as well, and if so, whether new question formats perform better than the standard one. Second, social desirability is usually assumed to be increased in the presence of an interviewer but also in the presence of other people when answering the questions, the data collection settings, or third party disclosure (Tourangeau, Rips, & Rasinski, 2009; Tourangeau & Yan, 2007). As there is no interviewer present in an online survey, social desirability should in general be reduced in this self-administered survey (Chang & Krosnick, 2009; De Leeuw & Collins, 1997; Tourangeau & Smith, 1996). Any observed differences amongst the various question formats might thus reflect the dissolution of the individual’s personal cognitive dissonance on assumed social behavior.

¹⁰ Holbrook and Krosnick’s (2010) tested the Item Count Technique in an online survey, but the experiment failed to reduce self-reported turnout. When using a telephone survey they were however successful.

The survey experiment was conducted as a separate part in a survey on Internet service providers, which was coordinated by TNS opinion¹¹ and hosted by TOLUNA.¹² On average 1.000 respondents per country completed the questionnaire using the technique of Computer Assisted Web Interviewing (CAWI). The data collection took place December 1-21, 2011. For each country, TOLUNA builds samples on multiple recruitment strategies, via telephone and online recruitment. Panelists were invited to participate based on soft quotas on gender, age and urbanization. The samples show an acceptable distribution in terms of gender, age and urbanization.

Concerning the multi topic survey, possible respondents have been screened with respect to their interacting with Internet service providers in the last three years. To ensure eligibility of respondents for the experiment, a couple of filter questions were used to provide that respondents are citizens of a country. The participants were randomized into one of the three treatment groups. Respondents had to answer each question to be able to proceed to the next question. However, respondent were offered an escape option for each question: for all three treatment groups the “don’t know”-option was displayed when respondents tried to skip the question. The overall length of interview did not exceed 12 minutes.

Since the respondents of the TOLUNA samples are recruited and not chosen using a probability sample, it is important to note that the respondents are not in any way representative for the

¹¹ TNS opinion is a Brussels based research firm that is specialized in conducting multi-country studies in all parts of the world, working with and for clients such as European Commission, World Bank, European Central Bank, European Bank for Regional for Reconstruction and Development. Furthermore, TNS opinion participates in the frame work programmes and Marie Curie Actions of the EC.

¹² TOLUNA is the world’s leading independent online panel and survey technology provider. For this survey the Amsterdam office of TOLUNA was involved. TOLUNA provides online sample for most European countries and uses a panel care approach that focuses on maximising panellist engagement to offer increased survey responsiveness and data reliability.

population in the countries. The survey respondents deviate from the population in the different countries in a number of ways. First of all, the respondents are all Internet users and probably regular or heavy internet users. Because we are mostly interested in the underlying relationships between variables, we consider the deviation in the sample with respect to the adult population and the lack of representativity to be less problematic. As we are dealing with a non-probability panel, we only focus on the effect of the experimental treatments for reported levels of turnout. It is important to note that our objective is not to estimate accurate population values since the data is not suited for such purpose. Our more limited purpose is to evaluate and test the effects of applying different question constructions.

We were able to conduct our experiment in those seven European democracies that were part of the online survey conducted by TNS: Austria, Belgium, Denmark, France, Germany, the UK and Ireland. Due to restrictions based on the commercial nature of the survey, we were not able to fully control the selection of countries. However, the country selection proved to fit our contextual needs as substantial variation in some of the aforementioned contextual factors can be found in these countries: The sample included countries with variation on overall turnout rates. For instance, Belgium has very high levels of turnout, whereas in the UK turnout is rather low. Furthermore, we have also variation in the time passed since the last election (Saris & Gallhofer, 2007): Denmark is the country with the most recent election. Meanwhile, France represents a country where the election took place a long time ago.

Table 1 gives an overview of the countries based on the time passed since the election and official turnout in each election.

Table 1: Countries and elections selected

	Election	Election day(s)	Official turnout
AUT	Nationalratswahl	28.09.2008	78.8
BE (fr.)	Élections législatives fédérales belges de 2010	13.06.2010	87.7 ¹³
DK	Folketingsvalg	15.09.2011	87.2
FR	Élections législatives	22.04.2007 06.05 2007	83,77 (1st round) 83,97 (2nd round)
GER	Bundestagswahl	27.09.2009	70.8
UK	British General Election	06.05.2010	65.1
IRE	Irish General elections	25.02.2011	70.1

5. Results

First, we are interested in analyzing whether the different question wordings do actually reduce reported turnout in the studied country. As we cannot validate turnout, we have to compare reported levels of turnout in each treatment group. If reported levels of turnout decrease, we assume that overreporting is decreased. Of course, it might happen that we overshoot the mark and press respondents to underreport turnout. But, as underreporting is known to be the less significant problem, we keep with interpreting lower reported rates as the more accurate ones. We will first compare the answers in treatments B and C to the answers in the reference group (treatment A) to check if the new question formats lead to different levels of reported turnout as the standard question. This approach has been used repeatedly to evaluate new question formats (Belli, et al., 2006; Holbrook & Krosnick, 2010). Second, we will also compare treatments B and

¹³ Regions considered: Hainaut, Namur, Walloon Brabant, Liège, Luxembourg (Lussimbork).

C to see which one behaves how in which contextual circumstances. For this aim, we present simple descriptive statistics. In Table 2 we can observe that the additional response options in treatments B and C are chosen by quite a number of respondents. In treatment B, we observe that in addition to simply admitting that they did not vote, a considerable share of respondents chose the response options that they thought about voting, but then did not (ranging between 2% in Belgium to 8% in the UK and Germany) or that they usually vote, but this time did not (ranging from 3% in the UK, Germany and Denmark to 9% in France). The descriptive results suggest that the simple “I did not vote” option is chosen by about the same share of respondents in treatments B and A in all countries – with the exception of Germany. In general, the additional response options for admitting non-voting in treatment B lead to a much higher total share of declared non-voters.

The same holds true for treatment C: In sum, between 4% (Denmark and Ireland) and 16% (Austria) of respondents choose the response options that they are not sure if they voted or not. Denmark and Ireland are the two countries with the most recent elections and are also the two countries with the lowest shares of respondents that admit they cannot remember if they voted or not. This adds to the interpretation raised by Zeglovits and Kritzinger (2011) that treatment C might be particularly useful for elections that took place a long time ago. In sum, additional response options seem to be successful in the sense that they are chosen by a considerable share of respondents, but this share differs across countries.

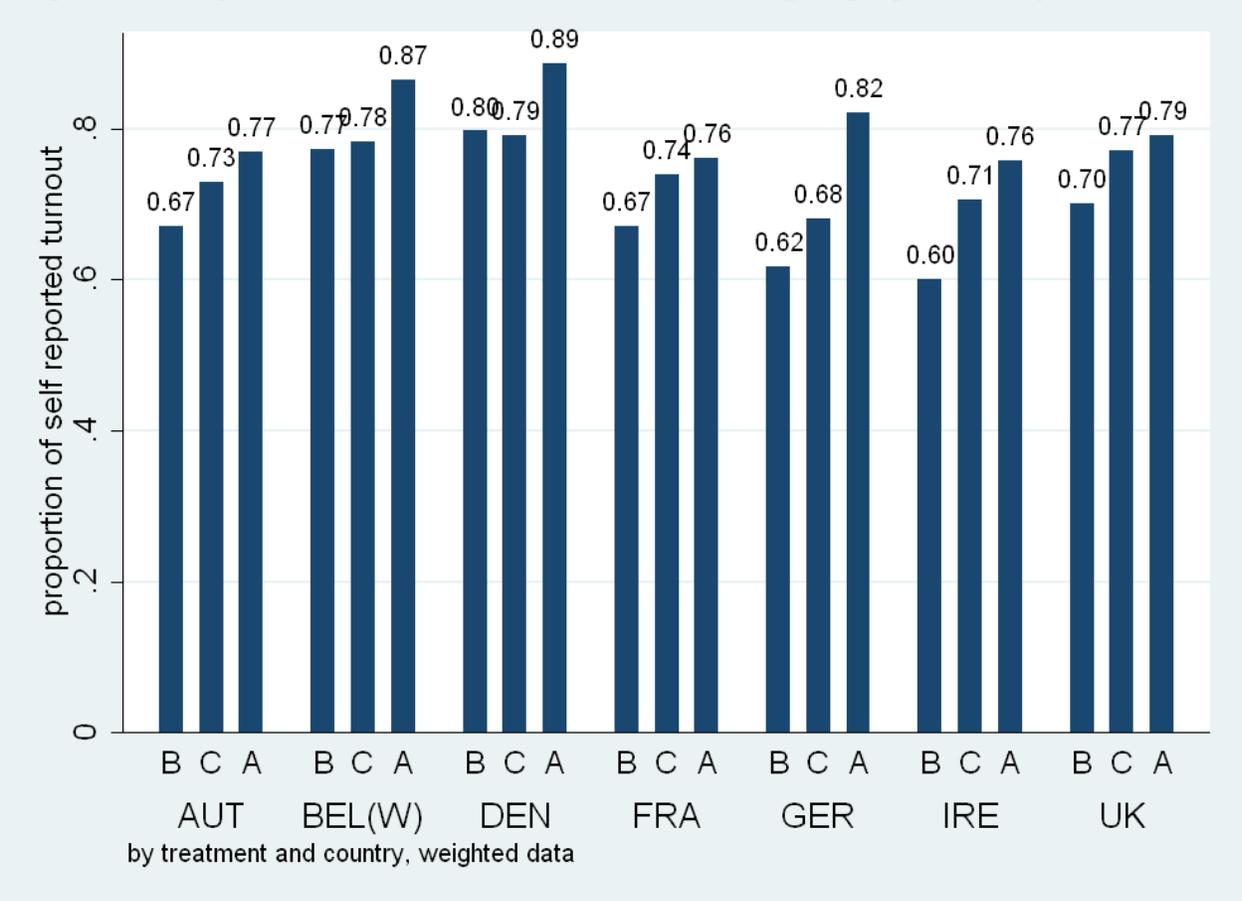
Table 2: Self-reported turnout for three question formats, by country

Treatment B		Treatment C		Treatment A	
Austria					
did not vote	24%	sure I did not vote	20%	did not vote	23%
thought about voting, but did not usually vote, but this time not	5%	presumably not	7%		
	4%	presumably yes	9%		
sure I voted	67%	sure I voted	64%	voted	77%
Belgium					
did not vote	15%	sure I did not vote	18%	did not vote	13%
thought about voting, but did not usually vote, but this time not	2%	presumably not	4%		
	6%	presumably yes	4%		
sure I voted	77%	sure I voted	74%	voted	87%
Denmark					
did not vote	12%	sure I did not vote	18%	did not vote	11%
thought about voting, but did not usually vote, but this time not	5%	presumably not	3%		
	3%	presumably yes	1%		
sure I voted	80%	sure I voted	78%	voted	89%
France					
did not vote	21%	sure I did not vote	20%	did not vote	24%
thought about voting, but did not usually vote, but this time not	3%	presumably not	6%		
	9%	presumably yes	7%		
sure I voted	67%	sure I voted	67%	voted	76%
Germany					
did not vote	27%	sure I did not vote	27%	did not vote	18%
thought about voting, but did not usually vote, but this time not	8%	presumably not	5%		
	3%	presumably yes	3%		
sure I voted	62%	sure I voted	65%	voted	82%
Ireland					
did not vote	25%	sure I did not vote	26%	did not vote	24%
thought about voting, but did not usually vote, but this time not	7%	presumably not	3%		
	8%	presumably yes	1%		
sure I voted	60%	sure I voted	69%	voted	76%
UK					
did not vote	19%	sure I did not vote	20%	did not vote	21%
thought about voting, but did not usually vote, but this time not	8%	presumably not	3%		
	3%	presumably yes	2%		
sure I voted	70%	sure I voted	75%	voted	79%

Note: weighted data

Next, we recoded the questions to a simple indicator variable, indicating self-reported turnout (1), compared to self reported abstention (0). For treatment C, we had to make an assumption how to classify people who reported that they were not sure, if they had voted or not. We counted people, who reported that they were not sure but presumably did vote as voters, and the ones who were not sure but presumably did not vote as nonvoters. This estimation for C simplifies the results. Graph 1 displays self-reported turnout for each country comparing the three treatments. Moreover, the estimation for C allows running simple z-tests (Table 3) to analyze whether the proportion of self-reported voters was lower in treatments B and C than in A in each country.

Graph 1: Self-reported turnout in the different treatment groups, per country



Both Graph 1 and Table 3 show that treatment B reduces self-reported turnout significantly compared to the standard treatment A in all countries. The estimated turnout of treatment C is significantly lower than reported turnout in treatment A in four countries: French speaking Belgium, Denmark, Germany and Ireland.

Table 3: Z-Tests, comparing share of reported turnout in treatments B and C to treatment A; plus comparing B to C

		proportion of reported turnout	Standard error of proportion	Z-statistic: B and C compared to A	p-value (one- sided)	Z-statistic: B compared to C	p-value (two sided)
AUT	B	0.672	0.026	-2.748	0.001	-1.578	0.057
	C	0.729	0.025	-1.179	0.060		
	A	0.769	0.024				
BEL	B	0.774	0.032	-2.225	0.007	-0.210	0.417
	C	0.783	0.030	-2.050	0.010		
	A	0.865	0.026				
DEN	B	0.797	0.023	-3.046	0.001	0.184	0.573
	C	0.791	0.023	-3.267	0.000		
	A	0.886	0.018				
FRA	B	0.671	0.024	-2.718	0.002	-2.100	0.018
	C	0.740	0.023	-0.651	0.129		
	A	0.761	0.023				
GER	B	0.618	0.026	-6.299	0.000	-1.797	0.036
	C	0.682	0.024	-4.452	0.000		
	A	0.821	0.020				
IRE	B	0.601	0.030	-3.950	0.000	-2.599	0.005
	C	0.706	0.027	-1.356	0.044		
	A	0.757	0.026				
UK	B	0.700	0.025	-2.751	0.001	-2.118	0.017
	C	0.771	0.022	-0.665	0.127		
	A	0.791	0.022				

Finally, we ran several logistic regression models.¹⁴ We ran three models to test whether treatments work at all, and if so whether they work the same in all countries. The dependent variable is self-reported turnout, 1 indicating that a person reported that he or she had voted, 0 indicating self-reported non-voting (as in Graph 1). As independent variables we use variables indicating the treatments B and C, with the standard question in treatment A forming the reference group (*Model 1 – treatment only*). Next, we add interaction terms between countries and treatments meaning that we allow treatments to work differently in the countries (*Model 2 – treatment in countries*). Here, Ireland is the country of reference. To check whether the effects observed hold true we control for individual characteristics that are known to affect turnout and overreporting (*Model 3 – with controls on the individual level*).

Table 4: Logit models; Ireland as the reference group

Variable	treatment only	treatment in countries	with controls on the individual level
treatment B	-0.643***	-1.044***	-0.995***
treatment C	-0.416***	-0.569***	-0.498**
interaction B * AUT		0.344***	0.288***
interaction B * BEL		0.802***	0.733***
interaction B * DEN		1.029***	1.198***
interaction B * FRA		0.290***	0.035
interaction B * GER		0.142***	-0.02
interaction B * UK		0.512***	0.105*
interaction C * AUT		0.092***	0.122**
interaction C * BEL		0.349***	0.204***
interaction C * DEN		0.460***	0.628***
interaction C * FRA		0.119***	-0.143***
interaction C * GER		-0.094***	-0.317***
interaction C * UK		0.287***	-0.041

¹⁴ We could not go for multilevel modeling as we only have 7 countries in our data. Random effects cannot successfully be modeled with that small n on the second level (Rabe-Hesketh & Skrondal, 2008, p.124). So we went for simple logistic models taking the clustered structure of the data into account when calculation standards errors, with the option `vce(clustered)` in the STATA command.

freq following news			0.175***
efficacy			0.246***
age			0.050*
age squared			0
female			0.091
education medium			0.302**
education high			0.525***
migrant			-0.841***
income			-0.054
currently working			0.269*
area big city			-0.117
area small city			-0.132*
<hr/> _cons	1.421***	1.421***	-1.811***
<hr/> N	6779	6779	6775
<hr/> ll	-3.833.716	-3.802.520	-3.487.807

Note: * p<0.05; ** p<0.01; *** p<0.001

Turning to the results of the logistic regression models (Table 4), in all three models we can observe that both alternative question wordings are successful in reducing respondents' tendency to report positive turnout. There are country specific differences in *how* treatments B and C reduce reported turnout. In Ireland, treatment B works “best” in the sense that reported turnout is reduced most severely. Treatment C has the greatest effect in Germany with reported turnout being reduced even more than in Ireland. Running a linear test to check if the treatment reduces self-reported turnout within each country,¹⁵ we can observe that both treatments significantly reduce reported turnout in five of the seven countries, namely Austria, France, Germany, Ireland and the UK, where treatment effects plus country specific correction terms are significantly lower than zero (both in models 2 and 3). In French speaking Belgium the reduction is not significant.¹⁶ In Denmark there is no significant effect of the treatments left when controlling for

¹⁵ Thus, we tested if the sum of the coefficient of a treatment and the interaction term between the treatment and the variable indicating a specific country are significantly different from zero.

¹⁶ This might be due the small n for Belgium. The p-value for the sum of the effects for treatment C is smaller than 0.10, but from our perspective still not significant.

the individual characteristics. Thus, self-reported turnout could be reduced in five out of seven countries when introducing new question formats.

6. Discussion and Conclusion

We have shown that reported turnout can be reduced substantially even in an online survey where no interviewers are present. Our experiment is thus the first online study where an attempt to reduce overreporting was tested successfully.

Moreover, we have shown that question wordings work indeed differently in different contexts. Depending on contextual circumstances, question formats perform differently. The descriptive results of our experiment suggest that the new approach introduced by Zeglovits and Kritzinger (2011) is convenient for surveys where elections took place a long time ago, while the Belli et al. (2006) approach is more favorable for more recent elections. However, future research taking into account a larger number of countries will need to test also analytically how country specific context variables change the effects of the treatments.

Of course, there are important limitations to our findings. As we used a non-probability sample consisting of an online access panel we cannot draw any conclusions on whether our findings can be generalized to the entire population of the surveyed countries. To analyze whether our results also hold true for the country population data collection needs to be based on random sampling techniques.

Finally, so far we have not tackled the problem of language differences in comparative research. For future research, we need to raise the very general question whether questions translated

carefully from one language to another always capture the same phenomenon in the different countries. Our small scaled experiment shows considerable variations between the countries regarding the effects of the different question wordings. It might be possible that one question form is simply more suitable from a language perspective in one country, whereas another question format captures turnout linguistically in a more accurate form in another country.

Acknowledgment

We would like to thank *TOLUNA* for their help in conducting the experiment and the entire technical setup, as well as *TNS opinion* for providing translations and implementing the survey.

References

- Abelson, R. P., Loftus, E. F., & Greenwald, A. G. (1992). Attempt to Improve the Accuracy of Self-Reports in Voting. In J. M. Tanur (Ed.), *Questions about Questions. Inquiries into the Cognitive Bases of Surveys*. New York: Russell Sage Foundation.
- Belli, R. F., Moore, S. E., & VanHoewyk, J. (2006). An experimental comparison of question forms used to reduce vote overreporting. *Electoral Studies*, 25(4), 751-759.
- Belli, R. F., Traugott, M. W., Young, M., & McGonagle, K. A. (1999). Reducing Overreporting in Surveys: Social Desirability, Memory Failure, and Source Monitoring. *Public Opinion Quarterly*, 63(1), 90-108.
- Bernstein, R., Chadha, A., & Montjoy, R. (2001). Overreporting Voting: Why It Happens and Why It Matters. *Public Opinion Quarterly*, 65(1), 22-44.
- Cassel, C. A. (2003). Overreporting and Electoral Participation Research. *American Journal of Politics*, 31(1), 81-92.
- Chang, & Krosnick, J. A. (2009). National Surveys Via Rdd Telephone Interviewing Versus the Internet Comparing Sample Representativeness and Response Quality *Public Opinion Quarterly*.
- De Leeuw, E., & Collins, M. (1997). Data Collection Methods and Survey Quality: an Overview. In L. Lyberg, P. P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz & D. Trewin (Eds.), *Survey Measurement and Process Quality* (pp. 199-220). New York: Wiley.
- Granberg, D., & Holmberg, S. (1991). Self-Reported Turnout and Voter Validation. *American Journal of Political Science*, 35(2), 448-459.
- Hill, K. Q., & Hurley, P. (1984). Nonvoters in Voters' Clothing: The Impact of Voting Behaviour Misreporting on Voting Behaviour Research. *Social Science Quarterly*, 65(1), 199-206.
- Holbrook, A. L., & Krosnick, J. A. (2010). Social Desirability Bias in Voter Turnout Reports. *Public Opinion Quarterly*, 74(1), 37-67.
- Karp, J. A., & Brockington, D. (2005). Social Desirability and Response Validity: A Comparative Analysis of Overreporting Voter Turnout in Five Countries. *Journal of Politics*, 67(3), 825-840.
- Presser, S., & Traugott, M. W. (1992). Littel White Lies and Social-Science Models - Correlated Response Errors in a Panel Study of Voting. *Public Opinion Quarterly*, 56(1), 77-86.
- Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and longitudinal Modeling Using Stata* (2nd ed.). College Station (Texas): Stata Press.
- Saris, W. E., & Gallhofer, I. N. (2007). *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. Hoboken (NJ): Wiley.
- Silver, B. D., Anderson, B. A., & Abramson, P. R. (1986). Who overreports voting? *American Political Science Review*, 80(2), 613-624.
- Stockè, V. (2007). Response Privacy and Elapsed Time Since Election Day as Determinants for Vote Overreporting. *International Journal of Public Opinion Research*, 19(2), 237-246
- Stockè, V., & Stark, T. (2007). Political involvement and memory failure as interdependent determinants of vote overreporting. *Applied Cognitive Psychology*, 21(2), 239-257.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2009). *The psychology of survey response* (10th ed.). Cambridge: Cambridge University Press.
- Tourangeau, R., & Smith, T. W. (1996). Asking Sensitive Questions: The impact of Data Collection Mode, Question Format, and Question Context. *Public Opinion Quarterly*, 60(2), 275-304.
- Tourangeau, R., & Yan, T. (2007). Sensitive Questions in Surveys. *Psychological Bulletin*, 133(5), 859-883.
- Traugott, M. W., & Katosh, J. P. (1979). Response Validity in Surveys of Voting Behavior. *The Public Opinion Quarterly*, 43(3), 359-377.
- Zeglovits, E., & Kritzinger, S. (2011). *Reducing Overreporting in the Voter Turnout Question*. Paper presented at the Fourth Conference of the European Survey Research Association (ESRA).