

Assessing the Validity of Survey Measures for News Exposure Through Digital Footprints: Evidence from Spain and the UK

Ana S. Cardenal (corresponding author)

School of Law and Political Science, Universitat Oberta de Catalunya

acardenal@uoc.edu

María Victoria-Mas

School of Communication, Universitat Internacional de Catalunya

mvictoria@uic.es

Silvia Majó-Vázquez

Reuters Institute for the Study of Journalism, University of Oxford

silvia.majo-vazquez@politics.ox.ac.uk

Iván Lacasa-Mas

School of Communication, Universitat Internacional de Catalunya

lacasa@uic.es

Forthcoming in Political Communication

This paper assesses the performance of three commonly used type of questions –open-ended, check-all and forced choice— for capturing retrospective online news exposure, combining both survey and web-tracking data. It examines the performance of these different survey questions considering both systematic and random error in two unexplored non-US contexts: Spain and the UK. Results show that the check-all question produces on average the most accurate –i.e. less biased— estimates of observed exposure. Some motivational and cognitive factors underlying bias in self-reports are explored. Findings reveal that the characteristics of outlets are associated with systematic error. Finally, we find that media systems matter for accuracy –where media fragmentation is high (Spain), accuracy is low across all questions; where it is low (UK), accuracy is high across all questions. In a final section, we highlight the methodological and theoretical contributions of our study and provide some recommendations.

Keywords: media exposure, survey measures, digital footprints, overreporting, cognitive bias.

Introduction

Although we know that survey-based measures of media exposure are imperfect (Guess, 2015; Konitzer et al., 2020; Parry et al., 2021, Prior, 2009b; Vraga et al., 2016; Vraga & Tully, 2020), there is widespread scholarly agreement that doing away with these measures is simply not an option (Dilliplane et al., 2013; Guess, 2015). Surveys (i.e., asking people to retrospectively report their media exposure) continue to be valuable measurement instruments of media exposure for many reasons. They are less costly to implement than relying on more technologically advanced techniques to obtain tracking data (Stier et al., 2020). They can be more easily implemented across a large number of countries, contributing to cross-national comparisons of media use and media effects, and they can assess the role of media in real-world studies of political behavior for large

populations, overcoming problems of external validity (Guess, 2015, Prior, 2009a).

Accordingly, several endeavors to improve on survey-based measures of exposure have been taking place since 2007. Some endeavors have been carried out autonomously by scholars, who have taken advantage of technological developments and innovative research designs to test the validity of self-reported measures (Guess, 2015; Parry et al., 2021; Vraga et al., 2016). These efforts have been prompted by two interrelated, though distinct, recent developments: technological advancements, which now allow to trace individual's news consumption behavior unobtrusively (Stier et al., 2019), and a changing and more complex media environment with a growing number of news outlets and side-doors to news, which calls for an adaptation of self-reported measures of exposure (Althaus & Tewksbury, 2007; de Vreese & Neijens, 2016).

Here we join this collective effort to improve upon existing survey-based measures with a focus on measures of online news consumption. Despite the fact that Internet already surpasses TV for access to news in many countries, including the UK and Spain, (Newman et al., 2021), most efforts have focused on improving survey-based measures of TV (Andersen et al., 2016; Prior, 2009b) with only a few works devoted to online news sources (Guess, 2015; Vraga & Tully, 2020).

In this study, we assess the performance of three of the most frequently used survey measures of online media exposure. Our results show that the check-all question provides on average the most accurate –i.e., less biased—estimates of observed exposure. We study some of the motivational and cognitive factors underlying bias in self-reports, showing that the characteristics of outlets are associated with systematic error. Finally, we find that media systems matter for accuracy –where media fragmentation is high (Spain), accuracy is low across all type of questions to measure online media exposure; where it is low (UK), accuracy is high across all questions.

We rely on passively collected digital trace data to construct direct, unobtrusive measures of exposure, which we then use as a benchmark to test the validity of these survey measures. Our study stands on the shoulders of previous research (see especially Guess 2015), but it innovates on several fronts. First, in contrast to previous research, which has largely focus on the US (Guess, 2015; Prior, 2009b; Vraga et al., 2016; Vraga & Tully, 2020), our study tests the performance of survey measures of exposure in two European countries: Spain and the UK. Hence, we expand existing research on the topic by bringing in evidence from two non-US countries that represent two very different media systems.

Secondly, we focus on a period of abnormally high news consumption: the weeks following the outbreak of the COVID-19 pandemic in March 2020. Given that during this period online consumption of news also intensified, leading people to visit a higher than usual number of online news sources and spending more time on them (Van Aelst et al., 2021, Newman et al. 2020), our study can be seen as a hard test for the performance of survey-based measures of online news consumption. Although by the time we collected the data (May 2020), the maximum peak in news consumption had passed in both countries, people were still in lockdown (Spain) or moving out of it (UK), which likely affected their usual consumption behavior increasing the average time spent in news exposure, and still far from the levels seen before the pandemic.¹

¹ See the evolution of time spent in the five most visited outlets in the months surrounding the outbreak of the pandemic for each country in Figure A1 in the Online Supplementary Material (p. 2), which can be found here:

https://www.dropbox.com/s/zdpn4p4f1cg6pxx/Online%20Supplementary%20Material_R1%20.docx?dl=0

Third, following a recent strand of research, our study evaluates accuracy of survey measures considering both systematic and random error. Moreover, it assesses the performance of these questions by exploring the motivational and cognitive factors associated with bias in self-reports across conditions.

In the remainder of this paper, we proceed as follows. We first review the relevant literature highlighting the impact of question formats, individuals' characteristics and media systems on self-reported error. We then derive our hypotheses and proceed to describe our data and methods and present the results of our analyses. Finally, we conclude by discussing the methodological and theoretical contribution of our study and providing some recommendations.

Survey-based questions to tap online news consumption

A lot of attention has been placed on the challenges of correctly measuring media exposure, especially since Althaus and Tewksbury published an influential report in 2007. However, fewer efforts have been made on specifically testing measures of news exposure in the online environment, where the complexity of news consumption is at the highest (an important exception here is Guess 2015). Recently, a growing strand of research has focused on misreporting specifically on social media platforms (Ernala, Burke, Leavitt, & Ellison, 2020; Guess, Munger, Nagler, & Tucker, 2019; Haenschen, 2020) and more generally on the Internet (Scharrow, 2016). Some of the insights from this line of research are useful for the purposes of this study in that they distinguish between systematic and random error in self-reports, and will guide our expectations and analyses about error affecting self-reporting on media diets.

Several exclusive features of the online news ecosystem complicate measurement of news and political information exposure in this environment --the sheer quantity of available sources and its distributed access, the speed and dynamism of social and connective media (Niederdeppe, 2016), the participation of citizens in what is considered “newsworthy” (Niederdeppe 2016), incidental exposure which is greater among online news users (Fletcher et. al 2017) and the blurring of content and genre across media platforms (Landert, 2014).

In order to advance which measures are better to capture online news consumption in this environment, we focus on three types of questions: The check-all, open-ended and forced-choice questions, long debated in previous research (Guess 2015). The check-all question provides respondents with a list of most visited sources after asking them to report their media exposure. It has a precedent in the program-list technique proposed by Dilliplane et al. (2013). This measure was first implemented in the National Annenberg Election Study in 2008 and later in the ANES in 2012 (Andersen et al. 2016; Guess, 2015). An adapted version of this technique has used a list of most visited sources to tap on media exposure during a campaign (Van Spanje & De Vreese, 2014) or on online news consumption (Cardenal et al., 2019a; 2019b; Guess, 2015).

The open-ended question searches for spontaneous answers when asking respondents to report the media or sources used to get political information. It has been used by the European Electoral Study (EES) and by at least one opinion poll in Spain. For example, in 2004, the EES asked respondents to name up to five channels or television news programs that they watched regularly. The Centre d’Estudis d’Opinió (CEO), the main Catalan opinion poll, asks regularly to name only one TV channel, radio station or newspaper that respondents frequently use to get political information.

Finally, the forced-choice question requires people to choose either “Yes” or “No” to each question in a sequence, which in this case is a sequence of brands corresponding to the list of news outlets. This type of question is mostly used in telephone surveys (Smyth et al., 2006).

The list technique has been advocated because it is expected to decrease the cognitive demands placed on respondents (Andersen et al., 2016; Dilliplane et al., 2013: 239; Nagler, 2017: 8) vis a vis the open-ended question. According to Dilliplane et al. (2013) it should be easier for respondents to recognize TV programs (or for our matter news sources) than “to mentally tally of time devoted” to TV exposure (or news). The check-all question decreases cognitive demands through recognition (Nagler, 2017: 8), which in turn may facilitate recall (Schwarz & Oyserman, 2001). In this respect, check-all questions may help minimize the kind of memory overload problems that drive misreporting (Prior 2009a), and which the greater complexity of a high-choice media environment have only exacerbated (De Vreese & Neijens 2016; Niederdeppe 2016; Andersen et al. 2016). However, by preselecting a list of choices, the check-all question may induce errors in judgment caused by the familiarity of a given item (Guess, 2015). This leads to the expectation that the check-all question will foster respondents’ tendency to overreport media exposure, at least compared to the open-ended format: *Overreporting will be higher in the check-all format compared to the open-ended question across both countries (H1a).*

The forced-choice format compels respondents to think and make decisions about each of the options, which may encourage deeper processing helping to produce more accurate results. However, it may foster even more overreporting than the check-all question by weakening a satisficing strategy (Krosnick, 1991; see also Guess 2015). In a study comparing these two types of questions, Smyth et al. (2006) found that

respondents took longer to answer the forced-choice format and endorsed more options than in the check-all question. This leads to the following hypothesis: *Overreporting will be higher in the forced-choice condition compared to the check-all question across both countries (H1b).*

Bias in self-reported exposure

The problem of error when using survey measures of exposure is well-documented (Prior 2009a, Konitzer et al. 2020). However, error *per se* may not be a problem if it is random. What is indeed problematic is that error may be systematic and likely to be related to underlying factors (confounders) that in turn influence the key opinion or behavior under study (see Scharkow, 2016 for an excellent illustration on this). In such case, effects may not be attributable to media exposure but to factors influencing accuracy (Prior 2009b).

There are two types of confounders that may bias effects in media studies (Nagler, 2017). The first one is motivational. Individuals interested in politics may be more motivated to report media use. Research suggests that having high interest in politics is often associated with overreporting exposure to media that may be considered pro-social, like news, while underreporting is more common for less desirable media, like video games or certain websites (Kahn et al., 2014). Kahn et al. (2014) suggest that how people view themselves (e.g., as regular news consumer) will influence self-reports about their behavior, so that their views and actions align as a means of resolving any potential cognitive dissonance. Another reason why politically interested individual might be more likely to overreport media use is related to attention. They are more likely to pay attention to news and hence to overreport reception (Price & Zaller, 1993; Zaller, 1992).

Beyond the news media ecosystem, research has found that more politically motivated individuals are also more prone to misreport their activity on social media (Guess et al., 2019). This work shows that those who are more partisan are significantly more likely to overreport their posting on social media. Yet, the role of political interest is contentious since Haenschen (2020) shows that heavy Facebook users may simply have better awareness of their activity.

Overall, looking at the research within the realm of news consumption we expect that: *Politically interested people will overreport more than people not interested in politics across conditions and contexts (H2).*

A second type of confounder that may bias studies on media effects is cognitive. Research shows that question wordings and answers can induce errors of judgement (e.g., Schwarz & Oyserman 2001, Sudman et al. 1996). Closed questions formats (i.e., check-all) might help respondents to recall the relevant behavior (e.g., Schwarz & Oyserman 2001, Ernala, 2020). Yet, they might also increase overreporting by facilitating accessibility or familiarity through easing the perception of an item (Tourangeau et al. 2000). In contrast, by removing the familiarity effect, open-ended questions might help counteract respondents' tendency to overreport (Guess 2015). This leads to expect more overreporting in check-all questions as formulated in H1.

However, given the complexity and high choice of the online media environment, not helping respondents to recall, can lead to substantial underreporting. Moreover, it can exacerbate bias in the "recalling" stage. Open-ended questions have been found to be good at reporting "salient" concerns (Geer, 1991) and "frequent" behavior (Sudman et al., 1996). In open-ended questions, where respondents are offered no aid accessibility bias may set in, and respondents may be more likely to recall information that is more easily retrieved from memory (i.e., salient, frequent, or

recurring) (Iyengar, 1990). When asked to report news sources people have visited, the “accessibility bias” should make it more likely to recall those sources that are easily retrieved from memory, usually those with stronger brand awareness, higher reach and partisan slant, even if respondents have not visited them. In their study, Vraga and Tully (2020) reach similar conclusions. They show that overreporting of news consumption is more prominent in the presence of political cues. Based on this discussion, we would expect self-reporting of news sources to increase with outlets’ reach and partisan slant in the open-ended condition. This leads to our third set of hypotheses: *In the open-ended format respondents will be more likely to self-report visits to outlets with greater reach (H3a) and a partisan slant (H3b).*

Country differences

Previous research has shown that context may influence self-reporting of certain types of political behavior (e.g., turnout) affected either by a social desirability bias (Karp & Brockington, 2005) or other problems (Prior 2009a). Yet, in general, the role of context in self-reporting behavior, and of news consumption in particular, has been largely underexplored. Most studies on self-reporting of news consumption are US-based, thus, we do not know whether the same pattern holds in other contexts. Moreover, existing studies have not compared the performance of self-reported measures of exposure across countries, and therefore we do not know how context may affect accuracy of survey-based measures.

Here, we have an opportunity to examine whether accuracy in self-reported measures of news consumption varies across two different regulatory models and media systems. Although with two cases only we cannot statistically test the influence of contextual variables in self-reporting, we can speculate about the amount of error across the two countries of study based on some key characteristics.

One factor that might affect accuracy in self-reporting is the scope of news consumption. We would expect that the lower the number of people consuming news the higher the room for overreporting (Karp & Brockington, 2005). In this respect, compared to the UK (and other Northern European models), Spain has a lower reliance on traditional news sources (Castro et al., 2021), stemming from lower levels of reach of the press market, a weaker journalistic professionalization, and the comparatively weaker role of the public service media (PSM). This readership gap concerning traditional media has been more recently compensated by the growth of online media. However, relying primarily on online news sources might seriously compromise people's capacity to recall (Niederdeppe, 2016), increasing misreporting.

A second factor likely to affect accuracy in self-reporting concerns the structure of the online news domain, particularly whether it is more or less fragmented.² A more populated media supply might increase overreporting by setting a higher reference point (or anchor) when it comes to reporting visits to online news sources (Prior 2009a, Andersen et al., 2016). At the same time, it might increase the difficulty to accurately recall news outlets (Niederdeppe, 2016), increasing underreporting when people are prompted to report news visits in an open-ended question. In general, greater fragmentation of the online news domain would be expected to decrease accuracy.

² Fragmentation here is not defined from an audience-centric approach (Ksiazek, 2011; Fletcher & Nielsen, 2017; Majó-Vázquez et al., 2019), but from a media perspective. At the media level, fragmentation takes on different values as a function of (i) the number of available sources and (ii) the distribution of the audience across these sources (Majó-Vázquez, 2017). The larger the set of available sources and the more widely distributed the audience is across these sources, the higher media fragmentation. Conversely, the smaller the number of available outlets and the more concentrated the audience is around a few outlets, the lower media fragmented.

Previous research, using an audience-centric approach, suggests that the Spanish (online) news domain tends to be more fragmented than the UK from a demand-side perspective (Majó-Vázquez et al., 2019; Fletcher & Nielsen, 2017). This essentially results from a high concentration of the UK news audience around the PSM, i.e., BBC and a more fragmented or scattered distribution of the Spanish news audience across several relevant sources. In brief, the greater fragmentation of the Spanish online news domain partly explained by a weaker public broadcasting service (Majó-Vázquez et al., 2019: 228), a stronger interest in local press (Schulz, 2020) and a fast proliferation of digital-born news outlets (Majó-Vázquez et al., 2019: 228; Nortes, 2019), might lead to expect lower accuracy in self-reporting. This leads to our last hypothesis: *Accuracy will be lower in Spain than in the UK (H4).*

Data, measures, and methods

Data

This study is based on three different datasets combining survey and web-tracking behavior data for both Spain (N= 652) and the UK (N = 508). As part of a larger project, between April 15 and April 28, we collected participants' socio-demographic attributes and media habits; we then tracked their online navigation patterns; and finally between May 15 and May 26, we asked participants to report all news sources they remembered having visited in the previous 15 days. For this study, we only used the web-tracking data from the 15 days before answering the second survey.

The recruitment of panelists and the data collection was done by two different market research companies, Netquest and YouGov. Both companies have consolidated panels in their countries of origin, Spain and UK, and extensive experience in working with web-tracking technology.

The characteristics of panelists tend to match those of the online population in that they tend to be slightly more educated and interested in politics than the general population (a summary of all the key variables is included in Table A1 of the Online Supplementary Material, p. 3). To check for the validity of our web-tracking data we cross-checked the most visited news outlets among our panelists with the top-most visited online news outlets according to an external source, Comscore. Using the Spearman rank-order method the correlation is respectively .91 and .92 for Spain and the UK, which speaks for the representativeness of our data.

In both countries we collected data of visits from mobile devices, including visits from news applications. In Spain, in addition to mobile, we collected visits from desktop. Hence, our web-tracking datasets have different structures: in the UK they include only visits from mobile devices, while in Spain they include visits from both mobile and desktop. We are aware that using different datasets may affect the results and hinder comparability. For this reason, we replied all the analyses using only mobile-tracked visits for Spain.³

In Spain, most participants (225) sent data from both devices, 156 participants had at least one URL recoded from the mobile device but no desktop, 145 participants produced at least one URL from the desktop but none from the mobile device, and 126

³ We replicated the analyses on two datasets. To create the first dataset, we filtered in individuals who visited outlets only from their mobile devices during the whole study period. To create the second dataset, we filtered in visits from mobile devices. This dataset, thus, includes individuals using both devices but keeps only the visits from mobile devices. A more detailed explanation of these datasets is offered in pages 19 and 24 of the Online Supplementary Material. To make datasets comparable to the UK dataset and avoid underestimating overreporting, in both datasets we included individuals that were inactive, who are more likely to overreport (see analyses in the [Online Supplementary Material, pp. 19-29](#); see also Prior 2009a).

participants produced no tracking data.⁴ In the UK, 108 participants produced no tracking data during the study period.

During the 15 day-time window, the mean number of visits to news outlets from the mobile device in Spain was 17,8, and from the desktop device, 37,7. The median visits for each was 2. In the UK, the mean number of visits (always from the mobile device) was 41,2. The median visits was 6. For Spain and the UK, the minimum number of visits was 0, and the maximum was, respectively, 1,557 (from desktop) and 1,254 (from mobile). This information is included in Table A1 of the Online Supplementary Material (p. 3).

Experimental Design

After having collected participants political attitudes and socio-demographics, in the second survey we randomly assigned participants to one of three condition, each consisting on a different question format to recall their past news consumption —open-ended, check-all, forced-choice. The distribution of participants across conditions for both countries can be found in Table A2 of the Online Supplementary Material (p. 4).

Measures of key variables

Open-ended. In the open-ended condition, participants were prompted to the following

⁴ Following Jürgens et al. (2020), we checked for differential sampling bias by device.

Individuals making use of their mobiles for news visits differ from those not using this device on college levels and ideology. On average, mobile news users are less educated (college education is 7% lower among this group) and less conservative (conservatism decreases 0.02 among this group). In most other relevant sociodemographic and attitudinal variables (age, gender, political interest, news interest....) mobile news users are indistinguishable from non-mobile users.

instructions: “Please list any websites or mobile apps that you have visited in the past 15 days for news. Take some time to ensure that you think of all the sites you have visited” (question formulation). Following it, participants saw a blank sheet with a set of empty spaces, which they had to fill in with the brand names of the sites they remembered having navigated. A screen capture of what respondents saw in each condition is shown in Figure A2 in the Online Supplementary Material (pp. 4-6).

Check-all. The check-all condition, adapted from Annenberg, read: “Which of these websites or mobile apps have you visited or used in the past 15 days for news, if any? Select ALL answers that apply”. Following the question, participants saw a list of 31 (Spain) and 28 (UK) news websites –the top-most visited news outlets in each country according to Comscore. Departing from previous work, here we used brand images exploiting the visual side of the web (Dillman et al. 2007) and therefore to increase the external validity of the question by easing recall through visual recognition.

Forced-choice. In the forced-choice condition, participants were shown an ordered list of news websites using their logos (the same list than in the check-all condition for each country) with the choices “Yes and “No” next to each news outlet’s logo. This followed the question *Have you visited or used any of these websites or mobile apps in the past 15 days for news?*

Observed exposure. Our measure of observed exposure is restricted to visits to the list of 31 (Spain) and 28 (UK) news websites provided by researchers in the check-all and forced-choice question. We have compiled these lists of media outlets using an external source data Comscore, which offers observed tracking data based on panels that include 25,000 individuals in Spain and 50,000 individuals in the UK. Using this external source has allowed us to include news media outlets with very small audience reach which are not usually identified in survey-based studies. For instance, this is the

case of vilaweb.cat in Spain with a reach of ~3% of the total digital population, or Breitbart.com in the UK with 0.3% of the total digital population. The full list of news media outlets and their reach can be seen in Table A3 of the OSM (p. 7). The same list of outlets have been used to automatically code the responses included in the open-ended question.

While participants may have visited news outlets in our list several times, our measure of observed exposure is binary and only accounts for whether they visited or not a given outlet (j) during the 15-day period (t). This matches the survey question, which does not ask about frequency. Observed exposure therefore is a dichotomous variable measured at both the individual and outlet level that adopts a 1 if individual i visited outlet j during the study period t , and a 0, otherwise.

We used three different measures of accuracy – overreporting, underreporting and misreporting -- summarized in Table 1 (here we followed Guess 2015).

TABLE 1 about here

Each line in Table 1 represents one of four possible combinations of reported and observed exposure to a given site for each individual. Based on these four possible combinations, each of the three measures of accuracy receives a 1 or a 0. Note that only in two of the four situations, accuracy measures may receive a 1, when an individual has visited a site j but does not report having visited it (underreporting) and when he/she has not visited a site j but reports having visited it (overreporting). In both these instances, misreporting receives a 1.

Our measure of outlet size is the market share according to a well-reputed market media company, Comscore. Market share according to Comscore is highly correlated to outlet reach in our data (Spain: .93, UK = .75, using Pearson correlation). This can be seen in Figure A3 in the Online Supplementary Material (p. 8), which plots

market share according to Comscore against the percentage of visitors (i.e., reach) for each outlet in our database, for both countries.

To classify outlets as slanted (or partisan) and neutral we used input information provided by our panelists. In the survey's first wave, participants were asked to locate each media outlet (provided they had visited it) on a 0-10 scale, where 0 was "Very left-wing", and 10, "Very right-wing". We classified an outlet as partisan if it fulfilled two conditions: 1) at least 50% of respondents perceived it as right- (> 5 in the scale) or left- (< 5 in the scale) leaning and 2) the outlet's mean position in the scale was either below .45 (left) or above .55 (right). Following these criteria, in Spain, 17 outlets of the list of 31 (60%) were classified as partisan; in the UK, 17 outlets of the list of 28 (65%) were classified as partisans. (Table A4 in p. 9 of the Online Supplementary Material offers the details of the classification).⁵

Finally, political interest is assessed by asking: "How much you would say you are interested in politics: Very much; quite interested; hardly interested; not at all?"

Controls

We use a series of controls in our models that include basic socio-demographics (e.g. age, gender, college education), political attitudes (left-right self-placement, interest in news), and frequency of news consumption measured by total observed visits to news

⁵ We also computed several audience-based measures with our observed and reported data following previous studies (Flaxman et al., 2016; Fletcher et al., 2019; Cardenal et al., 2019), but found several inconsistencies, which we attributed to the unconventional period of our analysis. To cross-validate our slant measure, we compared our results with an external survey, the Digital News Report. We found that a few results differ. But based on our knowledge of the media systems that we studied, we believe that our measure is consistent.

outlets. In Spain, to control for the effect of the device on accuracy observed visits are disaggregated by device (i.e., desktop, mobile). All of these variables can be seen as affecting the path between observed visits and retrospective self-reports.

Methods

To analyze accuracy, we compare reported and observed visits by outlet at the aggregate level and then use OLS regression to more easily and reliably quantify accuracy across questions. We also use multi-level logistic regression to test for the influence of outlets' characteristics on accuracy.

Analysis

Question performance

The analysis we present here (and in the following sections) are based on both desktop and mobile visits for Spain, but the results do not change when we only use mobile-tracked visits for this country (these results can be accessed in pages 19-29 of the OSM).

H1a and H1b state that overreporting will be highest in the forced-choice format, followed by the check-all question, and the open-ended question. As a quick glimpse to Figure 1 reveals our expectations overall fit quite well the pattern for Spain (top panel).

FIGURE 1 ABOUT HERE

In Spain, none of the questions seem to give very accurate estimates of observed visits. Yet, two of them provide more biased estimates of observed exposure, by either systematically underreporting observed visits (the open-ended question) or systematically overreporting it (forced-choice question). In contrast, inaccuracy is not as pronounced in the UK.

To further see this, we computed the mean differences between observed and reported visits by outlet and the uncertainty around these differences using the underlying data of observed and reported visits per outlet and respondent. Figure 2 plots these differences and the uncertainty around them using a 95% confidence interval.

For Spain (top panel of this figure), estimates of these differences and ci in the forced-choice and open-ended conditions fall, respectively, above or below the horizontal line (which stands for perfect matching between observed and reported visits). In contrast, in the check-all condition, most estimates and ci hit or fall around the horizontal line, implying that on average they hit the target more frequently than in the other two conditions (note that accuracy focuses on how likely estimates are to hit the target not on the precision of these estimates). In the UK (bottom panel), results are similar. Estimates of differences and ci for the open-ended question fall systematically below the horizontal line, implying that this question produces systematic underreporting. However, the differences in accuracy between the check-all and forced-choice seem not to be significant.

FIGURE 2 ABOUT HERE

To quantify accuracy, we regressed our three accuracy measures (overreporting, underreporting and misreporting) on the type of question and a set of controls (see methods section). We used OLS regression to obtain a direct interpretation of the coefficients, which read as the increase in overreporting, underreporting, and misreporting of moving from the check-all question (the reference category) to the open-ended and forced-choice conditions. Here, we report only the effects for overreporting (the effects for misreporting as well as the tables with the coefficients for our three dependent variables are included in the Online Supplementary Material [p. 10-11])

FIGURE 3 ABOUT HERE

In Figure 3 we can see that the effect on overreporting of moving from the check-all question to the open-ended and forced-choice conditions is much larger in Spain than in the UK. In Spain, moving from the check-all question to the open-ended condition decreases overreporting by more than two outlets (2.13), and moving from the same condition to the forced-choice format increases overreporting by 2 outlets and a quarter (2.25).

In contrast in the UK the effect of moving from the check-all question to the open-ended and forced-choice conditions is much smaller. In particular, moving from the check-all question to the open-ended condition decreases overreporting by almost one outlet (0.9), while moving from the check-all format to the forced-choice condition increases it by a little bit more than a half (0.52).

Political interest and accuracy

H2 stated that politically interested people would overreport more. Figure 3 shows the coefficients for several individual predictors, among them political interest. As we can see political interest significantly increases overreporting in both countries, and roughly by the same amount –by less than half an outlet (Spain: $B: .40$, $se: .21$; UK: $B: .39$, $se: .17$). This means that moving from the bottom to the top of the scale in political interest (from none to a lot) increases overreporting by more than one outlet in both countries ($\cong 1,2$). In the UK, however, political interest also increases underreporting by the same amount ($.32$), and the effect is statistically significant.

As for other significant predictors of accuracy, age is negatively associated to both overreporting and misreporting in Spain, and this result is in line with previous studies which found age a predictor of error in reporting activity on social media

(Ernala, 2020) and more precisely showed that younger people were more prone to misreport their time spent on Facebook. Also, in the UK going to college is negatively associated to misreporting (see Figure A4 in the Online Supplementary Material, p. 10).

Finally, it is also worth noting that the number of visits from the mobile device is positively associated to misreporting in Spain. This effect is larger and more significant ($B = .021$; $se = .004$) than the one for desktop ($B = .004$; $se = .002$)

Outlets as a source of bias in self-reporting

Our third set of hypotheses state that overreporting will increase with outlet reach (H3a) and partisan slant (H3b) in the open-ended condition. To test these effects, we ran three interaction models: in the first, we interacted the type of question with outlet reach; in the second, with both reach and reach squared (to allow for non-linear effects); in the third, with the outlet's partisan condition.

We found that self-reporting of sources increases with reach in all conditions but especially in the open-ended one. However, whether in the UK these effects are linear across all conditions, in Spain they are only linear in the open-ended condition. In other words, whereas in Spain in the open-ended condition the probability of reporting an outlet increases monotonically with reach, in the other two conditions the likelihood of reporting an outlet is maximized at medium levels of reach (see Figure A5 in p. 16 of the Online Supplementary Material). This suggests that in more fragmented online media environments check-all and forced-choice questions might help to recall medium sized outlets.

We also found that the likelihood of reporting an outlet that is partisan increases especially in the open-ended condition. In Spain, being prompted to the open-ended condition increases the likelihood of reporting a partisan outlet vis a vis the check-all question by 6%, although the effect is not significant. In the UK, in contrast, being

prompted to the open-ended question increases the likelihood of reporting partisan media by 30%, and the effect is highly significant (see Figure A6 in p. 16 of the Online Supplementary Material for a graphical visualization of these effects).

Media system influence

Our main rationale for our expectations concerning country differences on accuracy (H4) was that Spain has a more fragmented online news domain than the UK from a supply perspective. To see this, we plotted the distribution of the number of outlets visited by individuals during the study period for both countries (see Figure A7 in the Online Supplementary Material, p. 17). We found that the distribution for Spain is slightly flatter and has a longer tail, implying that more people tend to visit a higher number of outlets. The mean and maximum number of outlets visited during this period further confirms this. In Spain, panelists visited an average of 4.6 outlets, and a maximum of 23, while in the UK, the average was 3, and the maximum 14. This provides some support for the underlying mechanism in H4.

Discussion

Surveys are imperfect measurement tools of human behavior. Yet, scholars continue relying primarily on surveys to learn about media effects, so more effort needs to be placed on improving survey-based measures of media exposure. Here, we have used true exposure as a benchmark to test the performance of three survey question formats to measure news exposure online across two contexts, Spain and the UK. Below we highlight the methodological and theoretical contributions of this study and provide some recommendations.

Methodological contributions

Our study makes several methodological contributions. First, it shows that the check-all question provides on average the most accurate (non-biased) estimates of observed exposure, although not necessarily the most precise ones. Indeed, the open-ended question tends to produce more precise estimates (i.e., reducing random error) but biased ones, since they systematically underreport news exposure (i.e., increasing systematic error). As for the forced-choice question it tends to increase both systematic and random error, although the performance of this question depends very much on the context.

Second, in line with previous research, this study shows that political interest systematically increases misreporting. Third, our results show that outlets' characteristics may increase systematic error in self-reporting. For example, in the open-ended question outlets that have the largest reach and are partisan are more likely to be reported because they are more easily recalled. In contrast, due to the familiarity bias, medium-sized outlets are more likely to be overreported in the check-all and forced-choice.

Theoretical contributions

The main theoretical contribution of this study is that context matters for accuracy. Overall, accuracy was much lower in Spain than in the UK across all three questions. We suggest that this is mainly related to the level of fragmentation of the online news environment, since a higher number of new entrants increases the probability of error in reporting visits to news sources.

Recommendations

Based on this study we can offer the following set of recommendations. First, in fragmented online news domains (e.g., Spain), always use the check-all question. This

question not only provides more accurate (i.e., less biased) estimates of observed exposure, but also tends to reflect the fragmentation of the online environment more faithfully by increasing the likelihood of reporting medium -sized outlets. Second, in concentrated online news domains (e.g., UK), the choice of the question format matters less, since all questions perform quite well by increasing the likelihood of reporting the most popular outlets. Third, avoid using open-ended questions if the outcome of interest is correlated with the partisan condition of outlets (e.g., selective exposure), since partisan media are more likely to be reported in this type of question.

As any other study, this one also suffers from limitations. The most important one is that two different companies, with slightly different technologies and practices, performed the tracking and the data collection. We know that different practices and technologies might produce different kinds of errors and biases in the web-tracking data (Jürgens et al., 2020). Moreover, the two companies provided different datasets for observed exposure. To minimize this problem, we first checked for selection bias across devices in Spain and found that most differences between mobile and desktop users were insignificant. Second, we replicated all the analysis on two mobile datasets (one based on mobile users and another on mobile visits) for Spain, and the results did not change. In spite of these and other limitations, we believe our study contributes to a burgeoning literature assessing the validity of exposure measures through the use of “digital footprints” and adds an important layer to the collective effort of improving survey measures of (online) news exposure by highlighting the significance of context.

References

- Althaus, S. L., & Tewksbury, D. H. (2007). *Toward a new generation of media use measures for the ANES*. Report to the Board of Overseers of the ANES.
- Andersen, K., H. de Vreese, C., & Albæk, E. (2016). Measuring media diet in a high-choice environment-Testing the list-frequency technique. *Communication Methods and Measures*, 10(2-3), 81-98.
- Cardenal, A. S., Aguilar-Paredes, C., Cristancho, C., & Majó-Vázquez, S. (2019a). Echo-chambers in online news consumption: Evidence from survey and navigation data in Spain. *European Journal of Communication*, 34(4), 360-376.
- Cardenal, A. S., Aguilar-Paredes, C., Galais, C., & Pérez-Montoro, M. (2019b). Digital Technologies and Selective Exposure: How Choice and Filter Bubbles Shape News Media Exposure. *The International Journal of Press/Politics*, 24(4), 465-486.
- Castro, L., Strömbäck, J., Esser, F., Van Aelst, P., de Vreese, C., Aalberg, T., ... & Theocharis, Y. (2021). Navigating high-choice European political information environments: A comparative analysis of news user profiles and political knowledge. *The International Journal of Press/Politics*.
<https://doi.org/10.1177/19401612211012572>
- de Vreese, C. H., & Neijens, P. (2016). *Measuring media exposure in a changing communications environment*. Taylor & Francis.
- Dilliplane, S., Goldman, S. K., & Mutz, D. C. (2013). Televised exposure to politics: New measures for a fragmented media environment. *American Journal of Political Science*, 57(1), 236-248.
- Dillman, D. A., & Smyth, J. D. (2007). Design effects in the transition to web-based surveys. *American journal of preventive medicine*, 32(5), S90-S96.

- Ernala, S. K., Burke, M., Leavitt, A., & Ellison, N. B. (2020). How well do people report time spent on Facebook? An evaluation of established survey questions with recommendations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–14).
- Fletcher, R., & Nielsen, R. K. (2017). Are news audiences increasingly fragmented? A cross-national comparative analysis of cross-platform news audience fragmentation and duplication. *Journal of Communication*, 67(4), 476-498.
- Geer, J. G. (1991). Do open-ended questions measure “salient” issues?. *Public Opinion Quarterly*, 55(3), 360-370.
- Guess, A. M. (2015). Measure for measure: An experimental test of online political media exposure. *Political Analysis*, 23(1), 59-75.
- Guess, A., Munger, K., Nagler, J., & Tucker, J. (2019). How accurate are survey responses on social media and politics? *Political Communication*, 36(2), 241–258.
- Haenschen, K. (2020). Self-reported versus digitally recorded: Measuring political activity on Facebook. *Social Science Computer Review*, 38(5), 567–583.
- Iyengar, S. (1990). The accessibility bias in politics: Television news and public opinion. *International Journal of Public Opinion Research*, 2(1), 1-15.
- Jürgens, P., Stark, B., & Magin, M. (2020). Two half-truths make a whole? On bias in self-reports and tracking data. *Social Science Computer Review*, 38(5), 600-615.
- Kahn, A. S., Ratan, R., & Williams, D. (2014). Why we distort in self-report: Predictors of self-report errors in video game play. *Journal of Computer-Mediated Communication*, 19(4), 1010-1023.
- Kalogeropoulos, A., Fletcher, R., & Nielsen, R. K. (2019). News brand attribution in distributed environments: Do people know where they get their news? *New Media & Society*, 21(3), 583-601.

- Karp, J. A., & Brockington, D. (2005). Social desirability and response validity: A comparative analysis of overreporting voter turnout in five countries. *The Journal of Politics*, 67(3), 825-840.
- Ksiazek, T. B. (2011). A network analytic approach to understanding cross-platform audience behavior. *Journal of Media Economics*, 24(4), 237-251.
- Konitzer, T., Allen, J., Eckman, S., Howland, B., Mobius, M. M., Rothschild, D. M., & Watts, D. (2020). Comparing estimates of news consumption from survey and passively collected behavioral data. *Available at SSRN 3548690*.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology*, 5(3), 213-236.
- Landert, D. (2014). Blurring the boundaries of mass media communication? Interaction and user-generated content on online news sites. *Studies in Variation, Contacts and Change in English*, 15.
- Majó-Vázquez, S. (2017). Digital News In Spain: Characteristics and Effects of Online News Production and Consumption. Ph.D Thesis, Universitat Oberta de Catalonia, April 2017.
- Majó-Vázquez, S., Nielsen, R. K., & González-Bailón, S. (2019). The backbone structure of audience networks: A new approach to comparing online news consumption across countries. *Political Communication*, 36(2), 227-240.
- Naab, T. K., Karnowski, V., & Schlütz, D. (2019). Reporting mobile social media use: How survey and experience sampling measures differ. *Communication Methods and Measures*, 13(2), 126-147.
- Nagler, R. H. (2017). Measurement of media exposure. *The International Encyclopedia of Communication Research Methods*, 1-21.

- Newman, N., Fletcher, R., Schulz, A., Andi, S., & Nielsen, R. K. (2021). Digital news report 2020. *Reuters Institute*, 2020-06.
- Niederdeppe, J. (2016). Meeting the challenge of Measuring Communication exposure in the Digital Age. *Communication Methods and Measures*, 10(2-3), 170-172.
- Nortes, S. (2019). Making local neighbourhoods great again: In Spain hyperlocal newspapers are all the rage because they involve readers. *Index on Censorship*, 48(1), 60-61.
- Parry, D. A., Davidson, B. I., Sewall, C. J., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour*, 1-13.
- Price, V., & Zaller, J. (1993). Who gets the news? Alternative measures of news reception and their implications for research. *Public opinion quarterly*, 57(2), 133-164.
- Prior, M. (2009a). Improving media effects research through better measurement of news exposure. *The Journal of Politics*, 71(3), 893-908.
- Prior, M. (2009b). The immensely inflated news audience: Assessing bias in self-reported news exposure. *Public Opinion Quarterly*, 73(1), 130-143.
- Schulz, A. (2020). Global Turmoil in the Neighbourhood: Problems Mount for Regional and Local News. *Reuters Institute Digital News Report 2020*, 47-51.
- Scharkow, M. (2016). The accuracy of self-reported internet use—A validation study using client log data. *Communication Methods and Measures*, 10(1), 13–27.
- Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *The American Journal of Evaluation*, 22(2), 127-160.

- Smyth, J. D., Dillman, D. A., Christian, L. M., & Stern, M. J. (2006). Comparing check-all and forced-choice question formats in web surveys. *Public Opinion Quarterly*, 70(1), 66-77.
- Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). Integrating Survey Data and Digital Trace Data: Key Issues in Developing an Emerging Field. *Social Science Computer Review*, 38(5), 503-516.
- Sudman, S., Bradburn, N.M., Schwarz, N. (1996). *Thinking about Answers. The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*.
- Van Aelst, P., Toth, F., Castro, L., Štětka, V., Vreese, C. D., Aalberg, T., ... & Theocharis, Y. (2021). Does a Crisis Change News Habits? A Comparative Study of the Effects of COVID-19 on News Media Use in 17 European Countries. *Digital Journalism*. <https://doi.org/10.1080/21670811.2021.1943481>
- Van Spanje, J., & De Vreese, C. (2014). Europhile media and Eurosceptic voting: Effects of news media coverage on Eurosceptic voting in the 2009 European parliamentary elections. *Political Communication*, 31(2), 325-354.
- Vraga, E., Bode, L., & Troller-Renfree, S. (2016). Beyond self-reports: Using eye tracking to measure topic and style differences in attention to social media content. *Communication Methods and Measures*, 10(2-3), 149-164.
- Vraga, E. K., & Tully, M. (2020). Who is exposed to news? It depends on how you measure: Examining self-reported versus behavioral news exposure measures. *Social Science Computer Review*, 38(5), 550-566.
- Zaller, J. (1992). *The nature and origins of mass opinion*. New York: Cambridge University Press.

Table 1. Summary of accuracy measures

Exposure		Measures		
Observed	Reported	Overreporting	Undereporting	Misreporting
0	0	0	0	0
1	0	0	1	1
0	1	1	0	1
1	1	0	0	0

Figure 1. Difference between reported and observed visits by outlet for Spain (above) and UK (below).

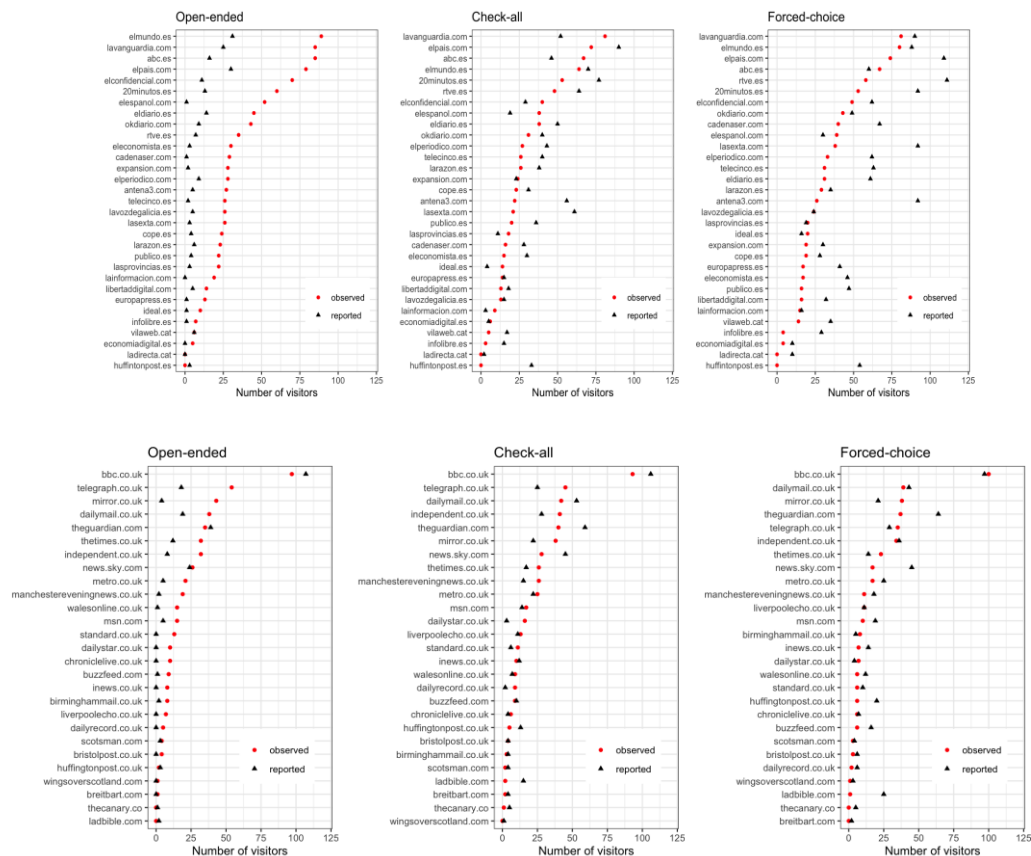


Figure 2. Mean difference between observed and reported visits by outlet for Spain (above) and the UK (below).

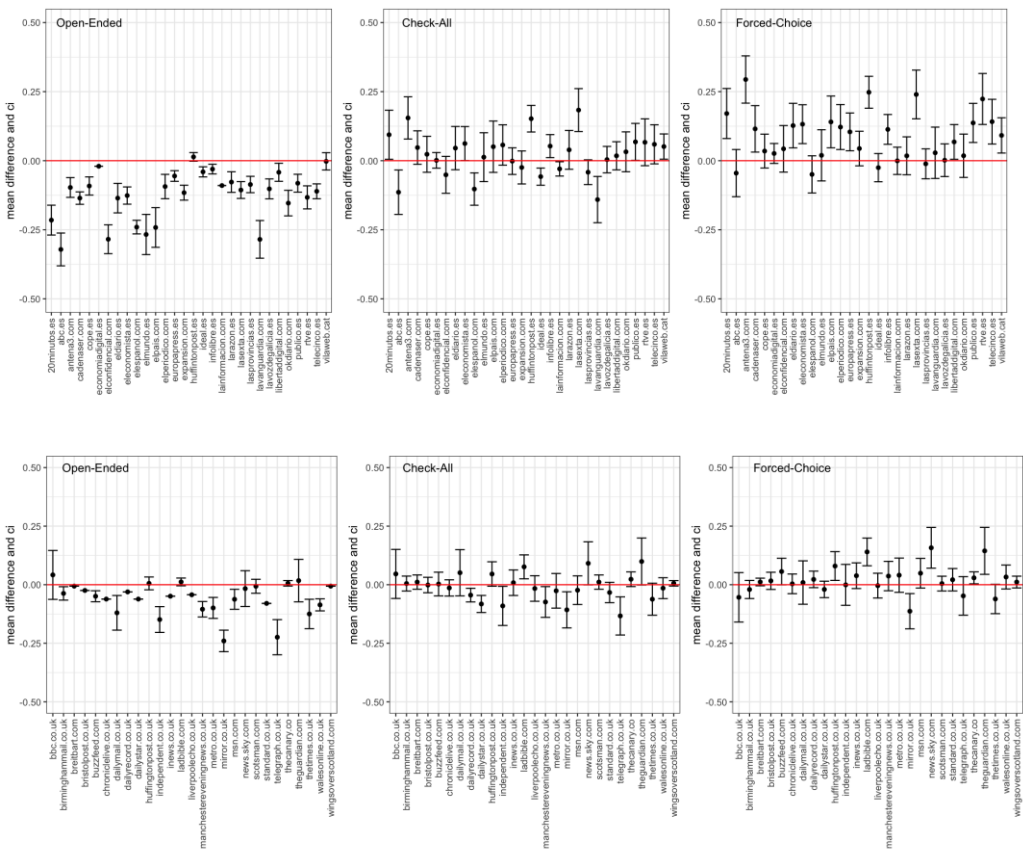


Figure 3. OLS coefficients for overreporting in Spain and the UK

