



## RESEARCH NOTE

# The Impact of Confidentiality Assurances on Participants' Responses to Sensitive Questions

Carmen M. Leon <sup>1</sup>, Eva Aizpurua <sup>2,3</sup>, and  
Sophie van der Valk<sup>3</sup>

<sup>1</sup>School of Law, University of Castilla-La Mancha, Albacete, Spain <sup>2</sup>European Social Survey ERIC, University of London, London, UK <sup>3</sup>School of Law, Trinity College Dublin, Trinity Research in Social Sciences (TRiSS), Dublin, Ireland

Privacy, anonymity, and confidentiality are issues often addressed when collecting sensitive information (Tourangeau, 2017). Despite the importance of the concept, there is no consensus on what constitutes a sensitive question (Tourangeau, Rips, & Rasinski, 2000). According to Tourangeau and Yan (2007), a question is considered sensitive when it asks for a socially undesirable answer, requesting that the respondent admits that he or she has violated a social or legal norm. Due to the potential consequences of admitting some socially undesirable beliefs, behaviors, or activities (e.g., social or legal sanctioning), sensitive questions are often associated with systematic measurement error (Groves et al., 2004). For instance, research has shown that sensitive questions have comparatively higher item nonresponse rates than nonsensitive questions, and they are regarded as one of the major causes for socially desirable answers in surveys (Diekmann, 2003; Krumpal, 2013; Näher & Krumpal, 2012; Tourangeau et al., 2000).

Over the years, numerous methodological experiments have been conducted to analyze the impact of several factors on reducing measurement error in sensitive questions. Survey mode has been found to be one of the most important factors in explaining social desirability bias in sensitive questions (Holbrook, Green, & Krosnick, 2003; Tourangeau & Yan, 2007; Turner et al., 2005). It has been found that, in general, when interviewers collect the data or when they are present during data collection, responses are more socially desirable than in self-administered surveys (Kreuter, Presser, & Tourangeau, 2008; Tourangeau & Yan, 2007).

---

\*All correspondence concerning this article should be addressed to Carmen M. Leon, PhD, School of Law, University of Castilla-La Mancha, 1 Plaza de la Universidad, 02071 Albacete, Spain. E-mail: Carmen.Leon@uclm.es

Apart from self-administration, other strategies have been used to reduce social desirability bias when asking about sensitive topics. For example, a common approach is the randomized response technique (Warner, 1965), in which respondents employ a randomizing method to add probabilistic misclassification to their responses and thus conceal their true answers from the interviewers (Kuha & Jackson, 2014). The studies using this method have found mixed results and there is no consensus about its effectiveness (for a review, see Krumpal, 2013). Another strategy is the item count technique, also known as the list experiment (Glynn, 2013; Miller, 1984). This approach protects the anonymity of participants by giving respondents a list of items and asking them how many, not which ones, they support. Despite being heavily used, it is not clear that this strategy improves the accuracy of self-reported sensitive information (Holbrook & Krosnick, 2010; Lax, Phillips, & Stollwerk, 2016; Tourangeau & Yan, 2007). Question wording has also received great attention. Previous research has shown the importance of presenting questions in a neutral way (Barton, 1958; Groves et al., 2004), and the relevance of other contextual features, such as confidentiality assurances, on participants' responses (Singer, Hippler, & Schwarz, 1992; Singer, Von Thurn, & Miller, 1995). For instance, in a review analyzing research reports conducted primarily in the United States, Singer et al. (1995) found that confidentiality assurances generated low item nonresponse, high survey response rates, and high response quality for sensitive items. However, when items were nonsensitive, confidentiality assurances resulted in lower response rates (Singer et al., 1992). These inconclusive results might indicate that the effectiveness of this type of messages is mediated by the sensitivity of the questions and highlight the importance of further research, especially in other cultural contexts.

Although previous research has documented the influence of several features on the report of sensitive data (e.g., data collection mode, interviewer effects, presence of third parties), few studies have analyzed the impact of confidentiality assurances on participants' responses, especially outside of the United States. This is important as research shows that questions may have different degrees of sensitivity in different countries and cultures, across social groups within the same country, and over time (Andreenkova & Javénile, 2019). Furthermore, country-level characteristics play a direct role in explaining cultural differences in rates of interview privacy and moderate the effect of respondent-level characteristics on privacy (Mneimneh, Elliott, Tourangeau, & Heeringa, 2018). To date, the evidence is mixed and there is no consensus as to the benefit of including this type of messages in reducing social desirability bias. In addition, with the growing use of smartphones to participate in survey research, the inclusion of additional text might be problematic due to the limited screen space on these devices.

The aim of this study is to analyze the effect of including an extra sentence about the confidentiality of participants' responses immediately before a series of questions measuring sexism. Drawing on previous literature, it is expected that socially desirable responding will be more likely to occur in the group which did not receive the extra sentence about confidentiality when compared with the group which received it. The results of the study will provide recommendations for survey researchers and practitioners collecting sensitive data.

Figure 1.

Screenshots of the two conditions for smartphone and tablet respondents with confidentiality assurance text highlighted.

The figure displays two side-by-side screenshots of a survey interface. Both screens show the same text: "A continuación, se presentan una serie de afirmaciones sobre los hombres y las mujeres y sobre su relación en nuestra sociedad. Lee cuidadosamente cada una de estas afirmaciones e indica en qué grado estás de acuerdo o en desacuerdo con ellas." Below this is a statement: "Las mujeres, por lo general, no tienen mucho talento". Underneath the statement are five radio button options: "Muy en desacuerdo", "En desacuerdo", "Ni de acuerdo ni en desacuerdo", "De acuerdo", and "Muy de acuerdo".

The left screenshot has a black-bordered box highlighting the text: "Recuerda que tus respuestas son confidenciales y que no hay respuestas correctas o incorrectas." The right screenshot is identical but does not have this highlighted box.

## Material and Methods

### Data Collection

The online survey was conducted between January 7 and 29, 2019, using the Netquest panel in Spain. Quotas for age, gender, and habitat were used in order to obtain a sample distribution similar to the Spanish population. A total of 1,007 questionnaires were completed using PCs (57.8%), smartphones (36.5%), and tablets (5.7%). Overall, the questionnaire was completed by 92.3% of those panellists invited. The questionnaire was administered in Spanish and included 51 questions about perceptions and attitudes toward intimate partner violence. The questionnaires took approximately 12 mins to complete ( $M = 11.72$ ,  $SD = 6.32$ ). Table 2 provides the wording of the questions used in the analysis.

### Experimental Design and Variables of Interest

The study used a one-factor (extra sentence vs. no extra sentence) between-subjects design. Participants were randomly assigned to one of the two conditions (extra sentence,  $n = 503$ ; no sentence,  $n = 504$ ). The extra sentence was placed immediately before the first statement and read "remember that your answers are confidential and that there are no right or wrong answers" (screenshots of the two conditions for smartphone/tablet and PC responses are displayed in Figures 1 and 2—the order of items was randomized and PC respondents were presented with grids instead of individual questions).

Figure 2. Screenshots of the two conditions for PC respondents with confidentiality assurance text highlighted.

A continuación, se presentan una serie de afirmaciones sobre los hombres y las mujeres y sobre su relación en nuestra sociedad. Lee cuidadosamente cada una de estas afirmaciones e indica en qué grado estás de acuerdo o en desacuerdo con ellas.	Muy en desacuerdo	En desacuerdo	Ni de acuerdo ni en desacuerdo	De acuerdo	Muy de acuerdo
El movimiento feminista no sirve para nada.	0	0	0	0	0
Las mujeres están mejor preparadas para cuidar de los niños y ancianos.	0	0	0	0	0
El nivel educativo de las mujeres en la población es poco común.	0	0	0	0	0
La discriminación contra las mujeres ya no es un problema en España.	0	0	0	0	0
Los hombres y las mujeres necesitan igual oportunidades laborales.	0	0	0	0	0
Resulta sencillo comprender los puntos de vista que tienen los grupos de mujeres.	0	0	0	0	0
La sociedad trata por igual a hombres y mujeres.	0	0	0	0	0
El currículo escolar debe adaptarse a las necesidades que ve con las mujeres.	0	0	0	0	0
El Estado posee demasiados refugios en los temas que tienen las mujeres.	0	0	0	0	0
Las mujeres, por lo general, son buenas conductoras.	0	0	0	0	0
Prefero tener de jefe a un hombre antes que a una mujer.	0	0	0	0	0
El trabajo de un hombre es más importante que el trabajo de una mujer.	0	0	0	0	0
Me gustaría ver a una mujer como presidenta de España.	0	0	0	0	0
Las mujeres por lo general no tienen mucho talento.	0	0	0	0	0
Las mujeres son buenas con el razonamiento lógico.	0	0	0	0	0

A continuación, se presentan una serie de afirmaciones sobre los hombres y las mujeres y sobre su relación entre hombres y mujeres. Lee cuidadosamente cada una de estas afirmaciones e indica en qué grado estás de acuerdo o en desacuerdo con ellas.	Muy en desacuerdo	En desacuerdo	Ni de acuerdo ni en desacuerdo	De acuerdo	Muy de acuerdo
Se deben adoptar mejores medidas para lograr la igualdad laboral entre hombres y mujeres.	0	0	0	0	0
Las mujeres son buenas con el razonamiento lógico.	0	0	0	0	0
La sociedad trata por igual a hombres y mujeres.	0	0	0	0	0
La discriminación contra las mujeres ya no es un problema en España.	0	0	0	0	0
Me gustaría ver a una mujer como presidenta de España.	0	0	0	0	0
Los hombres y las mujeres necesitan igual oportunidades laborales.	0	0	0	0	0
Las mujeres, por lo general, son buenas conductoras.	0	0	0	0	0
El trabajo de un hombre es más importante que el trabajo de una mujer.	0	0	0	0	0
Resulta sencillo comprender los puntos de vista que tienen los grupos de mujeres.	0	0	0	0	0
El movimiento feminista no sirve para nada.	0	0	0	0	0
Prefero tener de jefe a un hombre antes que a una mujer.	0	0	0	0	0
El Estado posee demasiados refugios en los temas que tienen las mujeres.	0	0	0	0	0
Las mujeres por lo general no tienen mucho talento.	0	0	0	0	0
Las mujeres son buenas con el razonamiento lógico.	0	0	0	0	0

The variables of interest were 15 items examining sexism (Ekehammar, Akrami, & Araya, 2000).<sup>1</sup> Example indicators included “I prefer a male boss to a female boss,” or “women are better suited to look after children and old people.” Each item was measured on a five-point agree/disagree scale and six items were reverse coded. Responses were averaged to create a sexism scale ( $\alpha = .83$ ),<sup>2</sup> on which higher scores indicated greater sexist attitudes.

## Analytic Strategy

To explore how the confidentiality statement affected univariate distributions of each question, we conducted cross tabulation of the responses by experimental condition. Chi-square tests were used to explore differences and Cramer’s  $V$  was calculated to analyze effect sizes. Then, we compared both the average scores for each indicator and the composite scale by experimental condition. In this case, analysis of covariance (ANCOVA) was used to explore differences and partial eta squared ( $\eta_p^2$ ) was calculated to analyze effect sizes. Age and gender were included as covariates since these two variables have been identified as correlates of sexism, with men and older respondents generally displaying more sexist attitudes (Hammond & Overall, 2017; Lameiras, Rodríguez, & González, 2004). We also explored differences between smartphone/tablet and computer responses to examine if the effect of confidentiality assurances was mediated by the device.

## Results

Table 1 displays the sociodemographic characteristics of the sample and the device used to complete the questionnaire by experimental condition. As can be seen, significant differences were only found in marital status ( $X^2 = 7.69$ ,  $df = 2$ ,  $p = .02$ ). Respondents’ age ranged from 18 to 86, with an average of 45.4 years ( $SD = 15.2$ ). Approximately half of the respondents identified themselves as women (51.0%), indicated being married (51.5%), and having college education (55.0%). The majority of the sample held Spanish nationality (95.0%) and lived in urban areas (78.0%). Regarding device used to respond to the questionnaire, PCs were the preferred device, chosen by approximately 6 in 10 respondents (57.8%), followed by smartphones (36.5%) and tablets (5.7%).

Table 2 shows response distributions and gender- and age-adjusted means for both the 15 items and the composite scale in the two experimental conditions. Results indicate that differences based on the experimental condition were only significant for one indicator (“society treats men and women the same way”) ( $F(1, 1003) = 3.92$ ,  $p = .048$ ). Response distributions and adjusted means for both the rest of the items and the composite scale were comparable in both conditions.

Screen sizes and question presentation vary widely between smartphones/tablets and computers, with smartphones and tablets presenting individual item rather than

<sup>1</sup>We received the approval from the authors to use their instrument on October 15, 2018. Two bilingual researchers independently translated the items. The two versions were compared, and differences were discussed until consensus was reached, with feedback from a third researcher not involved in the translation.

<sup>2</sup>Scale average scores were calculated based on 14 indicators because the results from an Explanatory Factor Analysis revealed that one item (“the school curriculum should be adapted to girls’ need”) did not load in the same factor as the other ones. The scale with the 14 items showed good internal consistency ( $\alpha = .83$ ), with factor loadings between .31 and .75 (Eigenvalue = 4.99).

Table 1.  
*Sample Composition by Experimental Condition*

	Overall ( <i>N</i> = 1,007) % ( <i>n</i> )	Confidentiality assurance ( <i>n</i> = 503) % ( <i>n</i> )	No assurance ( <i>n</i> = 504) % ( <i>n</i> )	<i>X</i> <sup>2</sup>	<i>p</i>
Gender					
Women	51.0 (514)	48.9 (246)	49.0 (247)	0.00	.97
Men	49.0 (493)	51.1 (257)	51.0 (257)		
Age group					
18–24	11.8 (119)	10.1 (51)	13.5 (68)	5.39	.37
25–34	15.3 (154)	15.9 (80)	14.9 (74)		
35–44	22.3 (225)	23.7 (119)	21.0 (106)		
45–54	20.3 (204)	20.5 (103)	20.0 (101)		
55–65	17.2 (173)	18.1 (91)	16.3 (82)		
66+	13.1 (132)	11.7 (59)	14.5 (73)		
Spanish nationality	95.0 (957)	95.0 (478)	95.0 (479)	0.00	.99
Marital status					
Never married	37.9 (382)	35.8 (180)	40.1 (202)	7.69	.02
Married	51.5 (519)	51.1 (257)	52.0 (262)		
Separated, divorced, widowed	10.5 (106)	13.1 (66)	7.9 (40)		
Education					
High school/technical school or less	45.0 (453)	42.7 (215)	47.2 (238)	2.04	.15
College graduate or more	55.0 (554)	57.3 (288)	52.8 (266)		
Habitat					
Town/farm	22.0 (221)	21.7 (109)	22.2 (112)	0.46	.80
Small/medium city	39.7 (400)	39.0 (196)	40.5 (204)		
Large city	38.3 (386)	39.4 (198)	37.3 (188)		
Device					
PC	57.8 (582)	59.2 (298)	56.4 (284)	4.30	.12
Smartphone	36.5 (368)	34.0 (171)	39.1 (197)		
Tablet	5.7 (57)	6.8 (34)	4.6 (23)		

<sup>a</sup> College graduate or more included the categories 'lower tertiary education, BA level' and 'higher tertiary education, MA level and above' from the International Standard Classification of Education (ISCED).

Table 2.  
*Response Distributions and Gender- and Age-Adjusted Means (Items and Scale) by Experimental Condition*

Indicator	Response	Condition		Differences ( $X^2/F$ )	<i>p</i> -values	Effect sizes ( $V/\eta^2$ )
		Confidentiality assurance ( $n = 503$ ) % ( $n$ )	No assurance ( $n = 504$ ) % ( $n$ )			
I prefer a male boss to a female boss	Strongly disagree	23.7 (119)	23.8 (120)	3.62	.460	.060
	Disagree	18.9 (95)	18.8 (95)			
	Neither agree nor disagree	48.9 (246)	50.8 (256)			
	Agree	6.4 (32)	5.8 (29)			
	Strongly agree	2.2 (11)	0.8 (4)			
A man's work is more important than a woman's	Mean ( <i>SE</i> ) <sup>a</sup>	2.45 (0.04)	2.41 (0.04)	0.36	.548	.000
	Strongly disagree	71.2 (358)	66.9 (337)	6.87	.143	.083
	Disagree	17.7 (89)	24.0 (121)			
	Neither agree nor disagree	9.5 (48)	8.1 (41)			
	Agree	0.8 (4)	0.6 (3)			
Women are generally not very talented	Strongly agree	0.8 (4)	0.4 (2)			
	Mean ( <i>SE</i> )	1.42 (0.03)	1.44 (0.03)	0.09	.769	.000
	Strongly disagree	71.8 (361)	73.0 (368)	4.11	.391	.064
	Disagree	19.5 (98)	17.9 (90)			
	Neither agree nor disagree	7.2 (36)	6.7 (34)			
I would like to see a woman president in Spain (R)	Agree	1.4 (7)	1.2 (6)			
	Strongly agree	0.2 (1)	1.2 (6)			
	Mean ( <i>SE</i> )	1.39 (0.03)	1.40 (0.03)	0.04	.843	.000
	Strongly disagree	1.2 (6)	1.0 (5)	1.94	.746	.044
	Disagree	1.4 (7)	1.2 (6)			
I would like to see a woman president in Spain (R)	Neither agree nor disagree	31.2 (157)	31.7 (160)			
	Agree	28.6 (144)	31.9 (161)			
	Strongly agree	37.6 (189)	34.1 (172)			
	Mean ( <i>SE</i> )	4.00 (0.04)	3.97 (0.04)	0.29	.593	.000

Continued

Table 2. *Continued*

Indicator	Response	Condition		Differences ( $X^2 / F$ )	p-values	Effect sizes ( $V/\eta^2$ )
		Confidentiality assurance (n = 503) % (n)	No assurance (n = 504) % (n)			
Women are good at logical reasoning (R)	Strongly disagree	1.8 (9)	1.4 (7)	1.34	.854	.037
	Disagree	2.8 (14)	2.0 (10)			
	Neither agree nor disagree	41.4 (208)	41.3 (208)			
	Agree	33.2 (167)	35.3 (178)			
	Strongly agree	20.9 (105)	20.0 (101)			
Woman are better suited to look after children and old people	Mean (SE)	3.69 (0.04)	3.71 (0.04)	0.14	.708	.000
	Strongly disagree	15.7 (79)	12.3 (62)	3.31	.507	.057
	Disagree	22.3 (112)	23.0 (116)			
	Neither agree nor disagree	39.0 (196)	38.3 (193)			
	Agree	18.5 (93)	21.0 (106)			
Women are generally good drivers (R)	Strongly agree	4.6 (23)	5.4 (27)			
	Mean (SE)	2.74 (0.05)	2.84 (0.05)	2.42	.120	.002
	Strongly disagree	1.8 (9)	1.0 (5)	3.26	.515	.057
	Disagree	7.4 (37)	8.9 (45)			
	Neither agree nor disagree	46.9 (236)	44.6 (225)			
Discrimination of women is no longer a problem in Spain	Agree	29.0 (146)	31.9 (161)			
	Strongly agree	14.9 (75)	13.5 (68)			
	Mean (SE)	3.48 (0.04)	3.48 (0.04)	0.00	.975	.000
	Strongly disagree	31.6 (159)	29.6 (149)	5.90	.207	.077
	Disagree	41.0 (206)	45.2 (228)			
Neither agree nor disagree	Agree	16.9 (85)	12.9 (65)			
	Strongly agree	7.8 (39)	7.9 (40)			
	Disagree	2.8 (14)	4.4 (22)			
	Mean (SE)	2.09 (0.05)	2.12 (0.05)	0.25	.618	.000
	Strongly disagree	21.9 (110)	23.4 (118)	1.09	.895	.033
Disagree	38.8 (195)	35.9 (181)				



Humiliating treatment of women in adverts is unusual	Neither agree nor disagree	25.0 (126)	25.8 (130)	
	Agree	10.7 (54)	10.7 (54)	
Society treats men and women the same way	Strongly agree	3.6 (18)	4.2 (21)	.887
	Mean ( <i>SE</i> )	2.35 (0.05)	2.36 (0.05)	.000
	Strongly disagree	33.8 (170)	26.2 (132)	.092
	Disagree	47.9 (241)	54.8 (276)	.074
	Neither agree nor disagree	12.1 (61)	11.5 (58)	
	Agree	5.2 (26)	5.8 (29)	
The government puts too much emphasis on women's issues	Strongly agree	1.0 (5)	1.8 (9)	
	Mean ( <i>SE</i> )	1.92 (0.04)	2.02 (0.04)	.004
	Strongly disagree	14.7 (74)	13.5 (68)	.601
	Disagree	28.4 (143)	26.0 (131)	.052
	Neither agree nor disagree	31.0 (156)	35.1 (177)	
	Agree	18.9 (95)	19.6 (99)	
The women's movement serves no purpose and should be abolished	Strongly agree	7.0 (35)	5.8 (29)	
	Mean ( <i>SE</i> )	2.75 (0.05)	2.78 (0.05)	.000
	Strongly disagree	26.8 (135)	25.0 (126)	.884
	Disagree	35.6 (179)	35.7 (180)	
	Neither agree nor disagree	25.2 (127)	27.0 (136)	
	Agree	8.0 (40)	7.1 (36)	
It is easy to understand the viewpoints of women's groups (R)	Strongly agree	4.4 (22)	5.2 (26)	
	Mean ( <i>SE</i> )	2.27 (0.05)	2.32 (0.05)	.000
	Strongly disagree	4.8 (24)	4.0 (20)	.514
	Disagree	16.3 (82)	14.9 (75)	.521
	Neither agree nor disagree	39.0 (196)	43.8 (221)	
	Agree	29.2 (147)	28.6 (144)	
	Strongly agree	10.7 (54)	8.7 (44)	
	Mean ( <i>SE</i> )	3.25 (0.04)	3.23 (0.04)	.000
				0.07
				.790
				.043
				3.22
				.514
				.521
				.000
				.057
				.000
				.790

Continued

Table 2. *Continued*

Indicator	Response	Condition		Differences ( $\chi^2 / F$ )	<i>p</i> -values	Effect sizes ( $V / \eta^2$ )
		Confidentiality assurance ( <i>n</i> = 503) % ( <i>n</i> )	No assurance ( <i>n</i> = 504) % ( <i>n</i> )			
The school curriculum should be adapted to girls' needs (R)	Strongly disagree	30.8 (155)	26.0 (131)	3.42	.489	.058
	Disagree	25.2 (127)	27.0 (136)			
	Neither agree nor disagree	31.8 (160)	34.5 (174)			
	Agree	7.4 (37)	8.3 (42)			
	Strongly agree	4.8 (24)	4.2 (21)			
Better measures should be taken to achieve equality in workplaces (R)	Mean ( <i>SE</i> )	2.30 (0.05)	2.38 (0.05)	1.20	.273	.001
	Strongly disagree	1.8 (9)	1.2 (6)	2.71	.608	.052
	Disagree	3.0 (15)	3.0 (15)			
	Neither agree nor disagree	10.5 (53)	10.1 (51)			
	Agree	32.8 (165)	37.3 (188)			
Scale average score (14 indicators) <sup>2</sup>	Strongly agree	51.9 (261)	48.4 (244)			
	Mean ( <i>SE</i> )	4.30 (0.04)	4.29 (0.04)	0.05	.817	.000
		2.35 (0.02)	2.36 (0.02)	0.14	.710	.000

*Note:* *R* = reverse-coded item; *SE* = standard error;  $\chi^2$  = chi-square; *F* = ANCOVA; *V* = Gramer's *V*;  $\eta^2$  = partial eta squared.

<sup>a</sup>Age- and gender-adjusted means.

grids. Therefore, differences in response were examined by device. The results are consistent with the aggregated findings, suggesting no differences between the conditions in either device (smartphone and tablets vs. computers) (a table reporting these results is available upon request).

## Discussion

The aim of this research was to analyze the impact of including a confidentiality statement on responses to a series of questions assessing sexism in Spain. This statement highlighted the confidentiality and acceptability of the answers and was placed immediately before the items. The results show that this type of confidentiality assurance did not affect participants' responses about sexism in a self-administered web survey conducted in Spain. Contrary to the results found by previous research (Singer et al., 1995), the findings of this study indicate that these kinds of messages may not be very effective in reducing socially desirable responding in all situations. It was expected that confidentiality assurances would result in more sexist responses under the assumption that this type of messages might alleviate respondent concerns that the data could end up in the wrong hands (Tourangeau & Yan, 2007). Multiple reasons might contribute to explain these null findings. The first one refers to the data collection mode. As previous research has shown, self-administered web surveys (such as the one used in this study) generally provide less biased responses when compared with other modes in which interviewers are present during the administration of the questionnaires (Kreuter et al., 2008; Tourangeau & Yan, 2007). This could explain the differences between Singer and colleagues' study (1995) and the present one, since studies included in their meta-analysis did not involve online surveys. The operationalization of the confidentiality assurance in nonspecific terms and the fact that all respondents were exposed to a message at the beginning of the questionnaire advising them about the sensitivity and the confidentiality of the survey responses (see Appendix) might contribute to explain these null results as well. Another consideration is that respondents might not have paid much attention to the instructions given, which is especially true in self-administered web surveys, in which no interviewers are present (Vésteinsdóttir et al., 2019). In this regard, respondents might have skipped or skimmed the instructions (Brosnan, Babakhani, & Dolnicar, 2019). This could have been heightened because the sample was composed of panellists (Shamon & Berning, 2020). The location of the experiment within the questionnaire might also be relevant. In the current study, the confidentiality assurance was located at the beginning of the questionnaire and respondents might have been less sensitive to the effect of the experiment. Because the location of the experiment was not manipulated, it may be valuable to compare the effects of this manipulation at different stages of the questionnaire.

The results presented here show that confidentiality assurances might not be very effective in reducing socially desirable responding when asking about sexism. For this reason, it is recommended that future research examines the effects of other strategies for reducing socially desirable responding in sensitive questions, such as encouraging respondents to answer honestly rather than reminding them of the confidential nature of their responses (Vésteinsdóttir et al., 2019), and trying to enhance respondents' motivation to answer carefully and provide more accurate responses (Revilla, 2015; Tourangeau, Smith, & Rasinski, 1997). As Tourangeau et al. (1997) indicated, increasing the participants' motivation to respond accurately might have the potential to substantially improve their reporting.

Despite its contribution, the current study has limitations. The first one is the topic addressed. Sexism is a particularly highly charged topic which respondents may hold strong opinions on. Whether the results would remain stable for different topics with different levels of sensitivity needs to be examined. Because of the topic under study, validation data were not available, and we assumed that higher reports of socially undesirable responses (i.e., sexism) represented more honest responses. Although this approach is commonly adopted and the assumption behind it is often plausible, it is still an assumption (Tourangeau & Yan, 2007). Furthermore, the sample was a non-probabilistic one, and the generalizability of our findings is not assured. Panels have been shown to overrepresent individuals with certain demographic characteristics (e.g., white, college educated) and personality traits (e.g., openness, conscientiousness) (Unangst et al., 2020). The risk of self-selection biases is particularly relevant in this experiment, where certain characteristics associated with participating in a survey panel (e.g., less concerns about data protection) might be linked to the outcome of the experiment (i.e., disclosure of socially undesirable attitudes). Further research is needed to determine the robustness of these findings and the extent to which probabilistic samples produce equivalent results. Finally, another limitation that should be noted is that the study did not include other variables that may be related to socially desirable responding, such as the place from which the respondents answered the questionnaire (e.g., in a public place, at home, or in the workplace) or the presence of third parties (Aquilino, Wright, & Supple, 2000; Tourangeau & Yan, 2007).

## Conclusions

Surveys are the main source of data for many socially undesirable behaviors. Therefore, honest responding is paramount. The main objective of our study was to examine the effect of including a confidentiality assurance placed before a set of questions measuring sexism in Spain. Our null findings do not support the use of such strategy to reduce social desirability, at least not in studies such as ours using panellists and providing non-specific messages. Our results, in concert with those found in previous research, suggest considering new strategies for collecting accurate information on sensitive topics.

## Funding

This study was financed by the Department of Public and Business Law at the University of Castilla-La Mancha.

*Conflicts of interest:* The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this research note.

## Appendix

### *Screenshot of the sensitive screen*

Figure A1.

*Sensitive screen presented to the respondents immediately before starting the questionnaire.*



Below is the translation into English of the content:

#### **Intimate or personal questions:**

- This survey contains some questions of an intimate or personal nature.
- You can be sure that your responses are anonymous, and that the only purpose of this study is to compile statistics by gathering the opinions of all the participants.
- If you find any of the questions unsettling, we apologise in advance and remind you that you can leave the survey at any time.

#### **References**

- Andreenkova, A. V., & Javenile, D. (2019). Sensitive questions in comparative surveys. In T. P. Johnson, B. Pennell, I. A. L. Stoop, & B. Dorer (Eds.), *Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural Contexts (3MC)* (pp. 139–160). Hoboken: Wiley. .
- Aquilino, W. S., Wright, D. L., & Supple, A. J. (2000). Response effects due to bystander presence in CASI and paper-and-pencil surveys of drug use and alcohol use. *Substance Use and Misuse*, 35, 845–867.
- Barton, A. H. (1958). Asking the embarrassing question. *Public Opinion Quarterly*, 22, 67–68.

- Brosnan, K., Babakhani, N., & Dolnicar, S. (2019). "I know what you're going to ask me" Why respondents don't read survey questions. *International Journal of Market Research*, 61, 366–379.
- Diekmann, A. (2003). *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen*. Reinbek: Rowohlt Taschenbuch.
- Ekehammar, B., Akrami, N., & Araya, T. (2000). Development and validation of Swedish classical and modern sexism scales. *Scandinavian Journal of Psychology*, 41, 307–314.
- Glynn, A. N. (2013). What can we learn with statistical truth serum? Design and analysis of the list experiment. *Public Opinion Quarterly*, 77, 159–72.
- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey methodology*. Hoboken: Wiley.
- Hammond, M. D., & Overall, N. C. (2017). Dynamics within intimate relationships and the causes, consequences, and functions of sexist attitudes. *Current Directions in Psychological Science*, 26, 120–125.
- Holbrook, A. L., & Krosnick, J. A. (2010). Social desirability bias in voter turnout reports: Tests using the item count technique. *Public Opinion Quarterly*, 74, 37–67.
- Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone versus face-to-face interviewing of national probability samples with long questionnaires - Comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly*, 67, 79–125.
- Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and Web surveys: The effects of mode and question sensitivity. *Public Opinion Quarterly*, 72, 1–13.
- Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality and Quantity*, 47, 2025–2047.
- Kuha, J., & Jackson, J. (2014). The item count method for sensitive survey questions: Modelling criminal behaviour. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 63, 321–341.
- Lameiras, M., Rodríguez, Y., & González, L. (2004). Evolution of hostile sexism and benevolent sexism in a Spanish sample. *Social Indicators Research*, 66, 197–211.
- Lax, J. R., Phillips, J. H., & Stollwerk, A. F. (2016). Are survey respondents lying about their support for same-sex marriage? Lessons from a list experiment. *Public Opinion Quarterly*, 80, 510–533.
- Miller, J. D. (1984). *A new survey technique for studying deviant behavior* (PhD thesis, The George Washington University).
- Mneimneh, Z., Elliott, M. R., Tourangeau, R., & Heeringa, S. G. (2018). Cultural and interviewer effects on interview privacy: Individualism and national wealth. *Cross-Cultural Research*, 52, 1–28.
- Näher, A., & Krumpal, I. (2012). Asking sensitive questions: The impact of forgiving wording and question context on social desirability bias. *Quality and Quantity*, 46, 1601–1616.
- Revilla, M. (2015). Impact of raising awareness of respondents on the measurement quality in a web survey. *Quality & Quantity*, 50, 1469.
- Shamon, H. & Berning, C. C. (2020). Attention check items and instructions in online surveys with incentivized and non-incentivized samples: Boon or bane for data quality? *Survey Research Methods*, 14, 55–77.

- Singer, E., Hippler, H., & Schwarz, N. (1992). Confidentiality assurances in surveys: Reassurance or threat. *International Journal of Public Opinion Research*, 4, 256–268.
- Singer, E., Von Thurn, D., & Miller, E. (1995). Confidentiality assurances and response: A quantitative review of the experimental literature. *Public Opinion Quarterly*, 59, 66–77.
- Tourangeau, R. (2017). Confidentiality, privacy, and anonymity. In D. L. Vannette & J. A. Krosnick (Eds.), *The Palgrave Handbook of Survey Research* (pp. 501–507). Switzerland: Palgrave Macmillan.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (Eds.) (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.
- Tourangeau, R., Smith, T. W., & Rasinski, K. (1997). Motivation to report sensitive behaviors in surveys: Evidence from a bogus pipeline experiment. *Journal of Applied Social Psychology*, 27, 209–222.
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133, 859–883.
- Turner, C. F., Villarroel, M. A., Rogers, S. M., Eggleston, E., Ganapathi, L., Roman, A. M., & Al-Tayyib, A. (2005). Reducing bias in telephone survey estimates of the prevalence of drug use: A randomized trial of telephone audio-CASI. *Addiction*, 100, 1432–1444.
- Unangst, J., Amaya, A. E., Sanders, H. L., Howard J., Ferrell A., Karon, S., & Deber, J. A. (2020). A process for decomposing total survey error in probability and nonprobability surveys: A case study comparing health statistics in US internet panels. *Journal of Survey Statistics and Methodology*, 8, 62–88.
- Vésteinsdóttir, V., Joinson, A., Reips, U., Danielsdóttir, H. B., Thorarinsdóttir, E. A., & Thorsdóttir, F. (2019). Questions on honest responding. *Behavior Research Methods*, 51, 811–825.
- Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60, 63–69.

### Biographical Notes

**Carmen M. Leon**, PhD, is an Assistant Professor at the University of Castilla-La Mancha (Spain). Her current research interests include intimate partner violence against women, with particular emphasis on public perceptions and attitudes, sexism, and survey research methods.

**Eva Aizpurua**, PhD, is a Research Fellow at the European Social Survey Headquarters, and a Visiting Research Fellow at Trinity College Dublin. She earned her PhD in Criminology from the University of Castilla-La Mancha (Spain) and worked as a postdoctoral scholar at the University of Northern Iowa (USA, 2016–2018). Eva is currently serving on the editorial boards of *Criminal Justice and Behavior* and *Survey Practice*. Her main research interests are questionnaire design and pretesting in the context of cross-national surveys and data harmonization.

**Sophie van der Valk**, PhD candidate on the European-wide Project “Prisons: The Rule of Law, Accountability and Rights” (PRILA), School of Law, Trinity College Dublin. Her main research interests include human rights, prison law, socio-legal research methods and accountability mechanisms.