Field Methods
1-18

DOI: 10.1177/1525822X211012259
journals.sagepub.com/home/fmx
SAGE

Agree or Disagree: Does It Matter Which Comes First? An Examination of Scale Direction Effects in a Multi-device Online Survey

Carmen M. Leon[1], Eva Aizpurua[2,3], and Sophie van der Valk[3]

Abstract

Previous research shows that the direction of rating scales can influence participants' response behavior. Studies also suggest that the device used to complete online surveys might affect the susceptibility to these effects due to the different question layouts (e.g., horizontal grids vs. vertical individual questions). This article contributes to previous research by examining scale direction effects in an online multi-device survey conducted with panelists in Spain. In this experiment, respondents were randomly assigned to two groups where the scale direction was manipulated (incremental vs. decremental). Respondents completed the questionnaire using the device of their

[1] University of Castilla-La Mancha, Albacete, Spain
[2] ESS ERIC HQ, City, University of London, United Kingdom
[3] Trinity College Dublin, Ireland

Corresponding Author:
Carmen M. Leon, University of Castilla-La Mancha, 1, Plaza de la Universidad, 02071 Albacete, Spain.
Email: carmen.leon@uclm.es

choosing (57.8% used PCs; 36.5% used smartphones; and 5.7% used tablets). The results show that scale direction influenced response distributions but did not significantly affect data quality. In addition, our findings indicate that scale direction effects were comparable across devices. Findings are discussed and implications are highlighted.

Survey research has proved that many design features of scales affect how respondents process and use them to construct their responses (DeCastellarnau 2018; Menold and Bogner 2016; Yan and Keusch 2015). For instance, it is well established that the presentation order of categorical response options influences survey responses. This type of response bias, known as response-order effects, distinguishes between primacy and recency effects (Chang and Krosnick 2009; Galesic et al. 2008; Stern et al. 2007). Primacy effects refer to higher endorsements of response categories presented early in the list, while recency effects refer to higher endorsements of response categories presented later in the list (Schwarz and Hippler 2004).

Although previous research has widely documented the impact on survey responses produced by varying the order of categorical response options, fewer studies have examined response-order effects with ordinal scales, despite their extensive use in survey research. The first studies that analyzed these effects date back to the 1960s, when Belson (1966) found that survey responses tended to shift toward the starting point of the rating scales, regardless of scale length and respondent characteristics. More recent studies, however, have turned up mixed evidence. Some have reported a similar tendency for responses to be biased toward the starting point of scales (Garbarski et al. 2018; Höhne and Krebs 2017; Israel 2006; Toepoel et al. 2009; Yan and Keusch 2018), while others have found no effect of the direction of rating scales on survey responses (Krebs and Hoffmeyer-Zlotnik 2010; Rammstedt and Krebs 2007; Weng and Cheng 2000). In addition, some studies have reported scale direction effects on a limited number of questions and no effect on other questions within the same experiment (Christian et al. 2009; Elway 2013; Höhne and Krebs 2017).

Theoretical Explanations

Although several theoretical approaches have been used to explain response-order effects, the exact mechanism responsible for scale direction

effects has not been elucidated yet (Höhne and Krebs 2017; Keusch and Yan 2018; Salzberger and Koller 2019). Currently, there are three main approaches explaining how response-order effects occur: cognitive elaboration, survey satisficing, and interpretive heuristics.

The cognitive elaboration model (Schwarz et al. 1992) argues that response-order effects are caused by an interaction between the serial position of the response category (e.g., beginning, middle, end) and the presentation mode (i.e., visual or aural). Based on this, it is stated that recency effects emerge in visual formats (e.g., self-administered web questionnaires) while primacy effects are observed in aural formats (e.g., telephone surveys). This approach has, however, been criticized for overlooking the complexity of response categories and the cognitive ability and motivation of respondents (Bishop and Smith 2001).

A second approach is the satisficing theory (Krosnick 1991), according to which some respondents will not spend the necessary effort to provide optimal responses, taking shortcuts that will degrade the integrity of their responses to various degrees. This theory posits that the likelihood of satisficing depends on the difficulty of the task as well as the ability and motivation of the respondent. Krosnick (1991) discusses several forms of satisficing, including selecting the first response option that seems reasonable and not differentiating among response options for items using the same response scale.

More recently, research has emphasized the importance of the interpretive heuristics used by respondents while answering questionnaires (e.g., the assumption that the middle category represents the average or typical response). Specifically, studies have examined the anchoring and adjustment bias (Höhne and Lenzner 2015; Yan and Keush 2015) and the spatial proximity heuristic (Salzberger and Koller 2013, 2019).

Correlates of Scale Direction Effects

Recent studies suggest that scale direction effects may be moderated by both survey and personal variables. Related to the survey itself, research shows that the visual design of the scales and the devices used by the respondents to complete the questionnaire could moderate the scale direction effects. Regarding the visual design, studies have shown that, in self-administered surveys, response-order effects are more likely to occur in vertical than in horizontal arrangements of the response categories (Christian et al. 2009; Höhne and Lenzner 2015; Menold and Bogner 2014). For instance, Christian and colleagues (2009), in two web surveys,

found that scale direction had a significant impact only on questions that listed options vertically rather than horizontally. Höhne and Lenzner (2015) provided evidence of this interaction effect in an eye-tracking study with attitudinal questions. On the other hand, Höhne and Krebs's (2017) study using paper questionnaires revealed response-order effects with agree–disagree grids but not with item-specific questions. As far as the devices used to complete the survey are concerned, the experiment conducted by Stapleton (2013) pointed to an interaction effect between the scale order and the device, with smartphone respondents being more prone to scale direction effects. More recently, Krebs and Höhne (2019), in a study conducted using a German opt-in panel, found that both PCs and smartphones were robust against scale direction effects.

Scale Direction and Its Impact on Data Quality

In addition to the finding that respondents tend to select responses located at the beginning of the scales, it has been suggested that the direction of rating scales could affect response quality. Regarding completion times, Christian and colleagues (2009) found that response times were longer when the negative end of the scale was presented first in two web surveys conducted among undergraduate students. However, Garbarski et al. (2018) found no significant differences in response times across experimental factors (i.e., response option order and scale orientation) in an online multi-device experiment. In terms of acquiescence and extreme response bias, Liu and Keusch (2017), in a mixed-mode study using face-to-face and web surveys, found that acquiescence and extreme response styles existed in both scale directions and survey modes. In addition, Yan and Keusch (2018) did not find evidence that scale direction affects acquiescence, mid-point response style, straight-lining, and internal consistency in four surveys conducted face-to-face and online. Similarly, Keusch and Yan (2018) noted that scale direction did not influence reliability, measured as test–retest, in longitudinal settings in their experiment embedded in two web surveys.

Present Study

As highlighted by previous research examining the effects of scale direction, findings across the studies are mixed and inconclusive. Moreover, studies suggest that these effects may be affected by both survey and personal variables, such as the device used to complete the surveys (Stapleton 2013), the visual design of the scales (Höhne and Lenzner 2015; Liu and

Cernat 2018), and by linguistic and cultural norms (Yan and Hu 2018; Yan and Keusch 2015). To contribute to this field of research, this study examines scale direction effects on response distributions and average scores of 24 items measuring myths about intimate partner violence (10 items) and sexism (14 items). In addition, it analyzes the relationship between scale direction and data quality, measured using two indicators: nondifferentiation and internal consistency. Drawing on previous research, we anticipate that:

H_1: Respondents will be more likely to select the starting point of the rating scales (i.e., strongly agree and strongly disagree).

H_2: Scales presented in decremental order will produce higher average scores.

H_3: Scale direction will affect response distributions but no other indicators of data quality.

H_4: Scale direction effects will be strongest among smartphones/tablets respondents who were presented with vertical individual questions than PC respondents who received horizontal grid (matrix) questions with a common scale.

H_5: Regardless of scale directionality, we expect internal consistency to be higher and differentiation to be lower among PC respondents.

This last hypothesis is based on Höhne and Krebs's (2017) reasoning that questions on grids may promote the same answering task as identical response scales apply to all items, which can dismay respondents and discourage them from expending great effort when responding.

Materials and Methods

Data Collection

The online survey was conducted between January 7 and January 29, 2019, using the Netquest panel in Spain.[1] Quotas for age, gender, and habitat were used to obtain a sample distribution similar to the Spanish population. A total of 1,007 surveys were completed using PCs (57.8%), smartphones (36.5%), and tablets (5.7%). The questionnaire was completed by 92.3% of those panelists invited, who received incentives in the form of points for participating. The surveys took approximately 12 minutes to complete ($M = 11.72$, SD = 6.32). The questionnaire was pretested between October

10 and October 31, 2018, using expert reviews. The survey was administered in Spanish and included 51 questions about perceptions and attitudes toward intimate partner violence. The wording of the questions (translated into English) used in the analysis can be found as Supplemental Material.

Experimental Design

The study used a one-factor (incremental vs. decremental) between-subjects design. Respondents either randomly received the scales in an incremental order, from "strongly disagree" to "strongly agree" ($n = 503$), or in a decremental one, from "strongly agree" to "strongly disagree" ($n = 504$). Assignment to the condition was linked and respondents assigned to one condition received all the rating scales in the same order. The screenshots of the two conditions for smartphone/tablet and PC responses are displayed in Supplemental Material.

Variables of Interest

Myths about intimate partner violence. The first variable of interest was a 10-item scale developed ad hoc to examine gender myths and beliefs about intimate partner violence (e.g., "a large number of complaints alleging intimate partner violence against women are false," "in couples, women perpetrate as much violence as men"). Each item was measured on a five-point agree/disagree scale. Responses were averaged to produce a scale, on which higher scores indicated greater myth endorsement.

Classical and modern sexism scale. This scale was developed and validated in the Scandinavian context by Ekehammar et al. (2000).[2] Example items included "I prefer a male boss to a female," or "women are better suited to look after children and old people." Although the original scale is composed of 15 items, in the present study one indicator was removed because the results from an Exploratory Factor Analysis revealed that it did not load in the same factor as the others. Each item was measured on a five-point agree/disagree scale and six items were reverse-coded. All the responses were averaged to create a sexism scale, on which higher scores indicated greater sexist attitudes.

Response distributions and data quality indicators. Response distributions and average scores in the outcome variables were evaluated. In addition, non-

differentiation and internal consistency were the two indicators used to assess data quality. Information about the operationalization of these indicators is provided in Table 1.

Table 1. Response Distributions and Indicators of Data Quality.

Indicator	Definition
Average scores	Mean score for each scale (myths and sexism)
Response distributions	Number of occasions that respondents select each response option (e.g., strongly disagree) within the scales
Non-differentiation	Respondent-level standard deviation
Internal consistency	Consistency (Cronbach α) of responses across all items that composed each of two scales

Survey and respondent characteristics. Both survey and personal variables were included in the analyses. These involved gender (men, women), age (in ranges), urbanicity (town/farm, small/medium city, large city), device (PCs, smartphones, tablets), and experimental condition (incremental, decremental).

Analytic Strategy

First, we describe the characteristics of the sample. Because the device used to complete the questionnaire was not randomly assigned to respondents, we compare the composition of the sample by device (PCs vs. smartphones/tablets).[3] Chi-square and *t*-tests were used to explore differences and effect sizes were calculated (Cramer's *V* and Cohen's *d*, respectively). To explore whether the directionality of the scales affected both average scores and standard deviations and whether there was an interaction between the experimental condition and the device, separate ordinary least squares (OLS) regression models were estimated controlling by a series of variables. Cronbach's α coefficients were also compared between the groups using the "cocron" package in R (Diedenhofen 2016). Finally, to analyze the effect of scale direction on the number of occasions that respondents select each response option within the scales (i.e., count variables) and whether there was an interaction between experimental condition and device, negative binomial regressions were estimated.

Findings

Sample Composition

Table 2 displays the sociodemographic characteristics of the sample by device. Most of the sample indicated being married (51.5%), having college education (55.0%), and holding Spanish nationality (95.0%). Significant differences between PC and smartphone/tablet respondents were found in terms of gender, age, and urbanicity, although effect sizes were small in all cases. Women more often completed the survey using smartphones/tablets than men (56.5% vs. 47.1%). In addition, more smartphone/tablet respondents reported living in large cities than PC respondents (42.6% vs. 37.6%). Finally, those who used smartphones and tablets were approximately four years younger than those who used PCs ($M = 43.03$, SE $= 0.70$ vs. $M = 47.08$, SE $= 0.64$; $t = -4.22$, $df = 1{,}005$, $p \leq .001$).

Scale Direction Effects on Response Distributions and Data Quality

To examine whether scale directionality influences average and non-differentiation scores (*hypotheses 2* and *3*, respectively), we estimated OLS regression models controlling for those variables in which differences between devices were found (i.e., gender, age, and urbanicity), as well as the experimental condition and device used to complete the questionnaire. As shown in Table 3, for both scales average scores were higher among those who received the scales in decremental order ($b_{myths} = 0.07$, $b_{sexism} = 0.05$), although both coefficients were marginally significant ($p_{myths} = .057$, $p_{sexism} = .057$). To examine whether the relationship between scale directionality and scales' average scores was moderated by the device, two additional models were estimated including the main effects of device and experimental condition and their interaction. The interaction term was non-significant in both models (myths: $b = 0.07$, $p = .360$; sexism: $b = 0.03$, $p = .600$), indicating that the association between scale direction and scales' scores was comparable across devices (*hypothesis 4*).

In terms of non-differentiation, our results show that differentiation in the myths scale was lower among PC respondents—who received grids instead of individual items ($b = -0.00$, $p = .017$) (*hypothesis 5*). This difference, however, was not found in the sexism scale. For this scale, differentiation was marginally lower among respondents who received the scales in decremental order ($b = -0.03$, $p = .084$) (*hypothesis 3*). Once again, the results from the model incorporating an interaction term between the experimental condition and the device indicate that the effect of scale

Table 2. Sample Composition by Device.

	Overall ($N = 1{,}007$) % (n)	Smartphone/ Tablet ($n = 425$) % (n)	PC ($n = 582$) % (n)	X^2	Cramer's V
Gender					
Women	51.0% (514)	56.5% (240)	47.1% (274)	8.67**	−0.09
Men	49.0% (493)	43.5% (185)	52.9% (308)		
Age group					
18–24	11.8% (119)	13.9% (59)	10.3% (60)	23.83***	0.15
25–34	15.3% (154)	15.8% (67)	15.0% (87)		
35–44	22.3% (225)	26.8% (114)	19.1% (111)		
45–54	20.3% (204)	20.7% (88)	19.9% (116)		
55–65	17.2% (173)	13.9% (59)	19.6% (114)		
66+	13.1% (132)	8.9% (38)	16.2% (94)		
Spanish nationality	95.0% (957)	95.8% (407)	94.5% (550)	0.83	—
Marital status					
Never married	37.9% (382)	40.5% (172)	36.1% (210)	2.21	—
Married	51.5% (519)	48.9% (208)	53.4% (311)		
Separated, divorced, widowed	10.5% (106)	10.6% (45)	10.5% (61)		
Education					
High school/ technical school or less	45.0% (453)	43.1% (183)	46.4% (270)	1.10	—
College graduate or more[a]	55.0% (554)	56.9% (242)	53.6% (312)		
Urbanicity					
Town/farm	22.0% (221)	24.0% (102)	20.5% (119)	7.58*	0.09
Small/medium city	38.3% (386)	33.4% (142)	41.9% (244)		
Large city	39.7% (400)	42.6% (181)	37.6% (219)		
Condition					
Incremental	49.9% (503)	49.7% (211)	50.2% (292)	0.03	—
Decremental	50.1% (504)	50.3% (214)	49.8% (290)		

[a]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).

*$p \leq .05$; **$p \leq .01$; ***$p \leq .001$.

directionality on non-differentiation was comparable for smartphone/tablet and PC respondents (myths: $b = 0.05$, $p = .281$; sexism: $b = 0.07$, $p = .062$). Internal consistency was comparable between experimental conditions (*hypothesis 3*). In the myths scale, Cronbach's α was .77 in the incremental group and .74 in the decremental group ($X^2 = 2.42$, $p = .112$). Similarly, the internal consistency of the sexism scale was .83 in the incremental order and .82 in the decremental one ($X^2 = 0.52$, $p = .470$).

Table 3. Predictors of Average Scores and Standard Deviations by Scale (OLS Regression Models).

	Myths about IPVAW				Sexism			
	Average Score		Non-differentiation		Average Score		Non-differentiation	
Variables	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*
Condition (Ref. Incremental)	0.07†	0.04	0.02	0.02	0.05†	0.03	−0.03†	0.02
Device (Ref. Smartphone/tablet)	−0.04	0.04	−0.00*	0.00	0.02	0.03	−0.02	0.02
Gender (Ref. Women)	0.32***	0.04	−0.03	0.02	0.29***	0.03	−0.07***	0.02
Age	0.00	0.00	−0.00	0.00	0.00***	0.00	0.00	0.00
Urbanicity (Ref. Town/farm)								
Small/medium city	−0.02	0.05	−0.01	0.03	−0.03	0.04	0.01	0.02
Large city	0.03	0.05	−0.03	0.03	0.02	0.04	−0.01	0.02

Note: b = non-standardized coefficients, SE = standard errors.
†$p \leq .10$; *$p \leq .05$; ***$p \leq .001$.

Finally, negative binominal regression models were estimated to explore whether scale directionality affects response distributions (*hypothesis 1*) while controlling for a number of variables. Table 4 shows the results from these models, including the incidence rate ratios (IRR) and associated standard errors. For both scales, the number of strongly disagree responses was lower when participants received the scales in decremental order (myths: $IRR = 0.88$, $p = .039$; sexism: $IRR = 0.84$, $p = .001$). No differences, however, were found on the other extreme of the scale (strongly agree responses). Only for the sexism scale was the number of disagree and neutral responses marginally higher when scales were presented in

Table 4. Predictors of Response Distributions by Scale (Negative Binominal Regression Models).

Response Categories	Variables	Myths about IPVAW		Sexism	
		IRR	*SE*	*IRR*	*SE*
Strongly agree	Condition (Ref. Incremental)	1.02	0.09	0.99	0.06
	Device (Ref. Smartphone/tablet)	0.79**	0.07	0.85**	0.06
	Gender (Ref. Women)	1.26**	0.11	0.70***	0.05
	Age	1.00	0.00	1.00	0.00
	Urbanicity (Ref. Town/farm)				
	Small/medium city	0.94	0.11	1.09	0.09
	Large city	0.92	0.11	0.94	0.08
Agree	Condition (Ref. Incremental)	1.09	0.07	1.02	0.05
	Device (Ref. Smartphone/tablet)	0.97	0.06	1.07	0.05
	Gender (Ref. Women)	1.28***	0.08	1.22***	0.06
	Age	1.01†	0.00	1.00**	0.00
	Urbanicity (Ref. Town/farm)				
	Small/medium city	1.08	0.09	1.07	0.07
	Large city	1.10	0.09	1.06	0.07
Neither agree nor disagree	Condition (Ref. Incremental)	1.04	0.06	1.08†	0.05
	Device (Ref. Smartphone/tablet)	1.12	0.07	1.04	0.05
	Gender (Ref. Women)	1.43***	0.08	1.38***	0.06
	Age	1.00	0.00	1.00	0.00
	Urbanicity (Ref. Town/farm)				
	Small/medium city	0.93	0.07	0.88*	0.05
	Large city	0.97	0.07	0.99	0.06
Disagree	Condition (Ref. Incremental)	1.04	0.05	1.08†	0.05
	Device (Ref. Smartphone/tablet)	0.99	0.05	0.96	0.04
	Gender (Ref. Women)	0.97	0.04	1.07	0.05
	Age	1.00	0.00	1.00*	0.00
	Urbanicity (Ref. Town/farm)				
	Small/medium city	0.98	0.06	1.05	0.06
	Large city	1.06	0.06	1.06	0.06
Strongly disagree	Condition (Ref. Incremental)	0.88*	0.05	0.84***	0.04
	Device (Ref. Smartphone/tablet)	1.01	0.06	1.04	0.06
	Gender (Ref. Women)	0.62***	0.04	0.65***	0.03
	Age	1.00	0.00	0.99***	0.00
	Urbanicity (Ref. Town/farm)				
	Small/medium city	1.03	0.08	0.99	0.07
	Large city	0.91	0.07	0.93	0.06

Note: IRR = incidence rate ratios, *SE* = standard errors.
†$p \leq .10$; *$p \leq .05$; **$p \leq .01$; ***$p \leq .001$.

decremental order ($IRR = 1.08$, $p = .087$ and $IRR = 1.08$, $p = .072$, respectively). Additionally, the number of strongly agree answers was lower for PC respondents in both scales ($IRR = 0.79$, $p = .007$ and $IRR = 0.85$, $p \leq .001$, respectively). None of the interaction terms between scale directionality and device were significant in the models.

Discussion

The aim of this study was to examine the effects of scale directionality on response distributions, average scores, and two indicators of data quality. To our knowledge, this is one of the few studies that examines scale direction effects in an online multi-device survey outside of Germany and the United States. We do so by examining responses to 24 items grouped into two scales while controlling for a series of respondent characteristics and the device used to complete the questionnaire. We believe this study can be useful for survey practitioners by providing further evidence of the impact of scale characteristics.

"If respondents are carefully considering the item and reporting their true attitudes, then the ordering of the response choices should have no impact on the options that are selected" (Malhotra 2008:916). However, as previous research has shown, respondents do engage in satisficing behaviors that lead to suboptimal responses (Chang and Krosnick 2009; Galesic et al. 2008; Stern et al. 2007). Findings from this research show a tendency to this behavior, which partially confirms our first and second hypotheses. Specifically, we found that respondents were less likely to select strongly disagree responses when scales were presented in decremental order (i.e., starting from strongly agree). These results are consistent with previous research showing that participants tend to be biased toward the starting point of the rating scale (Smyth et al. 2019; Yan and Keusch 2018).

Although differences were found for disagree responses, there was no evidence that respondents selected strongly agree more often when the scale was presented in decremental order. These results are similar to those of Yan and Keusch (2018), who found higher levels of extreme response style when scales started with the agree, rather than the disagree side. They also concur with those of Smyth and colleagues (2019), with findings from those studies suggesting that scale direction effects may be attributed to heuristic processing and satisficing. In the current study, respondents selected disagree responses marginally more often in the sexism scale when the rating scale was presented in decremental order, suggesting that at least some respondents who disagreed were registering the first negative

response they encountered. However, it is important to highlight that we did expect respondents to be prone to disagree with the items since they represented sexist views and myths about intimate partner violence (there were some exceptions with reverse-coded items). This fact could also help explain our results.

Hypothesis 2 was partially supported by the data, with scales presented in decremental order having marginally higher average scores than those listed in incremental order. The fact that the average scores were different depending on the direction of the scale is an important finding because, in applied settings, scales like those included in this study are often used to measure constructs that are not directly observable (e.g., fear, trust, sexism) and the way in which these scales are often built is by computing their averages or sums. If the scores vary because of methodological decisions rather than differences in the construct of interest, this finding is important for applied researchers, who need to carefully assess the direction of their scales to avoid inadvertently influencing research findings as much as possible.

Partial support was found for *hypothesis 5*, with respondents differentiating their responses less in one of the two scales when they used PCs—and received grids—rather than smartphones/tablets. The fact that the interaction between scale directionality and device was not significant indicates that respondents differentiated more when the questions were presented vertically one by one than horizontally in the form of a grid, regardless of the scale direction. These findings are consistent with previous research (Mavletova et al. 2018) and suggest that differences in participants' responses are not only attributed to scale direction but also to the layout in which response options are presented within the different devices used to complete the questionnaire. Another finding supporting this assumption is that respondents who used PCs were less prone to select strongly agree responses in both myths and sexism scales, after controlling for key sociodemographic characteristics and scale direction. Regarding the sexism scale, our results show that respondents differentiated marginally less when they received the scale in decremental order. Overall, these findings suggest that scale direction have some impact on response distributions (specifically on negative responses), but the impact on data quality is limited (we did not find that scale direction affected internal consistency), providing partial support for *hypothesis 3*. This is consistent with the findings of Yan and Keusch (2018), who did not find that scale direction affected acquiescence, mid-point response style, straight-lining, and reliability in four surveys conducted using face-to-face and online modes in the United States.

In addition, findings from this research point to differences between PC and smartphone/tablet respondents on key sociodemographic characteristics, including age, gender, and urbanicity. This finding highlights the importance of controlling for differential self-selection effects when multiple devices are used to collect the data and the assignment is not randomized.

Further consideration is needed as to the effect scale direction could have for cross-cultural research, especially in light of recent research suggesting that cultural norms can impact on respondents' choice of answers (Yan and Hu 2018; Yan and Keusch 2015). It has been found that Hispanic and Mediterranean respondents, among which we place Spanish respondents, are more prone to extreme response styles than some other groups (Yan and Hu 2018). While the current study was conducted in Spanish and the findings are consistent with other research conducted in Western countries, this may not be the case in other languages and cultures since response patterns may differ systematically across groups.

Despite its contribution, the current study has important limitations. The first one refers to the topics addressed. Intimate partner violence and sexism are sensitive topics and whether the results replicate with more neutral topics needs to be examined. Further, the sample used in this the study was a non-probabilistic one, and the generalizability of our findings cannot be assured. Panels have been shown to overrepresent individuals with certain demographic characteristics and personality traits (e.g., openness, conscientiousness) (Unangst et al. 2020). In addition, there is evidence that data quality is generally lower among non-probability panels, with recent studies indicating that there is significantly more straight-lining in opt-in panels than in probability-based panels (Cornesse and Blom 2020). The extent to which scale direction effects differ between panelists and general populations is a question for future research. In our study, respondents chose the device of their preference, confounding measurement effects with selection effects. Future research could overcome these limitations by replicating experiments of this type, combining random sampling with random assignment.

It is also important to highlight that, in our study, the question layout (i.e., horizontal vs. vertical) was not experimentally manipulated and depended on the device used to complete the questionnaire. In addition, this study used five-point bipolar scales to test the effects of scale direction throughout the questionnaire and effects may be different for scales of different length, polarity, symmetry, and labeling options (Yan et al. 2018). Finally, the scales used were located toward the beginning and in

the middle of the questionnaire. Since some studies have shown that location might moderate scale direction effects (Yan and Keusch 2018), comparing the effects of scale directionality at different stages of the questionnaire is an important avenue for future research.

Conclusion and Implications

This research shows that the direction in which scales are presented influences response distributions but has limited effect on data quality, measured as internal consistency and differentiation. When presented with scales in decremental order, respondents were less likely to select strongly disagree, which was the last response option. Although scale direction had some impact on response distributions and average scores, non-differentiation was marginally affected by the order in which response options were presented in only one of the scales, suggesting that the scale direction does not consistently affect response quality. According to these findings, we refrain from making a firm recommendation regarding the preferred response scale directionality (incremental vs. decremental). However, as recent studies have pointed out, it is highly recommended that response order be kept in mind when designing questionnaires (Höhne and Krebs 2017; Salzberger and Koller 2019). Finally, since the device used to complete the survey had an impact on some of the indicators analyzed (i.e., lower response differentiation and a lower number of strongly agree answers among PC respondents), we recommend that the format be consistent throughout the devices, using item-specific questions in all cases to avoid unintended device effects (Revilla et al. 2017).

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research has been funded by the Department of Public and Business Law at the University of Castilla–La Mancha.

Supplemental Material

Supplemental material for this article is available online.

Notes

1. A brief description of the Netquest panel is available at: https://www.netquest.com/hubfs/docs/panelbooks/netquest_panel_book_EN.pdf Information about the panel in Spain is included on page 16.
2. We received the approval from the authors to use their instrument on October 15, 2018.
3. As there was a small proportion of respondents who chose tablets to complete the questionnaire (5.7%), and the question layout was the same as the one displayed for smartphones (i.e., vertical individual questions), we collapsed both groups.

References

Belson, W. A. 1966. The effects of reversing the presentation order of verbal rating scales. *Journal of Advertising Research* 6:30–37.

Bishop, G. F., and A. Smith. 2001. Response-order effects and the early Gallup split-ballots. *Public Opinion Quarterly* 65:479–505.

Chang, L., and J. A. Krosnick. 2009. National surveys via RDD telephone interviewing versus the Internet. Comparing simple representativeness and response quality. *Public Opinion Quarterly* 73:641–78.

Christian, L. M., N. L. Parsons, and D. A. Dillman. 2009. Designing scalar questions for web surveys. *Sociological Methods & Research* 37:393–425.

Cornesse, C., and A. G. Blom. 2020. Response quality in nonprobability and probability-based online panels. *Sociological Methods & Research* 1–30.

DeCastellarnau, A. 2018. A classification of response scale characteristics that affect data quality: A literature review. *Quality & Quantity* 52:1523–59.

Diedenhofen, B. 2016. cocron: Statistical comparisons of two or more alpha coefficients (Version 1.0-1).

Ekehammar, B., N. Akrami, and T. Araya. 2000. Development and validation of Swedish classical and modern sexism scales. *Scandinavian Journal of Psychology* 41:307–14.

Elway, H. S. 2013. *Satisfaction scale comparison. Internal report*. Seattle: Elway Research.

Galesic, M., R. Tourangeau, M. P. Couper, and F. Conrad. 2008. Eye-tracking data: New insights on response order effects and other signs of cognitive shortcuts in survey responding. *Public Opinion Quarterly* 72:892–913.

Garbarski, D., N. C. Schaeffer, and J. Dykema. 2018. The effects of features of survey measurement on self-rated health: Response option order and scale orientation. *Applied Research in Quality of Life* 14:545–60.

Höhne, J. K., and D. Krebs. 2017. Scale direction effects in agree/disagree and item-specific questions: A comparison of question formats. *International Journal of Social Research Methodology* 21:91–103.

Höhne, J. K., and T. Lenzner. 2015. Investigating response order effects in web surveys using eye-tracking. *Psihologij* 48:361–77.

Israel, G. D. 2006. *Visual cues and response format effects in mail surveys*. Revised version of the paper presented at the Annual Meeting of the Southern Rural Sociological Association, Orlando, FL, February 4–8.

Keusch, F., and T. Yan. 2018. *The influence of rating scale direction on reliability*. Paper presented at the annual conference of the American Association for Public Opinion Research, Denver, CO, May 16–19.

Krebs, D., and J. H. P. Hoffmeyer-Zlotnik. 2010. Positive first or negative first? Effects of the order of answering categories on response behavior. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* 6:118–27.

Krebs, D., and J. K. Höhne. 2019. Exploring scale direction effects and response behavior across PC and smartphone behavior across PC and smartphone surveys. *Journal of Survey Statistics* 1–19.

Krosnick, J. A. 1991. Response strategies for coping with the demands of attitude measures in surveys. *Applied Cognitive Psychology* 5:213–36.

Liu, M., and A. Cernat. 2018. Item-by-item versus matrix questions: A web survey experiment. *Social Science Computer Review* 36:690–706.

Liu, M., and F. Keusch. 2017. Effects of scale direction on response style of ordinal rating scales. *Journal of Official Statistics* 33:137–54.

Malhotra, N. 2008. Completion time and response order effects in web surveys. *Public Opinion Quarterly* 72:914–34.

Mavletova, A., M. P. Couper, and D. Lebedev. 2018. Grid and item-by-item formats in PC and mobile web surveys. *Social Science Computer Review* 36:647–68.

Menold, N., and K. Bogner. 2014. *Gestaltung von Ratingskalen in Fragebögen. SDM Survey Guidelines*. Mannheim, Germany: GESIS.

Menold, N., and K. Bogner. 2016. *Design of rating scales in questionnaires. GESIS Survey Guidelines*. Mannheim, Germany: GESIS—Leibniz Institute for the Social Sciences.

Rammstedt, B., and D. Krebs. 2007. Does response scale format affect the answering of personality scales? Assessing the big five dimensions of personality with different response scales in a dependent sample. *European Journal of Psychological Assessment* 23:32–38.

Revilla, M., D. Toninelli, and C. Ochoa. 2017. An experiment comparing grids and item-by-item formats in web surveys completed through PCs and smartphones. *Telematics and Informatics* 34:30–42.

Salzberger, T., and M. Koller. 2013. Towards a new paradigm of measurement in marketing. *Journal of Business Research* 66:1307–17.

Salzberger, T., and M. Koller. 2019. The direction of the response scale matters—accounting for the unit of measurement. *European Journal of Marketing* 53: 871–91.

Schwarz, N., and H.-J. Hippler. 2004. Response alternatives: The impact of their choice and presentation order. In *Measurement errors in surveys*, eds. P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman, 41–56. Hoboken, NJ: Wiley-Interscience.

Schwarz, N., H.-J. Hippler, and E. Noelle-Newmann. 1992. A cognitive model of response-order effects in survey measurement. In *Context effects in social and psychological research*, eds. N. Schwarz and S. Sudman, 187–201. New York: Springer.

Smyth, J. D., G. D. Israel, M. G. Newberry, and R. G. Hull. 2019. Effects of stem and response order on response patterns in satisfaction ratings. *Field Methods* 31: 260–76.

Stapleton, C. 2013. The smart(phone) way to collect survey data. *Survey Practice* 6: 1–8.

Stern, M., D. A. Dillman, and J. D. Smyth. 2007. Visual design, order effects, and respondent characteristics in a self-administered survey. *Journal of Official Statistics* 33:137–54.

Toepoel, V., M. Das, and A. van Soest. 2009. Design of web questionnaires: The effect of layout in rating scales. *Journal of Official Statistics* 25:509–28.

Unangst, J., A. E. Amaya, H. L. Sanders, J. Howard, A. Ferrell, S. Karon, and J. A. Deber. 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.

Weng, L.-J., and C.-P. Cheng. 2000. Effects of response order on Likert-type scales. *Educational and Psychological Measurement* 60:908–24.

Yan, T., and M. Hu. 2018. Examining translation and respondents' use of response scales in 3MC surveys: Multinational, multiregional, and multicultural contexts (3MC). In *Advances in comparative survey methods: Multinational, multiregional, and multicultural contexts (3MC)*, eds. T. P. Johnson, B. Pennell, I. A. L. Stoop, and B. Dorer, 501–18. Hoboken, NJ: John Wiley & Sons Inc.

Yan, T., and F. Keusch. 2015. The effects of the direction of rating scales on survey responses in a telephone survey. *Public Opinion Quarterly* 79:145–65.

Yan, T., and F. Keusch. 2018. *Direction of agree-disagree rating scales and data quality*. Paper presented at the annual conference of the American Association for Public Opinion Research. Denver, CO, May 16–19.

Yan, T., F. Keusch, and L. He. 2018. The impact of question and scale characteristics on scale direction effects. *Survey Practice* 11:1–9.