Exploring Scale Direction Effects and Response Behavior across PC and Smartphone Surveys

Dagmar Krebs
University of Gießen (Germany)

Jan Karem Höhne
University of Mannheim (Germany)
Universitat Pompeu Fabra (Spain)

Abstract
The effects of scale direction on response behavior are well-known in the survey literature, where a variety of theoretical approaches are discussed, and mixed empirical findings are reported. In addition, different types of survey completion devices seem to vary in their susceptibility to scale direction effects. In this study, we therefore investigate the effect of scale direction and device type on response behavior in PC and smartphone surveys. To do so, we conducted a web survey experiment in a German non-probability access panel (N = 3,401) using a two-step split-ballot design with four groups that are defined by device type (PC and smartphone) and scale direction (decremental and incremental). The results reveal that both PCs and smartphones are robust against scale direction effects. The results also show that response behavior differs substantially between PCs and smartphones indicating that the device type (PC or smartphone) matters. In particular, the findings show that the comparability of data obtained through multi-device surveys is limited.

Keywords: Latent means, measurement invariance, multi-device survey, rating scales, response behavior, scale direction

Introduction and background
This study addresses two research objectives that are central to quantitative social research: rating scale direction (decremental and incremental) and device type (PC and smartphone). In line with these two research objectives, we first address effects of rating scale direction and then effects of device type. At the end of this introductory chapter, we outline the scope of this experimental study.

Effects of rating scale direction
Following the early study by Rugg and Cantrill (1944), survey researchers began investigating response order effects on response behavior in the contexts of lists of unordered response categories (e.g., categories capturing child qualities; Krosnick and Alwin 1987, p. 205). Response order effects are, however, not limited to unordered response categories. Often

---

This document is a preprint and thus it may differ from the final version: Krebs, Dagmar, & Höhne, Jan K. (2021). Exploring scale direction effects and response behavior across PC and smartphone surveys. Journal of Survey Statistics and Methodology, 9, 477-495. DOI: 10.1093/jssam/smz058.

defined as a special case of response order effects, scale direction effects occur in the context of lists of ordered response categories (e.g., rating scales running from "applies completely" to "applies not at all"). While the survey literature abounds in studies on response order effects in lists with unordered categories, research on scale direction effects in rating scales is rather scarce.

In general, response order effects can be divided into two forms (Sudman, Bradburn and Schwarz 1996, p. 130). The first form, sometimes called "primacy effects", occurs when respondents tend to select the response categories at the beginning of the scale more frequently. The second form, sometimes called "recency effects", occurs when respondents tend to select the response categories at the end of the scale more frequently. While primacy effects occur more frequently in surveys with visual presentation, such as self-administered web surveys, recency effects occur more frequently in surveys with verbal presentation, such as telephone surveys (Krosnick 1991).[1]

When rating scales are used to measure attitudes and opinions, the results reported in the literature indicate that responses are more likely to accumulate at the beginning of the scale, indicating the occurrence of primacy effects (see Galesic, Tourangeau, Couper and Conrad 2008; Höhne and Krebs 2018; Höhne and Lenzner 2015; Höhne, Revilla and Lenzner 2018; Keusch and Yan 2018; Krebs 2012; Krebs and Bachner 2018; Krebs and Hoffmeyer-Zlotnik 2010; Mavletova 2013; Menold and Bogner 2015; Rammstedt and Krebs 2007; Sudman et al. 1996; Toepoel 2008; Yan and Keusch 2015). These findings are obtained irrespective of the survey presentation form (visual or verbal). Some studies report that these effects occur only in decremental scales running from the positive or high side to the negative or low side (Krebs and Hoffmeyer-Zlotnik 2010; Menold and Bogner 2015) and others report that these effects occur only in incremental scales running from the negative or low side to the positive or high side (Menold and Bogner 2015; Toepoel 2008).

According to Krosnick (1991) respondents either process the response categories at the beginning of the scale more intensively than the following ones, or they simply select the first acceptable response category without considering the following ones. In eye-tracking studies, Galesic et al. (2008) and Höhne and Lenzner (2015) found evidence for both types of response behaviors in lists with ordered and unordered response categories (visual presentation). These response behaviors are called satisficing and imply a superficial, rather than a thoughtful response process (Krosnick 1991).

The size of scale direction effects may depend on the format of the scale itself. According to Sudman et al. (1996), empirical evidence indicates that rating scales exhibit smaller effects than lists with unordered categories. In addition, the alignment of rating scales has an impact. Höhne and Lenzner (2015) showed that primacy effects are more common in vertically aligned rating scales than in horizontally aligned rating scales. In their study, eye-tracking data revealed that in vertical rating scales, respondents fixate on the categories at the beginning more frequently, but in horizontal rating scales, respondents fixate on the categories in the middle

[1] As mentioned by Keusch and Yan (2018), the terms primacy and recency effects including their theoretical frameworks are usually associated with unordered rather than ordered categories. We adopt the terms primacy and recency effects to simply describe respondents' response tendency – either to the beginning of the rating scale or to the end of the rating scale (see Sudman et al. 1996, p. 157 for a similar practice).

more frequently. This implies a connection between fixation behavior and category selection that depends on the alignment of the rating scales.

Effects of device type

In recent years, the number of web surveys has increased markedly (Mavletova 2013; Mavletova, Couper and Lebedev 2018; Tourangeau et al. 2017). This trend applies to web surveys completed on PCs, but also to those completed on smartphones (Gummer, Quoß and Roßmann 2018; Revilla, Toninelli, Ochoa and Loewe 2016). However, as Tourangeau et al. (2017) state, concerns remain about the quality of survey data collected via smartphones. The main reason behind these concerns is that the screen sizes of smartphones, which are small compared to those of PCs, decrease the visibility of survey questions and response categories. On smartphones, in order to provide their responses, respondents must frequently engage in vertical scrolling and screen tabs to locate the question stems and response categories. In contrast, on PCs respondents usually engage in less vertical scrolling and no screen tabs to locate question stems and response categories. The less arduous process on PCs may facilitate survey operation and navigation and thereby improves the quality of responses (see Couper and Peterson 2017 for a discussion of the impact of scrolling on response behavior). Research also suggests that respondents are more likely to select response categories that are immediately visible than response categories that require additional actions (e.g., scrolling) to make them visible (Couper, Tourangeau, Conrad and Crawford 2004; Galesic et al. 2008; Mavletova 2013).

Lugtig and Toepoel (2016) also refer to the process of data entry as a potential source of the differences between the responses obtained via PCs and smartphones. Responses on PCs are commonly given using mouse movements and clicks or by typing on a keyboard, while responses on smartphones are provided through finger movements and tabs on the screen. The authors argue that the input capabilities of smartphones are less precise and more effortful than those of PCs (p. 80).

In addition to device-related issues, such as the visibility of survey or question contents and input capabilities, context-related issues may also lead to differences in response behavior between PC and smartphone respondents. Smartphones allow respondents to participate in web surveys whenever and wherever they want with almost no locality, situation, or time restrictions (Mavletova 2013). Consequently, smartphone respondents – compared to PC respondents – are more frequently surrounded by other people or are engaged in multitasking behavior during survey completion (Toninelli and Revilla 2016a). As shown by Toninelli and Revilla (2016b), in comparison to PC respondents, smartphone respondents are more likely to report having distractions and engaging in multitasking behavior. This can foster the occurrence of response bias (Sendelbah, Vehovar, Slavec and Petrovčič 2016).

Both device-related and context-related issues induced a methodological discussion on response behavior across PCs and smartphones. For instance, several studies suggested that the occurrence of scale direction effects may depend on the device used in survey responding (Höhne et al. 2018, Mavletova 2013, Revilla and Couper 2018; Tourangeau et al. 2017). While some studies on PC and smartphone surveys reported that responding on smartphones may take longer than responding on PCs (Couper and Peterson 2017, p. 360), the findings on the occurrence of response bias, such as scale direction effects, are less clear. For instance, Höhne et al. (2018) investigated primacy effects across agree/disagree and item-specific questions and found slightly higher effects for smartphones than PCs. Mavletova (2013) investigated primacy effects in questions with rating scales and found almost no differences. Revilla and Couper (2018) and Tourangeau et al. (2017) found no clear pattern for primacy effects between PCs and smartphones. All in all, the survey literature consists of a small number of studies comparing PCs and smartphones with respect to scale direction effects. The scarcity of studies and the inconclusive results indicate that further research is necessary.

Scope of the study

To get a more comprehensive understanding of response behavior across PCs and smartphones, we investigate scale direction effects across the two device types using a survey experiment. To this end, we compare responses to decremental and incremental rating scales within each device type (PC and smartphone). In addition to previous research, we tackle device effects by comparing responses on PCs and smartphones within each scale direction (decremental and incremental). To do so, we randomly assign respondents to device types and scale directions, which results in four experimental groups.

In the next sections, we describe our research hypotheses, study design, survey questions, procedure of the study, sample, and analytical strategy. We then present the results of the statistical analyses. Finally, we provide a discussion of the theoretical and practical implications of our findings and suggest perspectives for future research.

Research hypotheses

Scale direction effects: In line with the previous argumentation, we assume that response behavior differs across PCs and smartphones due to device-related issues, such as screen size and input capabilities, and context-related issues, such as distractions and multitasking. More specifically, we expect the previously discussed device-related and context-related issues to foster a less intent and less diligent response behavior on smartphones than on PCs. This response behavior manifests itself in the form of stronger scale direction effects – i.e., a stronger tendency towards the beginning of the rating scale – in smartphones than PCs. We therefore hypothesize that shifts in latent means between decremental and incremental scales are more pronounced in smartphones than PCs (Hypothesis 1).

Device type effects: To ascertain that the hypothesized stronger scale direction effects for smartphones than PCs (see Hypothesis 1) do not imply differential measurement, we also test for measurement invariance between PCs and smartphones within scale directions. Although there is empirical evidence that smartphone respondents – compared to PC respondents – are more likely to select positive response categories on decremental than on incremental scales (Tourangeau et al. 2017), this does not necessarily imply differences in measurement properties (Meitinger 2017; Revilla 2013). Thus, we hypothesize measurement invariance between PCs and smartphones for both decremental (Hypothesis 2a) and incremental scales (Hypothesis 2b).

In line with our expectation that scale direction effects are stronger in smartphones than PCs (see Hypothesis 1), we hypothesize significant differences in latent means between PCs and smartphones for both decremental (Hypothesis 3a) and incremental scales (Hypothesis 3b).

Method

Study design

Only respondents who owned both a PC and a smartphone were invited to take part in this study. To identify these respondents, we used profiling information provided by the survey company. Before the survey started, we randomly assigned respondents to a device (PC or smartphone). If respondents tried to enter the survey with a different device than the one they were assigned, they were blocked from the survey and asked to switch to the correct device. In addition, we collected user-agent-strings that inform about device properties, such as device model and Internet browser (see Callegaro 2013).

After device assignment, we randomly assigned respondents to a scale direction (decremental or incremental) within devices. This means that we used a two-step split-ballot design with four experimental groups defined by device type (PC and smartphone) and scale direction (decremental and incremental). This resulted in a 2-by-2 factorial design, as shown in Table 1. The decremental scale ran from "applies completely" to "applies not at all" and the incremental scale ran from "applies not at all" to "applies completely". For statistical analyses, all responses were identically coded to values ranging from 1 "applies not at all" to 7 "applies completely".[2]

Table 1. Experimental design defined by device type and scale direction

Experimental group	Device type	Scale direction	Group size
1	PC	Decremental	837
2	PC	Incremental	843
3	Smartphone	Decremental	867
4	Smartphone	Incremental	854

Survey questions

The questions used in this study were adapted from those used in the Cross Cultural Survey for Work and Gender Attitudes (2010) and the German General Social Survey (2006). The advantage of using pre-existing questions from established social surveys is that these questions have been repeatedly tested. For this study, we used 16 questions. Five questions dealt with achievement motivation and were presented on one survey page. The remaining eleven questions dealt with job motivation (three on extrinsic, four on intrinsic, and four on social job motivation) and were presented on two survey pages. The questions were presented with an item-by-item presentation, seven-point, end-labeled rating scales, a vertical alignment, horizontally shaded response categories, and no numeric values. All questions were in German (see Appendix A for English translations), which was the mother tongue of 95.1% of the respondents. To improve comparability between PCs and smartphones, we used an optimized survey layout that avoids horizontal scrolling. Figure 1 illustrates the question design used for PCs and smartphones.

[2] The experiment was part of a larger study with several unrelated experiments, all of which were independently randomized limiting carryover effects.

Figure 1. Examples of achievement motivation questions used for PCs and smartphones

Note. The PC version is presented on the left side and the smartphone version on the right side (decremental direction only).

Procedure of the study

The data collection of this study was conducted by the survey company Respondi in Germany from September 25, 2018 to October 17, 2018. The company invited respondents varying in age from 18 to 70 years by email (in total, they sent out 36,585 email invitations). The email included an invitation to take part in the survey, an instruction on which device (PC or smartphone) to use for survey completion, and a URL link that directed respondents to the survey. Once there, an introductory page described the survey topic and procedure and informed respondents that their data would be treated confidentially. In case of participation, respondents received an incentive from the survey company, which was proportional to the length of the entire survey.

To draw a sample that is representative of the German population, we used cross quotas. Based on the German micro-census, we implemented cross quotas for age, education, and gender, which resulted in a 3×3×2 quota plan.

We collected several types of paradata, such as response times and user-agent-strings, using the open-source tool “Embedded Client Side Paradata (ECSP)” (Schlosser and Höhne 2018). Prior informed consent for the collection of paradata was obtained by Respondi as part of their registration process.

Sample

A total of 3,829 respondents reacted to the email invitation of the survey company. Some respondents were excluded from the study (n = 428) because they only visited the title page or they broke-off the survey before being asked any experimental questions. In total, n = 3,401 respondents (1,680 using PCs and 1,721 using smartphones) completed the survey, which corresponds to a participation rate of 9.3% (in relation to the total number of email invitations sent).

These participating respondents had a mean age of 46.6 (SD = 14.9) and 49.7% of them were female. In terms of education, 35.9% graduated from a lower secondary school, 29.6% from an intermediate secondary school, and 34.5% from a college preparatory secondary school or university.

To evaluate the effectiveness of the random assignment and the sample composition between the four experimental groups, we conducted chi-square tests. The results showed no statistically significant differences regarding age [$\chi^2(6) = 6.91$, p = .33], gender [$\chi^2(3) = .55$, p = .91], and education [$\chi^2(6) = 2{,}63$, p = .85].

Analytical strategies

The first step of our analysis was to calculate the means and standard deviations of the responses to the 16 survey questions (and their indices or composites) employed in this study.

To investigate scale direction effects, we conducted a multigroup confirmatory factor analysis (MG-CFA) that contained four latent variables with 16 indicators. We tested this dimensional structure for measurement invariance between decremental and incremental scale directions within PCs and smartphones. In the next step, we analyzed shifts in latent means to test for scale direction effects.

We then tested for measurement invariance between PCs and smartphones within decremental and incremental scale directions using the same dimensional structure (four latent variables with 16 indicators). Again, shifts in latent means between PCs and smartphones were analyzed.

The descriptive statistics were conducted using SPSS version 24 and the MG-CFAs were conducted using Mplus version 6.12 (see Appendix B for examples of the Mplus commands used). Since all indicators of the latent variables were measured on seven-point, end-labeled rating scales, we assumed a continuous scale level (Rhemtulla, Brosseau-Liard and Savalei 2012). Based on that assumption, we used the MLR (instead of MLM) estimator, which provides robust standard errors and accounts for non-normality.

Results

Descriptive statistics

To investigate scale direction effects and response behavior in PCs and smartphones, we first calculated the means and standard deviations of the responses to all 16 questions and their corresponding indices (composites). Table 2 reports the descriptive statistics for both scale directions (decremental and incremental) and both device types (PC and smartphone). It is to see that the means and standard deviations of decremental and incremental scales do not differ substantially – neither for PCs nor for smartphones. This observation applies to both PCs and smartphones. Comparing device types, we find that PCs have lower means than smartphones. This applies to both decremental and incremental scales. In contrast to means, standard deviations do not differ substantially between PCs and smartphones, irrespective of the scale direction. These results provide preliminary evidence that PCs and smartphones are equally robust against scale direction effects. However, it seems that there is a device effect at work influencing response behavior.[3]

[3] In Appendix C, we also report response times in seconds across the four experimental groups.

Table 2. Means and standard deviations of responses to questions (and their indices) for decremental and incremental scale directions within PCs and smartphones

	PCs				Smartphones			
	Decremental		Incremental		Decremental		Incremental	
Questions	Means	SDs	Means	SDs	Means	SDs	Means	SDs
Achieve-ment								
1	3.93	1.70	3.85	1.64	3.97	1.70	4.07	1.71
2	4.63	1.69	4.69	1.64	4.83	1.74	4.82	1.74
3	4.37	1.64	4.34	1.60	4.57	1.68	4.48	1.68
4	4.36	1.76	4.35	1.74	4.56	1.75	4.46	1.81
5	4.55	1.75	4.49	1.72	4.82	1.73	4.67	1.76
Index	*4.35*	*1.44*	*4.34*	*1.37*	*4.55*	*1.39*	*4.50*	*1.44*
Extrinsic								
1	4.75	1.47	4.68	1.48	4.97	1.50	4.84	1.59
2	4.35	1.68	4.21	1.72	4.54	1.75	4.56	1.72
3	5.42	1.47	5.35	1.51	5.55	1.44	5.47	1.40
Index	*4.84*	*1.28*	*4.75*	*1.27*	*5.02*	*1.28*	*5.00*	*1.33*
Intrinsic								
1	5.62	1.49	5.63	1.46	5.81	1.38	5.71	1.33
2	5.67	1.32	5.62	1.30	5.80	1.26	5.74	1.28
3	5.29	1.67	5.21	1.42	5.46	1.41	5.42	1.41
4	5.23	1.49	5.22	1.46	5.37	1.48	5.32	1.45
Index	*5.45*	*1.21*	*5.42*	*1.16*	*5.61*	*1.13*	*5.55*	*1.14*
Social								
1	5.55	1.42	5.42	1.45	5.60	1.42	5.54	1.37
2	5.24	1.47	5.13	1.56	5.45	1.50	5.42	1.42
3	4.97	1.51	4.89	1.56	5.07	1.65	5.07	1.54
4	6.25	1.31	6.16	1.31	6.38	1.18	6.37	1.15
Index	*5.50*	*1.13*	*5.39*	*1.14*	*5.63*	*1.12*	*5.60*	*1.07*

Note: Responses to decremental and incremental rating scales were coded to identical values from 1 “applies not at all” to 7 “applies completely”. Index states the means and standard deviations of the composites. Abbreviation: SDs = standard deviations.

Scale direction effects within device types

Scale direction effects can be detected by comparing latent means of responses across decremental and incremental scales within PCs and smartphones. To do that, we initially computed separate but identical confirmatory factor analysis (CFA) baseline models for each scale direction (decremental and incremental) within each device type. Each of these four baseline models contained four latent variables with 16 indicators. In each model, we admitted one error covariance per latent variable, which resulted in four error covariances between questions addressing similar contents. All baseline models had satisfactory goodness-of-fit statistics.

Next, we conducted multigroup confirmatory factor analyses (MG-CFA) within PCs and smartphones. To this end, we first tested configural invariance by simultaneously analyzing the baseline model for the two scale directions within each device type. Table 3 reports the statistical results. Given CFI-values higher than .95 and RMSEA-values of .05, configural invariance was accepted for rating scale directions within both PCs and smartphones. To test metric invariance, factor loadings were constrained to equality between decremental and incremental scale directions within device types. The model goodness-of-fit statistics were satisfactory and, thus, we accepted metric invariance. Finally, to compare latent means, scalar invariance was tested by imposing equality constraints on the intercepts. Again, scalar invariance holds for both rating scale directions in both device types.

There are two criteria for comparability between models with increasing equality constraints: non-significant differences between (mean-adjusted) chi-square values (Byrne 2012) and differences between CFI's and RMSEA's lower than .01 (Cheung and Rensvold 2002). These two criteria hold for all models in Table 3.

Table 3. Testing measurement invariance between decremental and incremental scale directions within PCs and smartphones

	Chi-square value	df	Chi-square difference test	CFI	RMSEA
PCs					
Configural	589.69 (1.39)	188		.956	.050
Metric	602.98 (1.38)	204	9.86	.956	.048
Scalar	624.57 (1.35)	220	11.43	.956	.047
Smartphones					
Configural	590.14 (1.42)	188		.954	.050
Metric	602.68 (1.42)	204	12.54	.954	.048
Scalar	630.61 (1.39)	220	20.59	.953	.047

Note. *p < .05. The results are based on MLR estimation. Scale correction factors for model comparisons are in parentheses.

Differences in latent means between responses across the rating scale directions were tested using the decremental direction as reference group. Since the responses for both scale directions were coded from 1 "applies not at all" to 7 "applies completely", estimates with negative signs indicate that responses on the incremental scales are more negative (i.e., the responses have lower values) than those on the decremental scales. Table 4 shows the results of the comparison of latent means for PCs [$\chi^2(216) = 617.00$ (1.36); CFI = .956; RMSEA = .047] and smartphones [$\chi^2(216) = 626.35$ (1.39); CFI = .953; RMSEA = .047]. In line with the descriptive statistics, the shifts in latent means between decremental and incremental scale directions are small for PCs and smartphones. There is only one exception to this trend, namely the social job motivation responses for PCs, which has a marginally significant latent mean shift. Overall, there is no supporting evidence for Hypothesis 1, which postulated stronger scale direction effects for smartphones than for PCs. This finding, in turn, indicates that PCs and smartphones are equally robust against scale direction effects.

Table 4. Latent mean differences between decremental and incremental scale directions within PCs and smartphones (unstandardized results)

PCs	Estimate	Standard error	Critical ratio	p-value
Achievement	–.008	.052	–.162	.872
Extrinsic	–.091	.060	–1.532	.126
Intrinsic	–.040	.053	–.75	.454
Social	–.111	.057	–1.948	.051
Smartphones				
Achievement	–.053	.051	–1.032	.302
Extrinsic	–.049	.049	–.860	.390
Intrinsic	–.054	.054	–1.033	.302
Social	–.038	.038	–.666	.506

Note. Responses to decremental and incremental rating scales were identically coded to values ranging from 1 "applies not at all" to 7 "applies completely". The reference group was the decremental scale direction.

Device effects within scale directions

To investigate device effects, we first tested whether responses on PCs and smartphones differ in terms of measurement invariance. This was done for decremental and incremental scale directions. Again, we conducted multigroup confirmatory factor analyses (MG-CFA), this time within scale directions. All analytical procedures and model characteristics are identical to those reported above. As the statistical results reported in Table 5 show, configural and metric invariance hold for both device types within decremental and incremental scales. However, scalar invariance holds only for the decremental scale direction. For the incremental scale direction, in contrast, only metric invariance can be accepted, as indicated by the significant chi-square difference test. Recall that we hypothesized that measurement invariance between PCs and smartphones holds for decremental (Hypothesis 2a) and incremental scales (Hypothesis 2b). Our findings, therefore, support Hypothesis 2a, but not Hypothesis 2b.

Table 5. Testing measurement invariance between PCs and smartphones within decremental and incremental scale directions

	Chi-square value	Df	Chi-square Difference Test	CFI	RMSEA
Decremental					
Configural	605.25 (1.41)	188		.954	.051
Metric	624.63 (1.40)	204	16.44	.954	.049
Scalar	654.78 (1.37)	220	22.85	.952	.048
Incremental					
Configural	574.55 (1.41)	188		.956	.049
Metric	598.55 (1.40)	204	21.72	.955	.048
Scalar	639.32 (1.37)	220	38.38*	.952	.047

Notes. *$p < .05$. The results are based on MLR estimation. Scale correction factors for model comparison are in parentheses.

Next, we compared the shifts in latent means between PCs and smartphones within the decremental scale direction. Table 6 summarizes the statistical results [$\chi^2(216) = 643.20$ (1.38); CFI = .953; RMSEA = .048]. The latent means exhibit significant shifts between PCs and smartphones for all four latent variables. This indicates that responses on smartphones are more positive (i.e., the responses have higher values) than responses on PCs. A strong device effect is, therefore, at play within decremental scales. Recall that we hypothesized significant differences in latent means between PCs and smartphones for decremental (Hypothesis 3a) and incremental scales (Hypothesis 3b). Thus, the results in this section support Hypothesis 3a. Due to the lack of scalar invariance, however, we did not test Hypothesis 3b on incremental scales. Nevertheless, the lack of scalar invariance for incremental scales indicates strong differences in response behavior between PCs and smartphones.

Table 6. Latent mean differences between PCs and smartphones within decremental scale direction (unstandardized results)

Decremental	Estimate	Standard error	Critical ratio	p-value
Achievement	.155	.051	3.016	.003
Extrinsic	.155	.058	2.664	.008
Intrinsic	.140	.053	2.657	.008
Social	.119	.056	2.115	.034

Note. Responses to decremental and incremental rating scales were identically coded to values ranging from 1 “applies not at all” to 7 “applies completely”. The reference group is PC.

Discussion and conclusion

The aim of this study was to investigate the occurrence of scale direction effects across PCs and smartphones. In addition, we investigated whether response behavior differs across device types. For this purpose, we conducted a web survey experiment with four groups defined by scale direction and device type and carried out two sets of comparisons. First, we compared decremental and incremental scale directions within PCs and smartphones. Second, we compared response behavior between PCs and smartphones within decremental and incremental scales. Our findings revealed the comparability of responses across different scale directions within both PCs and smartphones. The findings also showed substantial differences in response behavior between PCs and smartphones within each scale direction.

To test our first hypothesis, which postulated stronger scale direction effects for smartphones than for PCs, we first tested for measurement invariance between decremental and incremental scale directions. The results of a multigroup confirmatory factor analysis (MG-CFA) provided evidence that scalar measurement invariance holds for both scale directions in both device types. Contrary to the postulation of Hypothesis 1, the shifts in latent means between decremental and incremental scales were not consistently larger for smartphones than for PCs. In addition, the shifts were not significantly different within either device. This finding suggests that scale direction effects on response behavior are not affected by the device type used in web survey responding and corroborates the scholarly consensus that direction effects in rating scales are consistently small (Sudman et al. 1996), including cases where the scales are vertically aligned (Höhne and Lenzner 2015).

To test our Hypotheses 2a and 2b, which postulated that measurement invariance between PCs and smartphones holds for decremental and incremental scales, respectively, we tested for measurement invariance between PCs and smartphones within decremental and incremental scales. Scalar invariance could only be obtained for the decremental scales. For the incremental scales, in contrast, only metric invariance was obtained. This implies that for incremental scales, the intercepts differ between PCs and smartphones. Since intercepts can be seen as “additive constants” (Steinmetz 2013, p. 8), this intercept difference points to the presence of systematic measurement error. One potential cause, as an example, may be that incremental scales are perceived (or handled) differently by respondents completing questions on PCs and smartphones. This difference in perception (or handling) can lead, in turn, to a systematic distortion of respondents’ cognitive response process resulting in measurement error (Groves et al. 2004). This explanation, however, lacks empirical evidence. Thus, we suggest that future research scrutinizes this phenomenon, as it would allow to enhance the comparability of data collected in multi-device surveys.

In accordance with our Hypothesis 3a, we compared latent means between PCs and smartphones within decremental scales. The statistical results revealed significant shifts in latent means for all four latent variables. More specifically, responses on smartphones are considerably more positive than responses on PCs, which might point to a positivity bias (Tourangeau, Rips and Rasinski 2000) for smartphones. This result provides strong evidence that there is a device effect at work that affects response behavior to decremental scales. This result is also in line with Tourangeau et al. (2017), who found that smartphone respondents are more likely to select positive response categories in decremental scales than in incremental scales.

We did not test Hypothesis 3b, which postulated significant differences in latent means between PCs and smartphones for incremental scales, due to the lack of scalar invariance, which precluded the comparison. However, we take this lack of scalar invariance between PCs and smartphones as an indicator of device type-related differences within scale directions as well as scale direction-related differences between device types. There is a great need for research that systematically investigates device effects on response behavior across different scale directions.

Our study has some limitations. First, we used a quota sample (based on age, education, and gender) from a non-probability opt-in access panel. This does not decrease the internal validity of our study, but it might limit the generalizability of our empirical findings. Hence, it would be worthwhile to investigate scale direction and device effects using a probability-based sample to increase external validity. Second, respondents of this study were members of an opt-in access panel who participate in web surveys on a regular basis, so they have a high level of survey experience. Some research indicates that respondents with high survey experience differ from respondents with low survey experience in terms of response behavior (Toepoel, Das and van Soest 2008). For this reason, future studies should take survey experience into account. Third, in this study, we randomly assigned respondents to using a PC or smartphone for survey completion (see Keusch and Yan 2017 for the problem of requesting respondents to use a specific device type). Since participation in web surveys is still more common on PCs than on smartphones, break-offs due to non-compliance with the device assignment might lead to a self-selection bias so that only highly motivated smartphone respondents took part in the survey. Furthermore, forcing respondents to use a smartphone may change the context of survey participation so that the commonly expected context-related issues between PCs and smartphones do not apply. This could also explain the null-findings with respect to the scale direction effects. Finally, we chose a specific rating scale design, namely a seven-point, end-labeled rating scale where we systematically varied the scale direction (decremental and incremental). Hence, future research could vary additional scale design characteristics, such as scale length (e.g., five, six, and seven points) and the extent of verbal labels (e.g., fully and partially labeled).

Our findings showed that decremental and incremental scales are invariant within PCs and within smartphones. The findings also revealed that both device types are robust against scale direction effects. To the best of our knowledge, these findings have not been observed in previous studies dealing with scale direction effects in PCs and smartphones and, thus, they make a contribution to survey methodological research. All in all, it seems that scale direction effects pose only a minor threat to data quality in multi-device surveys, while the device type used in web survey completion poses a threat. Even if scalar invariance between PCs and smartphones can be obtained, as it is the case with the decremental scale direction, the existence of significant shifts in latent means must be heeded. The results of this study present strong evidence for a device effect on response behavior. Considering the increasing importance of multi-device surveys coupled with the importance of measuring respondents' attitudes and opinions for social research and adjacent research fields, it is crucial for future research to explore appropriate methods to prevent or minimize these device effects. This also means that decisions about scale directions in multi-device surveys should be well considered and the results from multi-device surveys should be interpreted with caution.

References

Byrne, B. M. (2012), *Structural Equation Modeling with Mplus. Basic Concepts, Applications, and Programming*, New York, NY: Routledge.

Callegaro, M. (2013), "Paradata in web surveys," in *Improving Surveys with Paradata. Analytic Uses of Process Information*, (pp. 261–280), ed. F. Kreuter, Hoboken, NJ: John Wiley and Sons.

Cheung, G. W., and Rensvold, R. B. (2002), "Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance," *Structural Equation Modeling: A Multidisciplinary Journal*, 9, 233–255.

Couper, M. P., and Peterson, G. J. (2017), "Why do Web Surveys Take Longer on Smartphones?" *Social Science Computer Review*, 35, 357–377.

Couper, M. P., Tourangeau, R., Conrad, F. G., and Crawford, S. D. (2004), "What They See Is What We Get: response Options for Web Surveys," *Social Science Computer Review*, 22, 111-127.

Galesic, M., Tourangeau, R., Couper, M. P., and Conrad, F. G. (2008), "Eye-Tracking Data: New Insights on Response Order Effects and other Cognitive Shortcuts in Survey Responding," *Public Opinion Quarterly*, 72(5), 892–913.

Groves, R. M., Fowler, F. L., Couper, M. P., Lepkowski, J. M., Singer, E., and Tourangeau, R. (2004), *Survey Methodology*, Hoboken, NJ: John Wiley and Sons.

Gummer, T., Quoß, F., and Roßmann, J. (2018), "Does increasing mobile device coverage reduce heterogeneity in completing web surveys on smartphones?" *Social Science Computer Review*. DOI: 10.1177/0894439318766836.

Höhne, J. K., and Krebs, D. (2018), “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 Lenzner, T. (2015), “Investigating Response Order Effects in Web Surveys Using Eye Tracking,” *Psihologija*, 48, 361–377.

Höhne, J. K., Revilla, M., and Lenzner, T. (2018), “Comparing the Performance of Agree/Disagree and Item-Specific Questions Across PCs and Smartphones,” *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 14, 109–118.

Keusch, F., and Yan, T. (2017), “Web Versus Mobile Web: An Experimental Study of Device Effects and Self-Selection Effects,” *Social Science Computer Review*, 35, 751–769.

Keusch, F., and Yan, T. (2018), “Is Satisficing Responsible for Response Order Effects in Rating Scale Questions?,” *Survey Research Methods*, 12(3), 259-270.

Krebs, D. (2012), “The Impact of Response Format on Attitude Measurement,” in *Methods, Theories, and Empirical, Applications in the Social Sciences*, (pp.105-113), eds. S. Salzborn, E. Davidov, and J. Reinecke, VS Verlag für Sozialwissenschaften.

Krebs, D., and Bachner, Y. G. (2018), “Effects of Rating Scale Direction under the Condition of Different Reading Direction,” *Methods, Data, Analyses*, 12(1), 21.

Krebs, D., and Hoffmeyer-Zlotnik, J. H. (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, 118–127.

Krosnick, J. A. (1991), “Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys,” *Applied Cognitive Psychology*, 5, 213–236.

Krosnick, J. A., and Alwin, D. F. (1987), “An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement,” *Public Opinion Quarterly*, 51, 201–219.

Lugtig, P., and Toepoel, V. (2016), “The use of PCs, Smartphones, and Tablets in a Probability-based Panel Survey. Effects on Survey Measurement Error,” *Social Science Computer Review*, 34, 78–94.

Mavletova, A. (2013), “Data Quality in PC and Mobile Web Surveys,” *Social Science Computer Review*, 31, 725–743.

Mavletova, A., Couper, M. P. and Lebedev, D. (2018), “Grid and item-by-item formats in PC and mobile web surveys,” *Social Science Computer Review*, 36, 647–668.

Meitinger, K. (2017), “Necessary but Insufficient: Why Measurement Invariance Tests Need Online Probing as a Complementary Tool,” *Public Opinion Quarterly*, 81, 447–472.

Menold, N., and Bogner, K. (2015), *Gestaltung von Ratingskalen in Fragebögen*. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (SDM Survey Guidelines).

Rammstedt, B., and Krebs, D. (2007), “Does Response Scale Format Affect the Answering of Personality Scales?” *European Journal of Psychological Assessment*, 23, 32–38.

Revilla, M. (2018), “Measurement Invariance and Quality of Composite Scores in a Face-to-Face and a Web Survey,” *Survey Research Methods*, 7, 17–28.

Revilla, M., and Couper, M. P. (2018), “Testing Different Rank Order Question Layouts for PC and Smartphone Respondents,” *International Journal of Social Research Methodology*, 21, 695–712.

Revilla, M., Toninelli, D., Ochoa, C., and Loewe, G. (2016), “Do Online Access Panels Need to Adapt Surveys for Mobile Devices?” *Internet Research*, 26, 1209–1227.

Rhemtulla, M., Brosseau-Liard, P. E., and Savalei, V. (2012), “When Can Categorical Variables Be Treated as Continuous? A Comparison of Robust Continuous and Categorical SEM Estimation Methods Under Suboptimal Conditions,” *Psychological Methods*, 17, 354–373.

Rugg, D., and Cantril, H. (1944), “The Wording of Questions,” in *Gauging Public Opinion* (pp. 23–50), ed. H. Cantril, Princeton, NJ: Princeton University Press.

Schlosser, S., and Höhne, J. K. (2018), “ECSP – Embedded Client Side Paradata,” *Zenodo*, DOI: 10.5281/zenodo.1218941.

Sendelbah, A., Vehovar, V., Slavec, A., and Petrovčič, A. (2016), “Investigating Respondent Multitasking in Web Surveys using Paradata,” *Computers in Human Behavior*, 55, 777–787.

Steinmetz, H. (2013), “Analyzing observed composite differences across groups: Is partial measurement invariance enough?” *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 9, 1–12.

Sudman, S., Bradburn, N. M., and Schwarz, N. (1996), *Thinking About Answers. The Application of Cognitive Processes to Survey Methodology.* San Francisco, CA: Jossey-Bass.

Toepoel, V. (2008), *A Closer Look at Web Questionnaire Design*, Tilburg: CentER, Center for Economic Research.

Toepoel, V., Das, M., and van Soest, A. (2008), “Effects of design in web surveys – Comparing trained and fresh respondents,” *Public Opinion Quarterly*, 72, 985–1007.

Toninelli, D., and Revilla, M. (2016a), “Smartphones vs PCs: Does the Device Affect the Web Survey Experience and the Measurement Error for Sensitive Topics? A Replication of the Mavletova and Couper’s 2013 Experiment,” *Survey Research Methods*, 10, 153–169.

Toninelli, D., and Revilla, M. (2016b), “Is the Smartphone Participation Affecting the Web Survey Experience?” *Proceedings of the 48th Scientific Meeting of the Italian Statistical Society*, Salerno. ISBN: 9788861970618.

Tourangeau, R., Maitland, A., Rivero, G., Sun, H., Williams, D., and Yan, T. (2017), “Web Surveys by Smartphones and Tablets. Effects on Survey Responses,” *Public Opinion Quarterly*, 81, 896–929.

Tourangeau, R., Rips, L. J., and Rasinski, K. (2000), *The Psychology of Survey Response*, Cambridge: Cambridge University Press.

Yan, T., and Keusch, F. (2015), “The Effects of the Direction of Rating Scales on Survey Responses in a Telephone Survey,” *Public Opinion Quarterly*, 79, 145–165.

Appendix A

English translations of the survey questions on achievement and job motivation with an item-by-item presentation for PCs and smartphones (see also Figure 1). Decremental direction only.

Questions on achievement motivation.

I like being in competition with other people. (Achievement 1)

It is satisfying when I achieve better results than other people. (Achievement 2)
I am always trying to perform better than other people. (Achievement 3)
I try harder when I am in competition with other people. (Achievement 4)
It is important for me to be the best at a task. (Achievement 5)
Response categories are 1 "applies completely" – 7 "applies not at all"

Questions on job motivation.
A job with a high income is important for me. (extrinsic 1)
A job with good promotion prospects is important for me. (extrinsic 2)
A job with clear career perspectives is important for me. (extrinsic 3)
A job that I can work autonomously on is important for me. (intrinsic 1)
A job that allows to make use of my skills and talents is important for me. (intrinsic 2)
A job where I have responsibilities for specific tasks is important for me. (intrinsic 3)
A job that allows me to develop my own ideas is important for me. (intrinsic 4)
A job that is recognized and respected is important for me. (social 1)
A job where I can help other people is important for me. (social 2)
A job that contributes to the society is important for me. (social 3)
A job with a good working atmosphere is important for me. (social 4)
Response categories are 1 "applies completely" – 7 "applies not at all"

Note. The order of the questions corresponds to the presentation order in Appendix A. The five questions on achievement motivation were presented on the same page and the eleven questions on job motivation were presented on two pages. The original German wordings of the questions are available from the second author on request.

Appendix B

Mplus commands to track the analyses of measurement invariance and latent means.

```
VARIABLE:
  NAMES ARE v1-v16 device;
  USEVARIABLES ARE v1-v16 device;
  GROUPING IS device (1 = PC_dec 2 = SP_dec);
ANALYSIS:
  ESTIMATOR IS MLR;
MODEL:
  F1 by v1-v5;
  F2 by v6-v8;
  F3 by v9-v12;
  F4 by v13-v16;
  v4 WITH v1;
  v7 WITH v6;
  v10 WITH v9;
  v15 WITH v14;
Model SP_dec:
```

Note. Abbreviations: v1 to v5 (achievement motivation); v6 to v8 (extrinsic job motivation); v9 to v12 (intrinsic job motivation); v13 to v16 (social job motivation); SP = smartphone; dec = decremental.

Appendix C

Average response times across the four experimental groups, which are defined by device type (PC and smartphone) and scale direction (decremental and incremental)

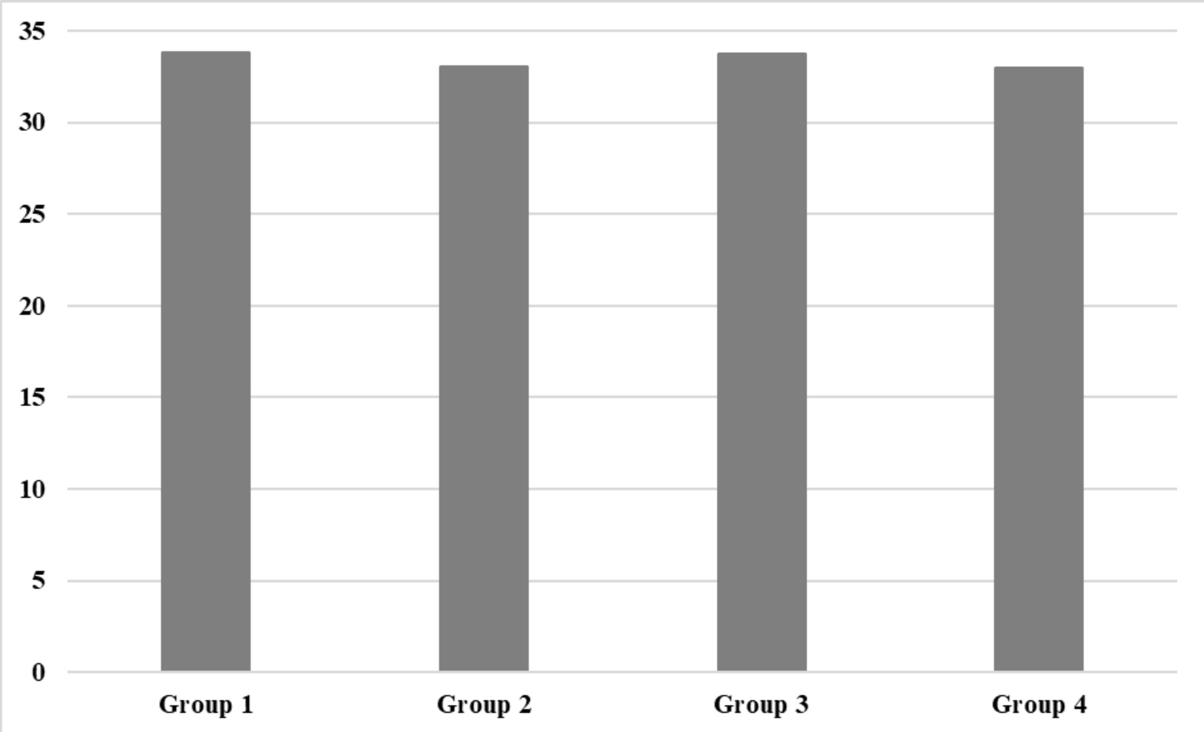

Figure 2. Average response times in seconds

Note. Experimental groups: 1) PC and decremental, 2) PC and incremental, 3) smartphone and decremental, and 4) smartphone and incremental. The result of a Kruskal-Wallis test indicates no significant differences.