

A comparison of question order effects on item-by-item and grid formats: visual layout matters

Stefkovics, Ádám; Kmetty, Zoltán

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:

Stefkovics, Á., & Kmetty, Z. (2022). A comparison of question order effects on item-by-item and grid formats: visual layout matters. *Measurement Instruments for the Social Sciences*, 4, 1-12. <https://doi.org/10.1186/s42409-022-00036-z>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:

<https://creativecommons.org/licenses/by/4.0>

ADVANCES IN METHODOLOGY

Open Access



A comparison of question order effects on item-by-item and grid formats: visual layout matters

Ádám Stefkovics^{1,2*}  and Zoltán Kmetty^{1,3}

Abstract

Question order effect refers to the phenomenon that previous questions may affect the cognitive response process and respondents' answers. Previous questions generate a context or frame in which questions are interpreted. At the same time, in online surveys, the visual design may also shift responses. Past empirical research has yielded considerable evidence supporting the impact of question order on measurement, but few studies have investigated how question order effects vary with the visual design. Our main research question was whether question order effects are different on item-by-item formats compared to grid formats. The study uses data from an online survey experiment conducted on a non-probability-based online panel in Hungary, in 2019. We used the welfare-related questions of the 8th wave of ESS. We manipulated the questionnaire by changing the position of a question that calls forth negative stereotypes about such social benefits and services. We further varied the visual design by presenting the questions in separate pages (item-by-item) or one grid. The results show that placing the priming questions right before the target item significantly changed respondents' attitudes in a negative way, but the effect was significant only when questions were presented on separate pages. A possible reason behind this finding may be that respondents engage in a deeper cognition when questions are presented separately. On the other hand, the grid format was robust against question order, in addition, we found little evidence of stronger satisficing on grids. The findings highlight that mixing item-by-item and grids formats in online surveys may introduce measurement inequivalence, especially when question order effects are expected.

Keywords: Online survey, Question order, Visual design, Questionnaire, ESS

Introduction

Question order effects or context effects refer to the idea that earlier items in a questionnaire can affect later responses (Suchman and Presser, 1981). For instance, earlier items "prime" certain beliefs or values, making it

easier for the respondents to access and retrieve these beliefs when answering later questions. Such effects are expected to be large in attitude surveys (Tourangeau et al., 2003). Various studies since the eighties have demonstrated that the size of question order effects may further depend on the topic of the questionnaire (Tourangeau et al., 2000), the respondents' characteristics (Tourangeau et al., 1989a, 1989b), or interviewer behavior (Snidero et al., 2009). Fewer studies, however, investigated how question order effects are related to the design features of a questionnaire.

As web surveys are increasingly used, the question of how the visual layout of the questionnaire impacts

The experiment was pre-registered before the fieldwork. It is available at the following link: <https://osf.io/3qbgm>

*Correspondence: adam.stefkovics@tatk.elte.hu

² Institute for Quantitative Social Sciences, Harvard University, Cambridge, USA

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

responses has attracted much attention in recent decades (Stern et al., 2007). One key decision in attitude surveys is whether to present a set of attitude questions in item-by-item or a grid or matrix format (Revilla and Couper, 2017). For instance, on mobiles or smartphones, the use of grids is problematic. Although the debate over the choice between the two design features remains open, most of the previous evidence suggests that the two formats do not lend the same results (for a review see e.g. Couper et al. (2017) or Revilla and Couper (2017)). Respondents, for instance, tend to spend less time answering questions and yield more correlated responses in a grid format. That may be due to the differences in the response process. First, in item-by-item formats, respondents process questions separately, which may result in deeper cognition. Second, some authors have argued that the cognitive load required may be higher for grid formats, as questions presented in a grid format can be more challenging (Liu and Cernat, 2018).

Anyhow, if differences arise in processing, the ordering of the questions may count in one format, but not in another. The fact that grids lead to higher intra-item correlations suggests that answers may be influenced by item order more strongly in grid formats compared to item-by-item formats. On the contrary, it is also possible that a deeper cognition in item-by-item format increases the “power” of priming and the likelihood of retrieving and accessing attitudes. Due to the lack of empirical evidence, however, these assumptions on the relationship between question order effects and visual layout are rather speculative. We argue that in the era when survey modes (De Leeuw, 2018), devices (Toepoel and Ludtigit, 2015), and visual layouts (Stern et al., 2007) are routinely mixed, understanding the effect of “mixing” on data quality is crucial.

To address this research gap, we collected online experimental data in Hungary. In a 3x2 design, we manipulated the position of a standard ESS question recalling negative stereotypes on social benefits and services both in item-by-item and grid format, to determine whether it would shift responses to a question about the efficiency of welfare services. We further investigated if the effect remains when an intervening item is placed between the priming and the target item.

Background and literature review

Question order or context effects in attitude surveys have been extensively researched in previous reports. Classical experiments in the early eighties have investigated the role of preceding questions (Schuman et al., 1981; Schuman and Ludwig, 1983; Schuman and Presser, 1981). Earlier questions may shift responses for the later question, for instance, by *priming* a context (Tourangeau

et al., 2000). As a result of priming, certain beliefs, values, and attitudes become more accessible when answering the following questions. According to the accessibility hypothesis (Bishop et al., 1982; Tourangeau et al., 1989a, 1989b), answers to attitude questions are not based on a systematic pool of relevant beliefs, instead, respondents quickly sample from beliefs. This sampling occurs under time pressure and is often influenced by several factors, for instance, by beliefs retrieved in an earlier question.

The retrieved information, however, is not always used in the evaluation of the next items. Wänke and Schwarz (1997) differentiates between information accessibility and information use. They argue that “if respondents are aware of a possible influence and consider it inappropriate, they will not use the accessible information” (p. 8.). When a specific question precedes a general question (part/whole questions), respondents often include the information gained in the specific question to answer the general question (assimilation effect, [Schwarz et al., 1991; Schwarz and Bless, 1991]). On the other hand, when a general question comes first, and questions are redundant, respondents tend to exclude previously gathered information and refer to other aspects of the question (contrast effect). When questions are on a similar level (part-part), previous judgments may serve as standards for later comparisons (judgmental or perceptual contrast (Tourangeau et al., 2000)). Such biased comparisons have been found, for instance, in the study of Silber et al. (2016), where respondents’ evaluated the European Union more positively, when it was not preceded by the evaluation of their home county (Germany). Earlier questions also provide an interpretative framework. When interpreting questions, previously answered items provide clues for unclear terms (Tourangeau et al., 2003).

Welfare attitudes have been shown to be sensitive to question order in earlier studies. Tourangeau et al. (1989a, 1989b) found that perceptions on the government’s welfare spending were influenced by priming pro- or anti-welfare attitudes. In the study of Schwarz and Hippler (1995), drawing attention to welfare spending influenced the willingness to donate to Russia. Thau et al. (2021) in a part/whole split ballot experiment found that overall satisfaction with public services increased if asked after specific service ratings. In another experiment, Nielsen and Kjær (2011) reported that participants’ willingness to pay for two potential policies that could reduce air pollution and thus increase life expectancy was subject to question order.

Although the mentioned cognitive processes underlying question order effects have been verified by several experiments, few of them have been observed “naturally” (Tourangeau et al., 2003), and some studies reported null results (e.g. Smith, 1988). In their recent study, Stark

et al. (2020) aimed to come to generalized conclusions on question order effects by replicating two classical experiments in the United States and 11 other countries. In contrast to another study (Klein et al., 2014), the authors found weak evidence of generalization of question order effects. One reason why question order effects are hard to generalize is because they are complex and depend on a number of features. For instance, cultural and country differences play a key role. In the study of Stark et al. (2020) question order effects were stronger in more individualistic countries than in more collectivistic countries. The topic and the related underlying values and attitudes may determine why question order matters a lot in some countries and not in others.

In self-administered surveys, question order effects may be dependent on the visual layout of the question. Specifically, in a set of attitude questions, researchers may choose to use an item-by-item format or a grid or matrix format. In item-by-item formats, questions are asked separately (either on one page or separate pages), whereas, in grid formats, questions (or items) are grouped in a matrix. A long line of research investigated measurement equivalence between the two formats. Many studies make such comparisons with device type, as the use of grids in mobile and smartphones can be even more problematic. The first such experiment by Couper et al. (2001) found higher intra-item correlations in the grid format with lower completion time compared to the item-by-item format. Other studies endorsed these findings (Iglesias et al., 2001; Roßmann et al., 2018; Tourangeau et al., 2004). Liu and Cernat (2018) found that the two formats yield the same results when the response options are below 7 options, but reveal differences when longer scales are used. Similar evidence has been found in the study of Mavletova et al. (2018). Research has also shown that item nonresponse (Couper, 2016; Iglesias et al., 2001; Liu and Cernat, 2018; Revilla and Couper, 2017; Toepoel et al., 2009), straightlining and nondifferentiation (Mavletova et al., 2018; Revilla and Couper, 2017; Roßmann et al., 2018; Tourangeau et al., 2004) is typically higher in grid formats. Research also suggests that respondents favor item-by-item formats or smaller grids above large grids (Grandmont et al., 2010; Thorndike et al., 2009; Toepoel et al., 2009). Nonetheless, we are aware of several studies reporting no differences between the formats (Bell et al., 2001; Callegaro et al., 2009; Höhne et al., 2017; Peterson et al., 2017; Thomas et al., 2015; Thorndike et al., 2009) or even opposing results to the above (McClain and Crawford, 2013).

Altogether, evidence on comparisons between item-by-item and grid formats remains mixed. Results on the common hypothesis about grids on smartphones are associated with lower data quality are also mixed (Revilla

and Couper, 2017). As argued by Revilla and Couper et al. (2017) this may be due to the wide range of details of visual layout that can influence responses. Grids can be very diverse in their visual appearance. Responses may be affected by the number of items (Grady et al., 2019; Mavletova et al., 2018; Toepoel et al., 2009), the length of the scale (Liu and Cernat, 2018; Mavletova et al., 2018), the labeling, the size of the grid, the horizontal-vertical layout (Revilla and Couper, 2017), and more. Clearly, no single study can involve all the possible combinations.

Research on how answering item-by-item and grid formats are different concerning the underlying mechanisms and required cognitive load is scarce. Based on the *cognitive load theory* (Sweller, 1988), cognitive load is the required effort of working memory when facing a certain task. More specifically, extrinsic cognitive load refers to how the task is visually presented, which seems relevant in this regard. It can be argued that grid formats require more extrinsic cognitive load from the respondents. According to Liu and Cernat (2018) “the way that a matrix presents questions may make it more daunting than an item-by-item question” (p. 4.). Similarly, Couper et al. (2013) suggested that grids *look* more complex.

Now, when cognitive load is high, respondents are likely to *satisfice* (Krosnick, 1991), by, for instance, skipping some of the steps of the response process or paying less attention to the questions. The fairly consistent finding of the previous literature on faster completion time in grid formats can be an indication of such behavior (Höhne et al., 2017). Grid formats provide ground for “participants engaging in less critical processing of survey questions and their own responses” (Grady et al., 2019, p. 2). Although, fast responding can also indicate strong opinions or an easy response task (Revilla and Couper, 2017). Respondents tend to give more positive ratings to item-by-item designs or shorter grids compared to large matrixes, which questions the assumption that answering grids would be easier (Thorndike et al., 2009; Toepoel et al., 2009).

Moreover, item-by-item formats may elicit *deeper cognition*. Using the *levels of processing* framework of Craik and Lockhart (1972), the amount of information retrieved from memory to answer a survey question is a function of the level of processing, and the way the information is encoded (e.g. reading questions thoroughly, mapping response options etc.). In item-by-item formats, respondents are required to spend time and consider each question separately, which may increase the cognitive effort spent on each question and reduce satisfying strategies. Disentangling high cognitive load and deeper cognition, however, can be challenging, since long time spent with a question may indicate both high cognitive load and deep cognition.

What do these considerations imply on question order? First, the manifestation of context effects can be subject to how well the priming question is processed. It seems plausible to assume that the deeper the information processing and the lower the cognitive load is, the more likely is to retrieve and access beliefs that are primed in the question. Hence, a deeper cognition in item-by-item formats may lead to stronger priming effects and thus, stronger question order effects.

At the same time, order effects can be strong in grid formats as well. According to the *interpretive heuristic* suggested by Tourangeau et al. (2004) respondents follow five heuristics when interpreting the visual layout: middle means typical; left and top mean first; near means related; up means good; and like means close. Especially, near means related heuristic has been assumed to affect responses in grid formats (Roßmann et al., 2018; Silber et al., 2018). In a grid format, questions are presented close to each other, thus, they may be interpreted as conceptually related. That, for instance, may lead to higher inter-item correlations, more straightlining, and nondifferentiation compared to item-by-item designs. Turning to question (or item) order, such behavior can easily lead to question order effects, but due to another heuristic compared to the one assumed in item-by-item formats. In grids, there is no reason to expect strong priming effects, but respondents may still be influenced by earlier questions due to spatial proximity.

We are aware of one study that tested the interaction of question order and question layout, however, in a telephone survey. Tourangeau et al. (1989a, 1989b) in a re-contact design investigated the effect of priming certain beliefs in four different issues. They further varied the mode of presentation: items presented in blocks or presented separately. When items were presented separately, issues were intermixed with one another, whereas in the block version, context and target items were presented in one block per issue, with the target item always coming as the last. Overall, they found that question order effects were significantly larger when related items were presented in a block than when they were scattered among other issues. This piece of evidence suggests that priming effects can be strong in grids as well. Although, in this study, question layout was only perceived by the interviewers. From the respondents' perspective layout instead, was a matter of whether items related to each other were asked in one block, after one another, or intermixed with other issues.

The current study and hypotheses

Measures

In this study, we conducted an online survey experiment (Mutz, 2011) to assess how question order effects

interact with question format. The welfare module of the 8th wave of the European Social Survey (ESS) was used in this study (ESS, 2016). Other, unrelated question blocks preceded these questions, manipulated questions were towards the second part of the questionnaire. Previous questions were in various public issues.

Our target item (E10) asked respondents “to what extent they agree or disagree that social benefits and services in Hungary prevent widespread poverty”.¹ In the original ESS questionnaire E10 question was placed in a 4-question grid format table:

What extent you agree or disagree that social benefits and services in [country]...

E9: ...place too great a strain on the economy?

E10: ...prevent widespread poverty?

E11: ...lead to a more equal society?

E12: ...cost businesses too much in taxes and charges?

The previous E8 question in the original question order is quite neutral as it asks about the government's responsibility for ensuring sufficient child care services.

We used E13 as our priming item, where respondents were asked to indicate “to what extent do they agree or disagree that social benefits and services in Hungary make people lazy”. With this question, we made the aspect of people's work motivations when receiving social benefits and services and called forth further negative stereotypes towards people receiving social benefits and services.

E9 was used as an intervening item. We choose E9 because this item introduces another context to judge social benefits and services, namely the country-level economic impacts.

All these items were presented with a five-point Likert scale (Agree strongly to Disagree strongly). All respondents were offered a *Refused* and *Don't know* button.

Experimental design

Respondents were randomly assigned to one of six experimental groups right before the welfare module. Our treatments varied the order of the items and the visual layout of the items. Question order treatment had three outcomes: respondents either received the priming item in its original place (after the target item), or right before the target item, or before the target item, but with the intervening item in between. Based on the visual treatment respondents received each item on a single page

¹ The original question was worded as follows: “Using this card, please tell me to what extent you agree or disagree that social benefits and services in [country]...” We dropped “using this card” to fit the online administration of the survey.

Table 1 Experimental design

	Group1 (ESS - control)	Group2	Group3	Group4	Group5	Group6
Question order treatment	priming after target	priming before intervening	priming before target	priming after target	priming before intervening	priming before target
Visual Treatment	Grid	Grid	Grid	Item-by-item	Item-by-item	Item-by-item
N	144	140	133	138	135	135

(item-by-item design) or all items (E9-E14) in one grid on one page. This resulted in a 3x2 factorial design. Table 1 provides a description of the six groups. We considered Group 1 as the control group because this version was the most similar to the original ESS version. In fact, ESS grouped these questions into three different blocks².

In our design decisions, we aimed to follow the ESS format. In both formats, scales were presented horizontally, and they were fully labeled. Grids contained six or seven items. In Appendix A1 we provided example screenshots of the different versions of the questionnaire.

Hypotheses

Drawing on the literature of priming effects (Tourangeau et al., 2000), and empirical findings suggesting the importance of question order in the measurement of well-being related questions (Nielsen and Kjær, 2011; Schwarz and Hippler, 1995; Thau et al., 2021; Tourangeau et al., 1989a, 1989b), we expect that our main priming item (E13) will shift responses in a negative way, namely, respondents who receive the priming item before the target item, will be less likely to consider welfare services efficient. The priming item can increase the accessibility (Bishop et al., 1982; Tourangeau et al., 1989a, 1989b) of possible negative aspects of providing social benefits, therefore the likelihood of engaging in a more critical assessment of welfare services is assumed in such cases. Admittedly, how people use accessed beliefs (Wänke and Schwarz, 1997) can be a function of attitude strength and the original opinion of the respondent about the priming question. We expect that the priming question will have a clear negative effect on the opinion of those, who do not hold a strong position in the laziness claim, who tend to agree with the laziness claim, and it may have no effect or even in some cases a positive effect on those who oppose the laziness claim.

H1: When the priming item is placed before the target item, the means of the target item will be higher (lower agreement with social benefits and services in Hungary prevent widespread poverty)

Several authors suggested that question order effects can be reduced by placing intervening (or buffer) items between the priming and the target item (Cantril, 1944; Schwarz and Schuman, 1997; Tourangeau et al., 1989a, 1989b; Wänke and Schwarz, 1997). In the classic study of Cantril (1944) the question that preceded the endorsement of U.S. citizens serving in the French or British army had an influence on the responses, but this order effect was reduced when the two items were separated with several buffer items. Schwarz and Schuman (1997) found that even one intervening item can change the context and the effect of preceding questions. Intervening items increase the distance between the primed beliefs and the target question and make other unrelated contexts, and beliefs accessible (Wänke and Schwarz, 1997). Depending on various factors intervening items may strengthen, weaken or even reverse priming effects. Our intervening item called forth a different aspect of the impact of social benefits and services: the economic system. We expected that judging the macroeconomic impact of social benefits and services will weaken the individualistic context primed by the original priming item. The intervening item, however, can still reinforce the negative effect of the priming item, since both items were negative. As we found no correlation between the intervening item and the target item in the 8th round of the Hungarian ESS, we rather expected that the priming effect would be stronger when the priming item was right before the target item compared to the design where the intervening item was placed between them.

H2: When the priming item is placed before the target item, with E9 as an intervening item mean differences stated in H1 will decrease or diminish.

Concerning visual design, our investigation is fairly explorative. Drawing on the literature, both formats may provide ground for question order effects. The item-by-item format may let respondents engage in a deeper cognition, which can lead to more prevalent beliefs primed by the earlier item. On the other hand, in grid formats, people may respond similarly due to spatial proximity. One could even assume that in grid formats the combination of priming effects and spatial proximity may add up and thus, lead to stronger question order effects.

² See the original questionnaire here: <https://www.europeansocialsurvey.org/data/download.html?r=8>

Nonetheless, we expect similar order effects on both formats, namely that the direction and size of the order effects will not be significantly different, both regarding H1 and H2.

H3a: When the priming item is placed before the target item, the means of the target item will be lower on both item-by-item and grid formats (lower agreement with social benefits and services in Hungary prevent widespread poverty).

H3b: When the priming item is placed before the target item, with E9 as an intervening item mean differences stated in H1 will decrease or diminish both on item-by-item and grid formats.

Finally, it has been widely suggested that question order effects are stronger when respondents do not have a strong opinion on the topic, or questions about an unfamiliar issue (Tourangeau et al., 2000). When respondents are in doubt or uncertain, they are more likely to look for cues in earlier questions (Moon et al., 2019; Tourangeau et al., 2003). We did not have direct measures of attitude strength or knowledge, we used political interest and educational level as proxies. Those with higher levels of political interest have been consistently found to hold a stronger opinion on public issues and more willing to express their opinion (e.g. Baldassare and Katz, 1996; Salmon and Neuwirth, 1990). Higher levels of education also predict stronger attitude strength and higher knowledge in public issues (Billiet et al., 2004; Krosnick and Abelson, 1992). We expect to find stronger priming effects among respondents with a lower level of political interest or lower level of education. Furthermore, as we assumed that question order effects may be subject to how well the questions are processed, we hypothesize that priming effects are associated with the conscientiousness of the respondents. Personality traits such as conscientiousness have been shown to be related to survey participation and data quality. People with higher conscientiousness were found to be more likely to participate in surveys (Cheng et al., 2020; Rogelberg et al., 2003), less likely to drop out (Brüggen and Dholakia, 2010; Nestler et al., 2015), and provide responses with higher quality (Brüggen and Dholakia, 2010). That may be because conscious individuals tend to be more devoted to tasks, more attentive, persistent, and careful (Roberts et al., 2005). Drawing on these findings, higher levels of conscientiousness may result in deeper cognition and less satisficing when answering survey questions.

H4: The effect stated in H1 will diminish among respondents with a higher level of political interest, higher level of education, or higher level of conscientiousness, and increase among respondents with a lower level of political interest, lower level of education, or lower level of conscientiousness.

tiousness, and increase among respondents with a lower level of political interest, lower level of education, or lower level of conscientiousness.

Although question order differences between the two formats were the main focus of our study, we intended to detect signs of measurement error on each format. As we discussed in our literature review, grids are often found to invite satisficing behavior. It has also been argued that such behavior may interact with the impact of question order. This analysis may help us to understand the underlying mechanisms better. Specifically, we expected to find higher item-nonresponse rates and higher non-differentiation on grids compared to item-by-item formats. We did not expect that question order would be related to such measurement error. We are aware that the two indicators used in this study were not able to capture the complexity of data quality. Still, we believe that these measures may be able to uncover if major differences arise between the two formats.

H5: Item-nonresponse rates of the target item will be higher on grids as compared to the item-by-item design

H6: Non-differentiation among the three questions (priming, target, intervening) will be higher on grids as compared to the item-by-item design

Data and analytical plan

Data collection

We draw on data collected in Hungary from a non-probability-based access panel by NetPanel of NRC Ltd. (<https://nrc.hu/netpanel/>), consisting of more than 140.000 internet users.³ Compared to general population statistics individuals with a high level of education and individuals from bigger cities are overrepresented in the panel. We used a quota sampling method (with gender, age, and region quotas) to ensure equal representation. The total sample size was 1100 and it is representative of the Hungarian internet users aged between 18 and 65. Our experimental approach ensured that the internal validity of our results remains high, the use of such non-probability-based data may limit the generalizability of our findings. The fieldwork was carried out at the end of 2018 between November 7 and November 22.

The questionnaire was optimized for users answering on mobile phones so that they did not get the grid

³ Subscribing for the panel is voluntary, panel recruitment occurs via advertising and panelists' word-of-mouth enrollment. By filling out a questionnaire, members normally get access to sweepstakes but do not necessarily receive prizes.

format, only the item-by-item format. Therefore, in their sub-sample, our experimental design was broken. After careful consideration, we decided to omit them from the analysis. Two hundred seventy-five respondents filled out the survey on a small screen device. The small screen respondents were overrepresented within the younger age cohort, people living in a small settlement, females, and lower educated people. We applied weights to adjust for the biased sample composition due to the exclusion of the mobile respondents, using an iterative weighting procedure. In the weighting process, we used gender, education level, age, and settlement type (capital city, county town, city, village).

Analytical plan

We test our hypothesis (H1-H4) with regression models fitted on the target variable E10. In these models, we included the treatment as independent variable. We used the original ESS design (priming after target in grid format) as the reference.

To test H4, these models were refitted with interactions between key demographic variables (educational level, political interest, and the conscientiousness psychological trait variable) and the treatment variable.

We measured education with a three-category variable (from “primary” to “university diploma”). Political interest was measured with an ordinal variable ranging from 1 (not interested) to 5 (very interested). We added one psychological trait variable to our models, the conscientiousness of the respondent. For that we used the short, 15-item version of Big Five Inventory (BFI-15; Benet-Martínez and John, 1998; O. John et al., 1991; O. P. John et al., 2008). Here we used an ad-hoc translation of the items which we had already used in our previous studies. We applied principal component analysis to calculate the trait variable with the related three items. We used a regression approach to extract the latent dimension. The factor score of the three items was the following:

- I’m thorough at work: -0.83
- I tend to be careless: -0.53
- I do things efficiently: 0.85

The reversed item had a much lower factor score, which might be the consequence of a straightlining in this block. A higher value means higher conscientiousness.

We included the questionnaire length as an additional covariate in the model in the pre-registration. Unfortunately, we only have the fill-out time of the complete questionnaire, not just the block we use in this analysis. After careful consideration, we decided to omit this variable from the models of this paper.

To test H5 we created a dummy variable to measure item non-response in the E10 question. We coded the valid answers to zero and the NA’s to one. To test H6, non-differentiation was measured also with a dummy variable. If a respondent answered with the same value for E9 to E14 we coded the variable to one, otherwise zero. To test the effect of treatment on item-nonresponse and the non-differentiation we fitted binomial logistic regressions.

Results

Effects of the treatments

We tested our H1-H4 hypothesis with OLS regression models. The dependent variable was E10 as described before, which measured – *to what extent people agree or disagree that social benefits and services in Hungary prevent widespread poverty [1: strongly agree – 5: Disagree strongly]*. In the case of the treatment variable, the ESS design was the reference. We also ran an additional simple variance analysis between the dependent variable and the treatment variable. The mean value of the target item was 3.7 and the standard deviation was 1.2 (see Figure A3 in the appendix).

The regression model was significant; the adjusted R^2 value was 0.06 see Table 2. Based on the results two treatment groups differed significantly from the control group, the one where we placed the priming item before the target item in the item-by-item design and the one where we used an intervening item in the item-by-item design. This result partly confirms our H1 hypothesis. Priming matters, but not on both layouts. We expected a similar effect with different layouts, but we found no effect on the grid design and found a significant effect on the item-by-item design. Based on this we can reject H3a-b hypotheses. We also have to reject H2 hypothesis, as we did not find differences in groups with or without an intervening item. We performed a power analysis to check the reliability of the results. The power level for the treatment variable was 0.94. On the other hand, the power level of education was just 0.12, so the results of education effect have to be treated with caution.

We ran an ordered-logit model with the same variables for a robustness check. This model also confirmed all the previous results (see Table A4 in the appendix).

In H4 we hypothesized that the effect stated in H1 would be stronger among respondents with a lower level of political interest, lower level of education, or lower conscientiousness, respectively. We tested how treatment effects differ within different groups of respondents. All the three independent variables had significant interaction with the treatment variable (see the models in the Appendix). We measured the lowest power level in the case of interaction of treatment and political interest

Table 2 Results of the linear regression fitted on the target E10 item – to what extent they agree or disagree that social benefits and services in Hungary prevent widespread poverty

	Unstandardized coefficient	2.5% Confidence interval	97.5% Confidence interval	p	BETA
(Intercept)	3.85	3.58	4.12	0.00	
Priming before intervening - grid	-0.02	-0.28	0.23	0.86	-0.01
Priming before target - grid	-0.00	-0.27	0.27	0.99	-0.00
No priming - separate	-0.02	-0.29	0.24	0.87	-0.01
Priming before intervening - separate	0.34	0.08	0.59	0.01	0.09
Priming before target - separate	0.38	0.11	0.64	0.01	0.10
Education - middle	0.19	0.01	0.36	0.04	0.07
Education - higher	0.30	0.08	0.52	0.01	0.09
Political Interest	-0.15	-0.21	-0.08	0.00	-0.15
Conscientiousness	-0.11	-0.18	-0.03	0.01	-0.10
N	778				
Adjusted R ²	0.06				

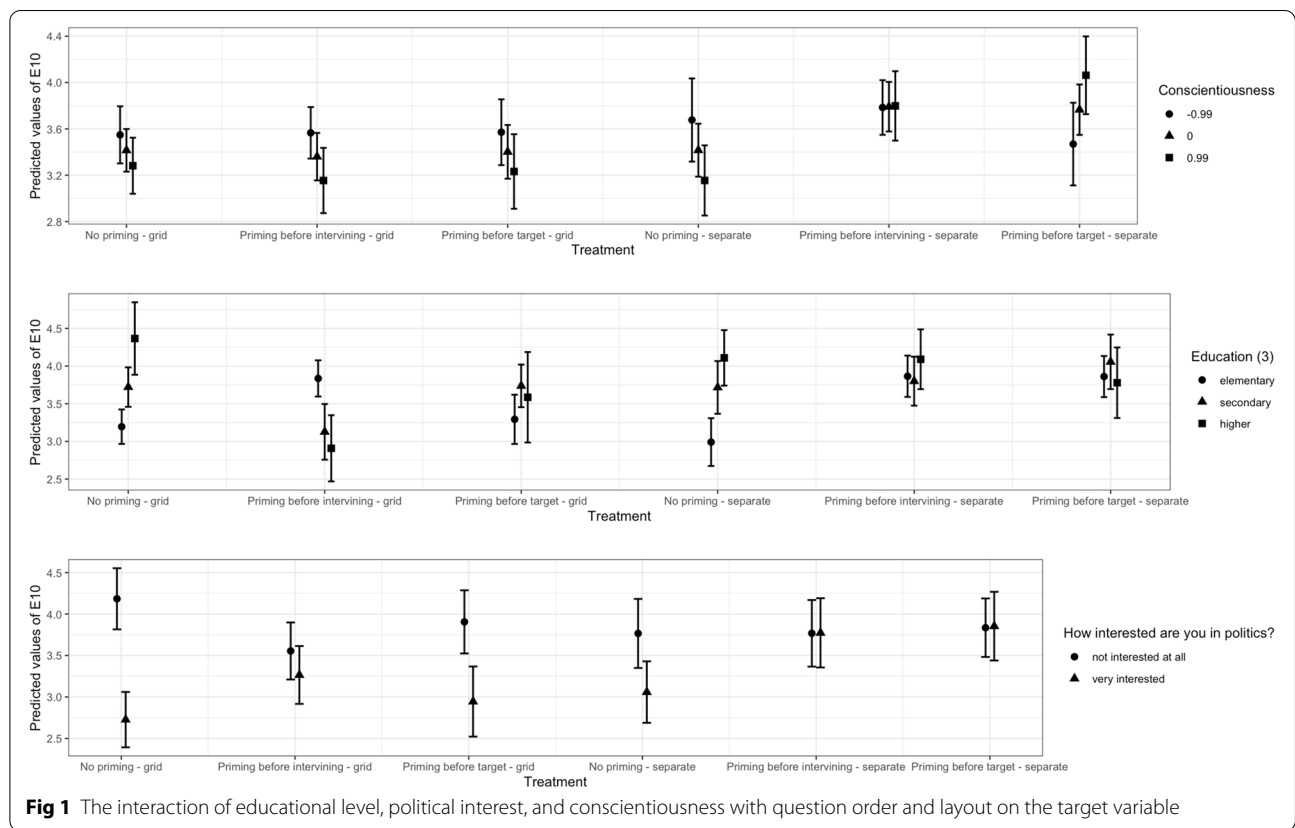


Fig 1 The interaction of educational level, political interest, and conscientiousness with question order and layout on the target variable

– 0.59. The power analysis result for the conscientiousness interaction was 0.62 and for the education interaction 0.77. We plotted the marginal effects for a more straightforward interpretation of the interaction terms (see Fig. 1). In the case of conscientiousness we plot the predicted value at the mean level, and \pm one standard

deviation from the mean. The effect of question order and layout was not significant in the case of people with low conscientiousness and low political interest. On the other hand, the higher the political interest was, the higher the mean of the target item after changing the question order, especially in the item-by-item layout

Table 3 Results of the binomial logistic regression models fitted on item-nonresponse

	Unstandardized coefficient	S.E.	p
(Intercept)	-1.78	0.47	0.00
Layout (0=item-by-item; 1=grid)	0.66	0.36	0.07
Education - middle	-0.71	0.40	0.08
Education - higher	-0.64	0.56	0.25
Political Interest	-0.49	0.15	0.00
Conscientiousness	0.17	0.16	0.29
N	810		
Pseudo R ² - McFadden	0.08		
BIC	302		

design. We detected the same effect among respondents with higher conscientiousness. However, results related to the educational level were contradictory. Among individuals with primary education, the mean value of the target item increased (lower agreement with social benefits and services in Hungary prevent widespread poverty) when we moved the priming item before the target item, regardless of the layout. Based on the models, the results were mixed regarding the effect of the treatment among different groups of respondents. Therefore, we reject the H4 hypothesis.

Data quality in the two layouts

To assess measurement error on the two layouts, we checked whether layout and question order affected item-nonresponse rates or non-differentiation rates. Overall 3.7 percent of the responses were in the NA category in the target question. We fitted a binomial logistic regression model to measure the treatment effect on the item nonresponse see Table 3.

From the demographic variables, gender, political interest, and education level affected the item-nonresponse. Individuals with lower political interest and lower education had a higher chance of skipping the target question. We found that the layout was not significant at the conventional 0.05 level.

Turning to non-differentiation, 9.5 percent of the respondents gave the same answer for all six questions. We fitted a similar binomial logistic regression model to measure the layout effect on non-differentiation see Table 4.

Respondents with lower conscientiousness had a higher probability of answering the same way for the six questions. This result confirms previous studies regarding satisficing. The layout was not significant in the model. Based on the results we could reject both H5 and H6 hypotheses.

Table 4 Results of the binomial logistic regression models fitted on non-differentiation

	Unstandardized coefficient	S.E.	p
(Intercept)	-1.97	0.38	0.00
Layout (0=item-by-item; 1=grid)	0.39	0.27	0.15
Education - middle	0.44	0.27	0.11
Education - higher	-0.54	0.47	0.25
Political Interest	-0.19	0.10	0.07
Conscientiousness	-0.29	0.12	0.01
N	692		
Pseudo R ² - McFadden	0.05		
BIC	406		

Discussion

This study was the first to examine the relationship between question order effects and the visual layout of the items in an experimental setting in a web survey. We manipulated the order of a set of standard welfare-related attitude questions of the ESS. We further varied whether items were presented in an item-by-item or a grid format. When our priming item was placed right before the target item, and items were presented separately on single pages, the priming item did shift responses in a negative way. We did not find a significant effect of question order in grid formats.

A possible reason behind the observed question order effects on the item-by-item format may be that respondents engage in a deeper cognition when the questions are presented on separate pages. We found a significant interaction effect between political interest and conscientiousness, indicating that those with a higher level of political interest or higher conscientiousness were more sensitive to question order effects. As high political interest is often attached to stronger attitudes, this result contradicts our expectations but resonates well with our main finding. Respondents with higher levels of political interest or higher conscientiousness may be more motivated to carefully read the questions (especially in item-by-item formats) which may result in stronger priming effects in this sub-sample despite their possible stronger attitudes or higher knowledge of the topic. On the other hand, the lack of question order effects among respondents with low political interest or lower levels of education may be due to poorly processed answers in general.

On the contrary, the grid format was not susceptible to question order effects. The hypothesis of spatial proximity causing question order effects was not supported. It may indicate that respondents put less effort into optimizing their answers in grids, and due to that, the previous item failed to prime a strong context. This seemed

plausible, as our items were reverse coded, in which case careless responding would even be more likely to weaken priming effects. Our additional analysis, however, does not provide strong evidence for that. Respondents were just as likely to skip questions and non-differentiate in the grid format as in the item-by-item format.

In line with several previous studies (Mavletova et al., 2018; Revilla et al., 2017; Roßmann et al., 2018), we observed measurement inequivalence between the two formats, which raises concerns about their simultaneous use in one survey design. In mixed-device web surveys, the visual layout is routinely optimized to screen size, for instance, some respondents receive an item in an item-by-item format, while others receive it in a grid, due to optimization. As our results indicate, in such designs, the order of the questions will cause measurement error in some cases, but not in other cases.

The solution, however, is not straightforward. The results of Mavletova et al. (2018) highlight that the use of grids without optimization on mobiles may lead to lower data quality and respondent satisfaction. If strong question order effects are expected, the use of grids on both devices could be considered. This, on one hand, would reduce measurement inequivalence, but on the other hand, it could also invite less optimized responses (Roßmann et al., 2018). Previous authors advise the use of smaller grids (maximum 5 items) and scale not longer than 5–7 scale points. One may also consider optimizing first for mobile respondents and applying that format to all devices. While this approach can secure measurement equivalence, some formats may not be feasible for larger devices, and one would miss the potential of the larger screen size.

One limitation of our study is that an opt-in online panel was used. This may weaken the generalizability of our results, as panel respondents are known to be more trained and motivated than an average respondent. This may provide explanations for the absence of higher satisficing on the grid format, and suggest that effect sizes would have been stronger among the general population. Furthermore, as found by Stark et al. (2020), question order effects may strongly vary with countries or cultures, which questions the generalizability of the findings based on a Hungarian sample. Although it is hard to determine in what sense the Hungarian population is special regarding the current experiment. Hungary was among the countries where disagreement with the target item was the highest (social benefits and services do not prevent poverty), and where agreement with the priming item was above average in the 8th round of ESS. These may imply that priming effects were higher in Hungary than they would have been in most of the other European

countries. Another limitation is that we were unable to provide insights into how these patterns may play out among mobile respondents, future research should experiment with different devices. Further research is required to better understand the differential cognitive mechanisms of the response process on the two formats. This may include eye-tracking analysis and qualitative assessment.

Our study provided additional evidence on two key design features of online questionnaires: the order and the visual layout of the questions. We showed that question order strongly matters in the measurement of welfare attitudes and that item-by-item and grid formats tend to yield different results. The findings underscore the importance of questionnaire design choices and draw attention to the potential measurement errors caused by mixing different visual layouts. The study provided some guidelines for the ESS survey as well. We argue that attention should be paid to question order effects in the future use of the well-being questions in the ESS, and carefully designed visual layouts are advised in the online measurement of these attitudes (e.g. in the Cross-National Online Survey [CRONOS] panel).

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s42409-022-00036-z>.

Additional file 1.

Acknowledgements

We thank Júlia Koltai for helping in designing the experiment.

Authors' contributions

Both authors contributed to the designing of the survey and were responsible for the data collection. The first author wrote the theoretical background, the current study and data and methods part of the manuscript. The second author was responsible for the analysis and the writing of the results section. Both authors contributed to the introduction and discussion sections. The authors read and approved the final manuscript.

Funding

Open access funding provided by Eötvös Loránd University. This work was supported by the Hungarian Research Fund (NKFIH / OTKA, grant number KI20070). The work of Zoltán Kmetty was funded by the Premium Postdoctoral Grant of the Hungarian Academy of Sciences.

Availability of data and materials

The dataset used in this article is available in the repository of the Hungarian Centre for Social Sciences Research Documentation Center, under the following link: https://openarchive.tk.mta.hu/378/22/mixed18-5_JB-378.zip

Competing interests

No potential conflict of interest was reported by the author(s).

Author details

¹Faculty of Social Sciences, Eötvös Loránd University, Budapest, Hungary.

²Institute for Quantitative Social Sciences, Harvard University, Cambridge, USA.

³Centre for Social Sciences, CSS-Recens research group, Budapest, Hungary.

Received: 5 January 2022 Accepted: 31 May 2022
Published online: 21 June 2022

References

- Baldassare, M., and Katz, C. (1996). Measures of Attitude Strength as Predictors of Willingness to Speak to the Media. *Journalism and Mass Communication Quarterly*, 73(1), 147–158.
- Bell, D. S., Mangione, C. M., and Kahn Jr., C. E. (2001). Randomized Testing of Alternative Survey Formats Using Anonymous Volunteers on the World Wide Web. *Journal of the American Medical Informatics Association*, 8(6), 616–620. <https://doi.org/10.1136/jamia.2001.0080616>.
- Benet-Martínez, V., and John, O. P. (1998). Los Cinco Grandes across cultures and ethnic groups: Multitrait-multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology*, 75(3), 729.
- Billiet, J., Swyngedouw, M., and Waeghe, H. (2004). Attitude strength and response stability of a quasi-balanced political alienation scale in a panel study. In W. E. Saris, and P. M. Sniderman (Eds.), *Studies in Public Opinion*, (pp. 268–294). Princeton University Press.
- Bishop, G. F., Oldendick, R. W., and Tuchfarber, A. J. (1982). Political information processing: Question order and context effects. *Political Behavior*, 4(2), 177–200.
- Brüggen, E., and Dholakia, U. M. (2010). Determinants of Participation and Response Effort in Web Panel Surveys. *Journal of Interactive Marketing*, 24(3), 239–250. <https://doi.org/10.1016/j.intmar.2010.04.004>.
- Callegaro, M., Shand-Lubbers, J., and Dennis, J. M. (2009). Presentation of a single item versus a grid: Effects on the vitality and mental health scales of the SF-36v2 health survey. *64th Annual Conference of the American Association for Public Opinion Research (AAPOR)*.
- Cantril, H. (1944). *Gauging public opinion*. Princeton University Press.
- Cheng, A., Zammaro, G., and Orriens, B. (2020). Personality as a predictor of unit nonresponse in an internet panel. *Sociological Methods and Research*, 49(3), 672–698.
- Couper, M. P., Antoun, C., and Mavletova, A. (2017). Mobile Web Surveys: A Total Survey Error Perspective. In P. P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, et al. (Eds.), *Total Survey Error in Practice*, (pp. 133–154). Wiley. <https://doi.org/10.1002/9781119041702.ch7>.
- Couper, M. P., Tourangeau, R., Conrad, F. G., and Zhang, C. (2013). The design of grids in web surveys. *Social Science Computer Review*, 31(3), 322–345.
- Couper, M. P., Traugott, M. W., and Lamias, M. J. (2001). Web Survey Design and Administration. *Public Opinion Quarterly*, 65(2), 230–253. <https://doi.org/10.1086/322199>.
- Couper, M. P. (2016). Grids versus item-by-item designs on smartphones. *GOR Conference, Dresden, Germany*.
- Craik, F. I., and Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684.
- De Leeuw, E. D. (2018). Mixed-Mode: Past, Present, and Future. *Survey Research Methods*, 12, 75–89.
- ESS. (2016). *European Social Survey. Source questionnaire*. https://www.europaeansocialsurvey.org/docs/round8/fieldwork/source/ESS8_source_questionnaires.pdf.
- Grady, R. H., Greenspan, R. L., and Liu, M. (2019). What Is the Best Size for Matrix-Style Questions in Online Surveys? *Social Science Computer Review*, 37(3), 435–445. <https://doi.org/10.1177/0894439318773733>.
- Grandmont, J., Graff, B., Goetzinger, L., and Dorbecker, K. (2010). Grappling with grids: How does question format affect data quality and respondent engagement. *Annual Meeting of the American Association for Public Opinion Research*. Retrieved from <http://www.Amstat.Org/Sections/SRMS/Proceedings/Y2010/Files/400116.Pdf>.
- Höhne, J. K., Schlosser, S., and Krebs, D. (2017). Investigating Cognitive Effort and Response Quality of Question Formats in Web Surveys Using Paradata. *Field Methods*, 29(4), 365–382. <https://doi.org/10.1177/1525822X1710640>.
- Iglesias, C. P., Birks, Y. F., and Torgerson, D. J. (2001). Improving the measurement of quality of life in older people: The York SF-12. *QJM: An International Journal of Medicine*, 94(12), 695–698. <https://doi.org/10.1093/qjmed/94.12.695>.
- John, O., Donahue, E., and Kentle, R. (1991). The big five inventory—Versions 4a and 54. In *Berkeley, Institute of Personality and Social Research*. Berkeley, CA: University of California.
- John, O. P., Naumann, L. P., and Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. *Handbook of Personality: Theory and Research*, 3(2), 114–158.
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., ... Nosek, B. A. (2014). Investigating Variation in Replicability: A “Many Labs” Replication Project. *Social Psychology*, 45(3), 142–152. <https://doi.org/10.1027/1864-9335/a000178>.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3), 213–236.
- Krosnick, J. A., and Abelson, R. P. (1992). The case for measuring attitude strength in surveys. In J. M. Tanur (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys*, (pp. 177–203). Russell Sage Foundation.
- Liu, M., and Cernat, A. (2018). Item-by-item Versus Matrix Questions: A Web Survey Experiment. *Social Science Computer Review*, 36(6), 690–706. <https://doi.org/10.1177/0894439316674459>.
- 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(6), 647–668.
- McClain, C., and Crawford, S. D. (2013). *Grid formats, data quality, and mobile device use: A questionnaire design approach*. Boston, USA: AAPOR Annual Conference.
- Moon, J. A., Keehner, M., and Katz, I. R. (2019). Affordances of Item Formats and Their Effects on Test-Taker Cognition under Uncertainty. *Educational Measurement: Issues and Practice*, 38(1), 54–62. <https://doi.org/10.1111/emip.12229>.
- Mutz, D. C. (2011). *Population-based survey experiments*. Princeton University Press.
- Nestler, S., Thielsch, M., Vasilev, E., and Back, M. D. (2015). Will they stay or will they go? Personality predictors of dropout in an online study. *International Journal of Internet Science*, 10(1), 37–48.
- Nielsen, J. S., and Kjær, T. (2011). Does question order influence sensitivity to scope? Empirical findings from a web-based contingent valuation study. *Journal of Environmental Planning and Management*, 54(3), 369–381. <https://doi.org/10.1080/09640568.2010.506087>.
- Peterson, G., Griffin, J., LaFrance, J., and Li, J. (2017). Smartphone participation in web surveys. *Total Survey Error in Practice*, 203–233.
- Revilla, M., and Couper, M. P. (2017). Comparing Grids With Vertical and Horizontal Item-by-Item Formats for PCs and Smartphones. *Social Science Computer Review*, 36(3), 349–368. <https://doi.org/10.1177/0894439317115626>.
- Revilla, M., Toninelli, D., and Ochoa, C. (2017). An experiment comparing grids and item-by-item formats in web surveys completed through PCs and smartphones. *Telematics and Informatics*, 34(1), 30–42. <https://doi.org/10.1016/j.tele.2016.04.002>.
- Roberts, B. W., Chernyshenko, O. S., Stark, S., and Goldberg, L. R. (2005). The structure of conscientiousness: An empirical investigation based on seven major personality questionnaires. *Personnel Psychology*, 58(1), 103–139.
- Rogelberg, S. G., Conway, J. M., Sederburg, M. E., Spitzmüller, C., Aziz, S., and Knight, W. E. (2003). Profiling Active and Passive Nonrespondents to an Organizational Survey. *Journal of Applied Psychology*, 88(6), 1104–1114. <https://doi.org/10.1037/0021-9010.88.6.1104>.
- Roßmann, J., Gummer, T., and Silber, H. (2018). Mitigating Satisficing in Cognitively Demanding Grid Questions: Evidence from Two Web-Based Experiments. *Journal of Survey Statistics and Methodology*, 6(3), 376–400. <https://doi.org/10.1093/jssam/smx020>.
- Salmon, C. T., and Neuwirth, K. (1990). Perceptions of opinion “climates” and willingness to discuss the issue of abortion. *Journalism Quarterly*, 67(3), 567–577.
- Schuman, H., Presser, S., and Ludwig, J. (1981). Context Effects on Survey Responses to Questions About Abortion. *Public Opinion Quarterly*, 45(2), 216. <https://doi.org/10.1086/268652>.
- Schuman, H., and Ludwig, J. (1983). The Norm of Even-Handedness in Surveys as in Life. *American Sociological Review*, 48(1), 112. <https://doi.org/https://doi.org/10.2307/2095149>.

- Schwarz, N., and Bless, H. (1991). *Constructing reality and its alternatives: An inclusion/exclusion model of assimilation and contrast effects in social judgment*.
- Schwarz, N., and Hippler, H.-J. (1995). Subsequent questions may influence answers to preceding questions in mail surveys. *Public Opinion Quarterly*, 59(1), 93–97. <https://doi.org/10.1086/269460>.
- Schwarz, N., and Schuman, H. (1997). Political knowledge, attribution, and inferred interest in politics: The operation of buffer items. *International Journal of Public Opinion Research*, 9(2), 191–195.
- Schwarz, N., Strack, F., and Mai, H.-P. (1991). Assimilation and contrast effects in part-whole question sequences: A conversational logic analysis. *Public Opinion Quarterly*, 55(1), 3–23.
- Silber, H., Höhne, J. K., and Schlosser, S. (2016). Question Order Experiments in the German-European Context. In *Survey Methods: Insights from the Field (SMIF)*. <https://doi.org/10.13094/SMIF-2016-00006>.
- Silber, H., Roßmann, J., and Gummer, T. (2018). When near means related: Evidence from three web survey experiments on inter-item correlations in grid questions. *International Journal of Social Research Methodology*, 21(3), 275–288. <https://doi.org/10.1080/13645579.2017.1381478>.
- Smith, T. W. (1988). Ballot position: An analysis of context effects related to rotation design. *GSS Methodological Re. No. 55*.
- Snidero, S., Zobec, F., Berchiolla, P., Corradetti, R., and Gregori, D. (2009). Question Order and Interviewer Effects in CATI Scale-up Surveys. *Sociological Methods and Research*, 38(2), 287–305. <https://doi.org/10.1177/0049124109346163>.
- Stark, T. H., Silber, H., Krosnick, J. A., Blom, A. G., Aoyagi, M., Belchior, A., ... Yu, R. (2020). Generalization of Classic Question Order Effects Across Cultures. *Sociological Methods and Research*, 49(3), 567–602. <https://doi.org/10.1177/0049124117747304>.
- Stern, M., Dillman, D. A., and Smyth, J. D. (2007). Visual design, order effects, and respondent characteristics in a self-administered survey. *Survey Research Methods*, 1, 121–138.
- Suchman, H., and Presser, S. (1981). *Questions and answers in attitude surveys: Experiments on question form, wording, and contexts*. New York: Academic Press.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285.
- Thau, M., Mikkelsen, M. F., Hjortskov, M., and Pedersen, M. J. (2021). Question order bias revisited: A split-ballot experiment on satisfaction with public services among experienced and professional users. *Public Administration*, 99(1), 189–204. <https://doi.org/10.1111/padm.12688>.
- Thomas, R. K., Barlas, F. M., Graham, P., and Subias, T. (2015). What they can't see can hurt you: Improving grids for mobile devices. In *Annual Conference of the American Association for Public Opinion Research*. Hollywood, FL.
- Thorndike, F. P., Carlbring, P., Smyth, F. L., Magee, J. C., Gonder-Frederick, L., Ost, L.-G., and Ritterband, L. M. (2009). Web-based measurement: Effect of completing single or multiple items per webpage. *Computers in Human Behavior*, 25(2), 393–401. <https://doi.org/10.1016/j.chb.2008.05.006>.
- Toepoel, V., Das, M., and Van Soest, A. (2009). Design of web questionnaires: The effects of the number of items per screen. *Field Methods*, 21(2), 200–213.
- Toepoel, V., and Ludtigi, P. (2015). Online surveys are mixed-device surveys. Issues associated with the use of different (mobile) devices in web surveys. *Methods, Data, Analyses*, 9(2), 8.
- Tourangeau, R., Couper, M. P., and Conrad, F. (2004). Spacing, Position, and Order: Interpretive Heuristics for Visual Features of Survey Questions. *Public Opinion Quarterly*, 68(3), 368–393. <https://doi.org/10.1093/poq/nfh035>.
- Tourangeau, R., Rasinski, K. A., Bradburn, N., and D'Andrade, R. (1989a). Belief accessibility and context effects in attitude measurement. *Journal of Experimental Social Psychology*, 25(5), 401–421.
- Tourangeau, R., Rasinski, K. A., Bradburn, N., and D'Andrade, R. (1989b). Carryover Effects in Attitude Surveys. *Public Opinion Quarterly*, 53(4), 495–524. <https://doi.org/10.1086/269169>.
- Tourangeau, R., Rips, L. J., and Rasinski, K. (2000). *The psychology of survey response*.
- Tourangeau, R., Singer, E., and Presser, S. (2003). Context Effects in Attitude Surveys: Effects on Remote Items and Impact on Predictive Validity. *Sociological Methods and Research*, 31(4), 486–513. <https://doi.org/10.1177/0049124103251950>.
- Wänke, M., and Schwarz, N. (1997). Reducing question order effects: The operation of buffer items. In L. Lyberg, P. Biemer, E. D. De Leeuw, C. Dippo, and N. Schwarz (Eds.), *Survey measurement and process quality*, (pp. 115–140). Wiley.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

