ARTICLES

The Impact of Question and Scale Characteristics on Scale Direction Effects

Ting Yan^{*}, Florian Keusch[†], Lirui He[‡]

Keywords: scale direction effects, questionnaire development, data collection mode, question length, question type, response scale

<https://doi.org/10.29115/SP-2018-0008>

Survey Practice

Vol. 11, Issue 2, 2018

Scale direction effects refer to the impact on survey responses of the direction in which a response scale is presented or read, holding other scale features constant. Scale direction tends to produce primacy effects. However, empirical research on scale direction effects turns up mixed evidence – scale direction effects are observed in some studies on some items, but not in other studies for other items. As a result, it is not clear under what circumstances scale direction effects can be expected. It is also unclear what questions are more prone to scale direction effects and what scale features induce or reduce scale direction effects. This paper advances the literature by examining empirically the impact of two question level characteristics (i.e., type of survey questions and location of questions in the questionnaire) and two scale features (i.e., type of response scales and the number of scale points) on scale direction effects. We found that scale direction effects are stronger for non-attitudinal items, earlier survey items, and items with longer scales. In addition, the moderating impact of question type, question location, and scale length on scale direction effects is more pronounced for items administered via Computer-Assisted Personal Interviewing than in self-administration. These findings have important implications for designing response scales in surveys.

INTRODUCTION

Studies have shown that many design features of response scales affect how survey respondents process a response scale and use the scale to construct their responses such as the number of scale points (Krosnick and Fabrigar 1997), the use of numeric labels (O’Muircheartaigh, Gaskell, and Wright 1995; Schwarz and Hippler 1995; Schwarz et al. 1991; Schwarz, Grayson, and Knauper 1998), the assignment of verbal labels to all or some of the scale points (Krosnick and Presser 2010), the spacing of response options (Daamen and de Bie 1992; Tourangeau 2004), the shading of response options (Tourangeau, Couper, and Conrad 2007), the alignment of the scale or the decision to present the scales horizontally or vertically on a screen or paper (Christian, Parsons, and Dillman 2009), and so on.

* Westat ORCID ID: 0000-0003-0187-3789

† University of Mannheim, Germany

‡ Jinan University, China

Scale direction is one design feature that has drawn less attention in the survey literature. A response scale could descend from high to low (e.g., “strongly agree” to “strongly disagree” or from “all of the time” to “never”). The same scale could also ascend from low to high (e.g., “strongly disagree” to “strongly agree” or from “never” to “all of the time”). Scale direction effects refer to changes in resultant responses due to the direction of the response scale while holding other features of a response scale constant.

Scale direction effects tend to take the form of primacy effects when scale direction is found to affect survey responses (Yan and Keusch 2015). Specifically, scale points are more likely to be selected when they are presented in the earlier part of a scale than when they are placed toward the end of the scale (i.e., when the scale is reversed). However, research on scale direction effects turns up mixed evidence — scale direction effects are observed in some studies on some items (e.g., Garbarski et al. 2015; Stapleton 2013; Toepoel et al. 2009), but not in other studies for other items (e.g., Israel and Taylor 1990; Krebs and Hoffmeyer-Zlotnik 2010; Malhotra 2009). Empirically, a key factor contributing to the mixed evidence of scale direction effects is that most of the experimental studies tested the scale direction effects on only one response scale with fixed scale features (e.g., Garbarski et al. 2015; Israel and Taylor 1990; Malhotra 2009; Stapleton 2013) and on one or more questions with similar wording (Krebs and Hoffmeyer-Zlotnik 2010; Toepoel, Das, and van Soest 2009). Question characteristics and scale features are neither systematically varied nor systematically examined. As a result, the studies failed to inform the survey field as to under what circumstances scale direction effects are observed, and there is not enough evidence to draw conclusions on what question- and scale-level characteristics are responsible for scale direction effects. This lack of attention to question- and scale-level characteristics results in a lack of clear guidance in the survey field on which scale direction works better under what circumstances. As a matter of fact, the decision on scale direction is considered to be “a matter of taste” (Rammstedt and Krebs 2007, 33).

The impact of question features and scale-level characteristics on scale direction effects is clearly overlooked and understudied. To overcome these limitations, this paper aims to explore the impact of question and scale characteristics on the scale direction effects through secondary data analysis. We selected question and scale features that are potentially associated with scale direction effects based on the satisficing and the anchoring and adjustment notion. Both accounts are cited as potential mechanisms underlying scale direction effects (Yan and Keusch 2015). In particular, we examine two question level characteristics — question type and the location of questions in the survey. The two scale features selected are the type of response scales and the number of scale points. We also selected one survey design feature — the mode of data collection.

We will answer the following three research questions:

1. What question level characteristics are associated with scale direction effects?
2. What scale features are more prone to scale direction effects?
3. Is there a mode difference with regards to the impact of question and scale characteristics on scale direction effects?

METHODS

Data are drawn from the American National Election Studies (ANES) 2008 and 2012 time series studies. The ANES is a national survey on political candidates, parties, American politics in general, and other related topics. The 2008 time series study was a face-to-face interview using Computer-Assisted Personal Interviewing (CAPI), including a Computer-Assisted Self-Interviewing (CASI) component. The American Association for Public Opinion Research (AAPOR) response rate (RR1) for the pre-election interviews is 59.5% and 53.9% for the post-election interviews (Lupia et al. 2009). The total number of completed interviews used for our analyses is 2,322.

The 2012 time series study was conducted in two modes (face-to-face and Web) independently, using separate samples. For the face-to-face mode, people sampled via address-based sampling were recruited and interviewed in person using CAPI, which also included a CASI component. The AAPOR RR1 for the pre-election study is 38.0% and 35.7% for the post-election study (American National Election Studies 2014). The Web sample constituted a representative sample separate from the face-to-face sample and was drawn from panel members of GfK Knowledge Networks. The response rate (AAPOR RR1) for the pre-election study is 2.0% and 1.8% for the post-election study (American National Election Studies 2014). The total number of completed interviews included in our analyses is 5,914.

An experiment was included in both years that varies the direction of response scales; a random half of the respondents were presented the scales in the “forward” order (all but 8 items employ a scale running from high to low), whereas the other half received the same scales in the “reversed” order (scales ran from low to high). There were 253 survey items from the 2008 pre- and post-election studies and 375 survey items from the 2012 pre- and post-election studies subjected to this scale direction experiment. All 628 survey items are double-coded on the following characteristics — question type, location of question in the survey, response scale type, and the number of scale points. When the two coders disagreed, the issues were brought for discussion with the first two authors, who reviewed the issues together with the two coders and resolved the discrepancy through discussion.

RESULTS

MAIN EFFECT OF SCALE DIRECTION

To provide a general idea on the main effects of scale direction, we first compared the average proportion of respondents selecting options appearing first in the “forward” order across all items by scale direction. For scales with an odd number of scale points, the first half of the scales includes the first scale point (for 3-point scales) or the first 2 scale points (for 5-point scales). For scales with an even number of scale points, the first half of the scales includes the first 2 scale points (for 4-point scales) and the first 3 scale points (for 6-point scales).

When scales are presented in the “forward” order, we found that an average of 46.0% respondents chose from the first half of the scales. However, only about 43.5% of respondents choose from that side when the same scales are presented in the reversed order (and that side became the latter half), producing a significant primacy effect ($F(1,1254)=4.6, p=0.03$).

We then looked at the proportion of survey items that exhibited significant scale direction effects (i.e., significant primacy effects). Across the 628 survey items from the 2008 and the 2012 studies, 25.2% of the survey items exhibited significant scale direction effects. In other words, scale direction significantly affects survey responses on a quarter of the survey items. However, 12 out of 628 items exhibited significant recency effects.

SCALE DIRECTION BY QUESTION TYPE

All survey items are coded as attitudinal ($n=564$) or non-attitudinal ($n=64$). The “non-attitudinal” category includes questions that encompass both subjective and behavioral components and cannot be easily coded as either attitudinal or behavioral. Significantly more non-attitudinal items (37.5%) exhibited scale direction effects than attitudinal items (23.8%) ($\chi^2(1)=5.8, p=0.02$). In addition, scale direction produced a larger primacy effect for non-attitudinal items than for attitudinal items; 4.0% more respondents chose from scale points presented earlier in the “forward” order than in the “reversed” order for non-attitudinal items, whereas scale direction only produced a difference of 2.3% in the proportion endorsing scale points appearing first in the “forward” order for attitudinal items ($F(1,626)=6.1, p=0.01$).

SCALE DIRECTION BY LOCATION OF QUESTION

We found stronger scale direction effects for questions located in the first half of the questionnaire than in the second half; about one-third (30.2%) of 311 survey items in the first half of the questionnaire exhibited scale direction effects, whereas only one-fifth of 317 survey items in the second half of the questionnaire did so ($\chi^2(1)=8.4, p=0.004$). Scale direction yielded a difference of 3.0% in the proportion endorsing scale points appearing first in the “forward” order for earlier items and a difference of 2.0% for later items ($F(1,626)=7.2, p=0.001$).

SCALE DIRECTION BY SCALE TYPE

Following Fowler's (1995) scheme, we coded response scales into one of the four types. The first type of response scales — attitudinal scales — includes in the scale labels both the domain to be evaluated and quantifiers (e.g., very satisfied-fairly satisfied-not very satisfied-not satisfied at all); 124 items use attitudinal scales. Fifty-two items are asked on a frequency scale (e.g., always-nearly always-part of the time-seldom). Quantity scales only contain quantifiers in the scale labels (e.g., a great deal-quite a bit-some-very little-none), and 205 questions use a scale in this type. The last type of scales — evaluative scales — is used for evaluation purpose (e.g., respondents are asked to evaluate if the phrase "He is moral" describes President Obama extremely well, very well, moderately well, slightly well, or not well at all). Sixty-six survey items are asked on evaluative scales. We did not find a moderating effect of scale type on scale direction effects; about equal percentages of survey items exhibited scale direction effects and the size of scale direction effects are comparable across scale types.

SCALE DIRECTION BY SCALE LENGTH

A total of 170 questions are asked on a 3-point scale and 62 on a 4-point scale. The 5-point scales are used on 213 questions. Only 2 questions use a 6-point scale. We found more pronounced scale direction effects by scale length. One-third of scales with 5 or 6 scale points (33.5%) showed scale direction effects, whereas 16.6% of 3-point scales and 13.5% of 4-point scales did so ($\chi^2(2)=26.4, p<0.0001$). Scale direction produced a larger difference in the proportion of endorsing earlier scale points in the "forward" order for longer scales (2.9%) than for shorter scales (2.2% for 3-point scales and 1.9% for 4-point scales) ($F(2,625)=2.6, p=0.08$).

SCALE DIRECTION BY MODE OF DATA COLLECTION

A total of 419 survey items were administered via CAPI, and 209 survey items were administered via one of the two self-administered modes (Web and audio CASI). Stronger scale direction effects are observed on items administered in a self-administered mode than in CAPI. About one third (34.9%) of self-administered surveys items produced scale direction effects, whereas 20.3% of items administered via CAPI did so ($\chi^2(1)=15.9, p<0.0001$). Scale direction induced 3.2% more endorsement of scale pointing appearing first in the "forward" for the self-administered items than for CAPI items (2.2%) ($F(1,626)=5.9, p=0.02$).

MODE DIFFERENCE IN IMPACT OF QUESTION CHARACTERISTICS AND SCALE FEATURES

We further examined whether or not there is a mode difference with regard to the impact of question and scale characteristics on scale direction effects. We found that the impact of question type, question location, and scale length on

scale direction effects is more pronounced for CAPI interviews, but less so for self-administered interviews. There is no mode difference with regard to the impact of scale type on scale direction effects.

DISCUSSION

This paper examines the impact of several question- and scale-level characteristics on scale direction effects through secondary data analyses. In the 2008 and 2012 ANES studies, scale direction has a significant main effect on respondents' answers; across all items, more respondents chose from options when they appear first (in the forward order) than when they appear later (in the reversed order).

More importantly, certain question- and scale-level characteristics are associated with the likelihood to observe scale direction effects. Non-attitudinal items, earlier items, and items with longer scales are shown to be more prone to scale direction effects than attitudinal items, later items, and shorter scales. Although there is no mode difference in terms of scale direction effects, the moderating effects of question type, question location, and scale length on scale direction effects are more pronounced in CAPI interviews than in self-administered modes.

The findings have important implications for questionnaire developers. First, although long scales with more scale points have the advantage of better differentiation, long scales (e.g., 7-point or 11-point agreement scales) have been found to yield data of worse quality than 5-point scales (Revilla, Saris, and Krosnick 2013). Long scales are more prone to scale direction in this current study. These two pieces of evidence together argue for more use of shorter scales than longer scales.

Second, survey items coded as “non-attitudinal” encompass both behavioral and subjective components in single items. We suspect that these items are probably hard to answer and advise questionnaire developers to avoid them.

Third, scale direction is shown to affect survey responses and its impact is stronger for earlier items than for later items. We recommend that questionnaire developers rotate the scale order and even question order when possible. Rotation allows researchers to examine and to statistically control for the effect of scale direction and the moderating effect of question order on scale direction effects.

Fourth, the findings shed light on the mechanism underlying scale direction effects. Although both the satisficing notion and the anchoring and adjustment heuristic predict that scale direction affects resultant answers, our findings of the moderating effect of question order on scale direction effects do not seem to offer strong evidence supporting the satisficing account, consistent with earlier research (Carp 1974; Mingay and Greenwell 1989; Yan and Keusch 2015).

The major limitation of this paper is that secondary data analysis has been conducted on one survey with a specific topic and sponsor. The findings are observational in nature. Also, we are not able to study the impact of other key question- and scale-level characteristics on scale direction effects such as scale labeling and scale polarity. Future research should involve experimental manipulation of important question and scale features that are expected to affect respondents' processing and use of response scales in order to advance our understanding on scale direction effects. Apparently, more empirical evidence is needed before the survey field can draw conclusions and make recommendations or guidelines on what the ideal scale direction should be.

Another limitation of the paper is that we only examined the impact of scale direction on survey responses; we did not consider other aspects of data quality induced by scale direction. Further research should also investigate whether or not scale direction affects the quality of the resultant answers such as reliability and validity.

REFERENCES

- American National Election Studies. 2014. *User's Guide and Codebook for the ANES 2012 Time Series Study*. Ann Arbor, MI and Palo Alto, CA: University of Michigan and Stanford University.
- Carp, F.M. 1974. "Position Effects on Interview Response." *Journal of Gerontology* 29 (5): 581–87.
- Christian, Leah Melani, Nicholas L. Parsons, and Don A. Dillman. 2009. "Designing Scalar Questions for Web Surveys." *Sociological Methods & Research* 37 (3): 393–425. <https://doi.org/10.1177/0049124108330004>.
- Daamen, D.D.L., and S.E. de Bie. 1992. "Serial Context Effects in Survey Items." In *Context Effects in Social and Psychological Research*, edited by Norbert Schwarz and S. Sudman. Springer, New York.
- Fowler, F.J. Jr. 1995. *Improving Survey Questions: Design and Evaluation*. Thousand Oaks, CA: Sage Publications.
- Garbarski, Dana, Nora Cate Schaeffer, and Jennifer Dykema. 2014. "The Effects of Response Option Order and Question Order on Self-Rated Health." *Quality of Life Research* 24 (6): 1443–53. <https://doi.org/10.1007/s11136-014-0861-y>.
- Israel, Glenn D., and C.L. Taylor. 1990. "Can Response Order Bias Evaluations?" *Evaluation and Program Planning* 13 (4): 365–71. [https://doi.org/10.1016/0149-7189\(90\)90021-n](https://doi.org/10.1016/0149-7189(90)90021-n).
- Krebs, Dagmar, and Juergen H.P. Hoffmeyer-Zlotnik. 2010. "Positive First or Negative First?" *Methodology* 6 (3): 118–27. <https://doi.org/10.1027/1614-2241/a000013>.
- Krosnick, J.A., and L.R. Fabrigar. 1997. "Designing Rating Scales for Effective Measurement in Surveys." In *Survey Measurement and Process Quality*, edited by L. Lyberg, P. Biemer, M. Collins, L. Decker, E. DeLeeuw, C. Dippo, Norbert Schwarz, and D. Trewin. New York: Wiley-Interscience.
- Krosnick, J.A., and S. Presser. 2010. "Questionnaire Design." In *Handbook of Survey Research*, edited by P.V. Marsden and J.D. Wright. West Yorkshire, England: Emerald Group.
- Lupia, A., J.A. Krosnick, P. Luevano, M. DeBell, and D. Donakowski. 2009. *User's Guide to the ANES 2008 Time Series Study*. Ann Arbor, MI and Palo Alto, CA: University of Michigan and Stanford University.
- Malhotra, N. 2009. "Order Effects in Complex and Simple Tasks." *Public Opinion Quarterly* 73 (1): 180–98.
- Mingay, D., and M.T. Greenwell. 1989. "Memory Bias and Response-Order Effects." *Journal of Official Statistics* 5 (3): 253–63.
- O'Muircheartaigh, C., G. Gaskell, and D.B. Wright. 1995. "Weighing Anchors: Verbal and Numeric Labels for Response Scales." *Journal of Official Statistics* 11 (3): 295–307.
- Rammstedt, Beatrice, and Dagmar Krebs. 2007. "Does Response Scale Format Affect the Answering of Personality Scales?" *European Journal of Psychological Assessment* 23 (1): 32–38. <https://doi.org/10.1027/1015-5759.23.1.32>.
- Revilla, Melanie A., Willem E. Saris, and Jon A. Krosnick. 2013. "Choosing the Number of Categories in Agree–Disagree Scales." *Sociological Methods & Research* 43 (1): 73–97. <https://doi.org/10.1177/0049124113509605>.

- Schwarz, Norbert, C. E. Grayson, and B. Knauper. 1998. "FORMAL FEATURES OF RATING SCALES AND THE INTERPRETATION OF QUESTION MEANING." *International Journal of Public Opinion Research* 10 (2): 177–83. <https://doi.org/10.1093/ijpor/10.2.177>.
- Schwarz, Norbert, and Hans-J. Hippler. 1995. "THE NUMERIC VALUES OF RATING SCALES: A COMPARISON OF THEIR IMPACT IN MAIL SURVEYS AND TELEPHONE INTERVIEWS." *International Journal of Public Opinion Research* 7 (1): 72–74. <https://doi.org/10.1093/ijpor/7.1.72>.
- Schwarz, Norbert, Barbel Knauper, Hans-J. Hippler, Elisabeth Noelle-Neumann, and Leslie Clark. 1991. "Rating Scales: Numeric Values May Change the Meaning of Scale Labels." *Public Opinion Quarterly* 55 (4): 570. <https://doi.org/10.1086/269282>.
- Stapleton, C.E. 2013. "The Smartphone Way to Collect Survey Data." *Survey Practice* 6 (2).
- Toepoel, V., M. Das, and A. van Soest. 2009. "Design of Web Questionnaires: The Effect of Layout in Rating Scales." *Journal of Official Statistics* 25 (4): 509–28.
- Tourangeau, Roger. 2004. "Spacing, Position, and Order: Interpretive Heuristics for Visual Features of Survey Questions." *Public Opinion Quarterly* 68 (3): 368–93. <https://doi.org/10.1093/poq/nfh035>.
- Tourangeau, Roger, Mick P. Couper, and Frederick Conrad. 2007. "Color, Labels, and Interpretive Heuristics for Response Scales." *Public Opinion Quarterly* 71 (1): 91–112. <https://doi.org/10.1093/poq/nfl046>.
- Yan, T., and F. Keusch. 2015. "The Effects of the Direction of Rating Scales on Survey Responses in a Telephone Survey." *Public Opinion Quarterly* 79 (1): 145–65. <https://doi.org/10.1093/poq/nfu062>.