# Mismatching middle options: consequences for attitude measurement in smartphone surveys

Jan Karem Höhne & Dagmar Krebs

Published online: 04 Aug 2020.

Submit your article to this journal

View related articles

View Crossmark data

Routledge
Taylor & Francis Group

Check for updates

# Mismatching middle options: consequences for attitude measurement in smartphone surveys

Jan Karem Höhne[a,b] and Dagmar Krebs[c]

[a]University of Mannheim, Collaborative Research Center 884 'Political Economy of Reforms', Mannheim, Germany; [b]Universitat Pompeu Fabra, Research and Expertise Centre for Survey Methodology, Barcelona, Spain; [c]University of Gießen, Faculty of Social Sciences and Cultural Studies, Institute for Sociology, Gießen, Germany

**ABSTRACT**
Measuring respondents' attitudes is a crucial task in numerous social science disciplines. A popular way to measure attitudes is to use survey questions with rating scales. However, research has shown that especially the design of rating scales can have a profound impact on respondents' answer behavior. While some scale design aspects, such as scale length and direction, are frequently researched, some other scale design aspects, such as scale midpoint and polarity, are under-researched. In this study, we therefore investigate the effects of mismatches between scale midpoints and scale polarity – i.e., unipolar 'moderately' vs. bipolar 'partly/partly' middle options in unipolar scales – on respondents' answer behavior. We conducted an experiment in a smartphone survey (N = 1,641) and randomly assigned respondents to one of two scale conditions (match vs. mismatch). The results reveal that mismatches between scale midpoints and scale polarity slightly affect respondents' answer behavior. More specifically, mismatches cause small but significant shifts in latent means. Thus, mismatches pose a threat to attitude measurement.

## Introduction

The measurement of attitudes is an important task in social science research to explore and explain social phenomena. To measure attitudes of respondents, researchers typically use rating scales (i.e., closed answer formats with a list of ordered options). When designing such scales, researchers must make careful design decisions, because they can influence respondents' answer behavior (see DeCastellarnau, 2018). For instance, researchers must decide whether to have a scale midpoint (i.e., an even or uneven number of scale points). Further scale characteristics that must be decided include the lengths of the scale (i.e., the number of scale points), the polarity of the scale (i.e., unipolar or bipolar), the verbalization of the scale (e.g., completely or end verbalized), the inclusion of numeric values (i.e., whether numbers accompany the scale points), the direction of the scale (i.e., decremental or incremental), and the alignment of the scale (i.e., horizontal or vertical). In completely verbalized scales, the aspects of a scale midpoint (or middle option) and scale polarity frequently conflict with each other. The reason is that scale midpoints do not match the polarity of the scales (Menold, 2019).

For instance, the English source questionnaire of the interviewer-administered International Social Survey Programme (ISSP; 2012) uses survey questions on gender roles with a completely labeled bipolar agreement/disagreement scale. The German version, however, employs comparable survey questions on gender roles with a unipolar agreement scale that contains a bipolar 'neither/

---

nor' instead of a unipolar 'moderate' middle option (see also Scholz & Jutz, 2014). Conceptually, the 'neither/nor' middle option does not match the unipolar scale. As shown by Höhne et al. (2020), even slightly differently labeled scales can have an impact on observational and latent answer distributions, affecting measurement properties (i.e., equidistance between scale points).

While unipolar midpoints (usually) suggest a moderate level of agreement, bipolar midpoints can have different meanings, depending on their formulation (see Menold, 2019; Wang & Krosnick, 2019). They can be conceived as indicating indifference or ambivalence (Menold, 2019; O'Muircheartaigh et al., 1995; Schaeffer & Presser, 2003). Indifference implies that respondents have either a neutral or no position at all towards the object under investigation (Menold, 2019; Sturgis et al., 2014). This applies when using middle options that consist of 'neither/nor' formulations. In contrast, ambivalence implies that respondents have both positive and negative feelings, preventing them from having a clear attitude towards the object under investigation (Menold, 2019; Schaeffer & Presser, 2003). This applies when using middle options that consist of 'partly/partly' formulations. These (linguistic/logical) differences between middle options have the potential to change the evaluative scale character (Höhne et al., 2020), particularly if middle options and scale polarity mismatch.

In contrast to research on other scale design aspects, such as the number of scale points and scale direction, research on the impact of the middle option on answer behavior and data quality is scarce. Menold (2019) conducted one notable study. In a web survey experiment (no device distinction), the author compared unipolar agreement scales with bipolar agreement/disagreement scales and systematically mismatched the respective middle options. The author shows that reliability decreases when the middle options do not match scale polarity. However, so far, very little is known about the consequences for attitude measurement when middle options and scale polarity mismatch. In this study, we attempt to fill this knowledge gap and address the following research question: How does the mismatch of a middle option in a unipolar scale affect respondents' answer behavior in terms of observational and latent answer distributions?

Unlike previous studies, we extend the current state of research by focusing on smartphone surveys. The reasons for this design decision are twofold: first, there is (almost) no research on the effects of middle options and scale polarity on respondents' answer behavior in smartphone surveys. Second, in recent years, the number of smartphone surveys has increased tremendously.

## Method

### Data source

Data collection was conducted by the survey company Respondi (Germany) in July and August 2019. The company invited respondents varying in age from 18 to 70 years by email. The email included an invitation to participate in a smartphone survey and a URL link that directed respondents to the survey. Clicking the link leads to an introductory page that described the survey topic and procedure and informed respondents that their data would be treated confidentially. Respondents who tried to enter the survey with a non-smartphone device were blocked and asked to use a smartphone for survey completion. We also collected user-agent-strings informing about device characteristics, such as device type and operating system. For this purpose, we used the open-source tool 'Embedded Client Side Paradata (ECSP)' developed by Schlosser and Höhne (2018, 2020).

If they participate, respondents received a financial compensation from the survey company that is proportional to the length of the entire survey.

## Sample

A total of 1,726 respondents started the smartphone survey, which took about 20 minutes. Some respondents were excluded from the study (n = 85) because they only visited the title page of the web survey or they dropped out before being asked any experimental questions. This leaves us with n = 1,641 respondents for statistical analyses. These respondents had a mean age of 43.3 (SD = 15.0), and 62.7% of them were female. In terms of education, 9.0% graduated from a lower secondary school, 37.2% from an intermediate secondary school, and 53.0% from a college preparatory secondary school or university. The remaining 0.8% still attended school, left school without a degree, or had a different degree from those mentioned above.

## Experimental design

Respondents were randomly assigned to one of two scale conditions. The first group (n = 834) received a unipolar agreement scale with a matching middle option (match condition; 'agree moderately'). The second group (n = 807) received a unipolar agreement scale with a mismatching middle option (mismatch condition; 'partly/partly').

To evaluate the sample composition between the groups, we conducted chi-square tests. The results showed no significant differences regarding age, gender, and education.

## Survey questions

The five questions used were adopted from those used in the Cross Cultural Survey for Work and Gender Attitudes (2010). The selected questions dealt with achievement motivation and were presented with a single presentation approach (i.e., one question per web survey page). The questions were asked with five-point, vertically aligned agreement scales. All questions and answer options were in German (see Appendix for English translations). To improve survey completion and navigation, we used an optimized survey layout that avoids horizontal scrolling.

# Results

## Answer distributions

We tested whether scales with matching and mismatching middle options differ regarding answer distributions. For this purpose, we conducted chi-square tests for each of the five questions. We also tested the percentages of the two differing middle options for statistical differences. Table 1 displays the results.

**Table 1.** Answer distributions in percentages.

| Options | Question 1: $\chi^2(4) = 3.35$, p = 0.50 | | Question 2: $\chi^2(4) = 2.93$, p = 0.57 | | Question 3: $\chi^2(4) = 15.98$, p < 0.01 | | Question 4: $\chi^2(4) = 8.42$, p = 0.07 | | Question 5: $\chi^2(4) = 2.00$, p = 0.74 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Match | Mismatch | Match | Mismatch | Match | Mismatch | Match | Mismatch | Match | Mismatch |
| 1 | 5 | 6 | 17 | 20 | 8 | 11 | 10 | 13 | 15 | 17 |
| 2 | 25 | 23 | 45 | 44 | 32 | 35 | 34 | 32 | 39 | 39 |
| 3 | 38 | 42 | 26 | 26 | **41** | **36** | 32 | 35 | 30 | 30 |
| 4 | 23 | 21 | 9 | 8 | 15 | 16 | 18 | 14 | 11 | 10 |
| 5 | 10 | 8 | 4 | 3 | 5 | 3 | 7 | 6 | 5 | 4 |

Bold indicates significant differences between middle options (p < 0.05). Due to rounding, the percentages may not add up to 100%. Verbal labels of the match condition: 1 'agree strongly', 2 'agree somewhat', 3 'agree moderately', 4 'agree hardly', and 5 'agree not at all'. Verbal labels of the mismatch condition: 1 'agree strongly', 2 'agree somewhat', 3 'partly/partly', 4 'agree hardly', and 5 'agree not at all'.

The chi-square tests show no significant differences for four out of five questions. The only exception is the third question. For this question, the matching middle option (i.e., 'agree moderately') is selected significantly more often than its mismatching counterpart (i.e., 'partly/partly').

## *Measurement invariance and latent mean differences*

We conducted confirmatory factor analyses (CFAs) using one latent variable and five indicators (questions on achievement motivation) for scales with matching and mismatching middle options. We then conducted multigroup confirmatory factor analyses (MG-CFAs) to test for configural, metric, and scalar invariance. We used non-significant differences between chi-square values (Byrne, 2012) and differences between CFIs (Comparative Fit Index) and RMSEAs (Root Mean Square Error of Approximation) lower than 0.01 (Cheung & Rensvold, 2002) as criteria for accepting measurement invariance. Since all indicators were measured with five-point scales, we assumed a continuous scale level (Rhemtulla et al., 2012) and used the MLR (Robust Maximum Likelihood) discrepancy function. Table 2 displays the results.

As Table 2 reveals, measurement invariance holds for scales with matching and mismatching middle options. The presence of measurement invariance is indicated by the non-significant results of the chi-square difference tests and implies that both scales are comparable.

We also examined latent mean differences between the scales with matching and mismatching middle options [$\chi^2(17) = 14.91$ (1.14), RMSEA = 0.00, CFI = 1.00]. The results show a significant latent mean difference of $-0.116$ (p = 0.031) between scale conditions. The scales with a mismatching 'partly/partly' midpoint produce significantly lower (more positive) answers than scales with a matching 'agree moderately' midpoint (reference group).

**Table 2.** Testing for measurement invariance.

| Invariance levels | $\chi^2$ values | Df | RMSEA | CFI | $\chi^2$ difference test |
|---|---|---|---|---|---|
| Configural | 9.17 (1.28) | 8 | 0.013 | 0.999 | |
| Metric | 10.94 (1.19) | 13 | 0 | 1 | 1.22 |
| Scalar | 19.09 (1.14) | 18 | 0.009 | 0.999 | 8.66 |

The results are based on the MLR discrepancy function. Scale correction factors are in parentheses.

## *Composite reliability*

We also estimated composite reliabilities of the scales with matching and mismatching middle options. For this purpose, we followed the method suggested by Raykov and Marcoulides (2011). Interestingly, the results reveal almost no differences between the two scales, which have reliability scores of 0.80 (match condition) and 0.81 (mismatch condition). Thus, mismatching midpoints do not reduce measurement quality in terms of composite reliability.

## Discussion and conclusion

The goal of this study was to investigate the consequences of mismatches between scale midpoints (or middle options) and scale polarity on attitude measurement in smartphone surveys. For this purpose, we conducted an experiment in a smartphone survey with two scale conditions (match vs. mismatch) and analyzed respondents' answer behavior in terms of observational and latent answer distributions. The results reveal only small differences between scales with a matching middle option and those with a mismatching middle option. This applies to the observational and latent level.

Overall, the percentages of answers to the scale with the matching middle option ('agree moderately') do not significantly differ from the percentages of answers to the scale with the mismatching middle option ('partly/partly'). This finding points to the fact that respondents may

not interpret or perceive the matching and mismatching middle options differently. In other words, both midpoints seem to imply the same meaning to respondents.

The analyses on the latent level show that scalar invariance holds for both scale conditions. This implies that mismatches between the middle option and scale polarity do not affect the intercepts and that the two scales are comparable. Scalar invariance is a prerequisite for comparing latent means. Although the latent mean differences between scales with matching and mismatching middle options are relatively small, they are significantly different. Scales with mismatching 'partly/partly' midpoints yield more positive answers than scales with matching 'agree moderately' midpoints. However, both scales do not differ in terms of composite reliability. Thus, our findings differ from those by Menold (2019).

This study has some limitations that provide avenues for future research. First, we only investigated the effects of mismatching middle options in unipolar agreement scales. However, it is also important to investigate the effects of mismatching middle options in bipolar agreement/disagreement scales. Second, we only tested a 'partly/partly' formulation, neglecting the possibility of a 'neither/nor' one. Third, we focused exclusively on questions dealing with achievement motivation. We suggest that future research covers a variety of further question topics. Finally, our study was solely conducted in a non-probability access panel in Germany and, thus, we cannot provide any conclusions on respondents' answer behavior in a cross-cultural or cross-national setting. It would also be worthwhile to build on this study by rerunning the experiment in a probability-based panel.

Our results show that respondents' answer behavior seems to be relatively robust against mismatches of middle options in unipolar agreement scales. However, mismatches, as in the case of the German questionnaire of the ISSP (2012), may have consequences for attitude measurement in the form of latent mean differences. We therefore encourage researchers to pay close attention to design aspects of rating scales for attitude measurement in smartphone surveys.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Notes on contributors

*Jan Karem Höhne* (hoehne@uni-mannheim.de) is a postdoctoral researcher at the Collaborative Research Center 884 'Political Economy of Reforms' at the University of Mannheim, Germany, and research fellow at the 'Research and Expertise Centre for Survey Methodology' at the Universitat Pompeu Fabra in Barcelona, Spain. His research combines survey methodology, psychology, computer science, and data science.

*Dagmar Krebs* (dagmar.krebs@sowi.uni-giessen.de) is a professor for empirical social research and statistics at the Institute for Sociology at the University of Giessen, Germany. Her research focuses on survey methodology in general and response format and response behavior in particular.

## References

Byrne, B. M. (2012). *Structural equation modeling with Mplus. Basic concepts, applications, and programming.* Routledge.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

DeCastellarnau, A. (2018). A classification of response scale characteristics that affect data quality: A literature review. *Quality and Quantity*, *52*(4), 1523–1559. https://doi.org/10.1007/s11135-017-0533-4

Höhne, J. K., Krebs, D., & Kühnel, S.-M. (2020). Measuring income (in)equality: Comparing survey questions with unipolar and bipolar scales in a probability-based online panel. *Social Science Computer Review*. https://doi.org/10.1177/0894439320902461

Menold, N. (2019). Response bias and reliability in verbal agreement rating scales: Does polarity and verbalization of the middle category matter? *Social Science Computer Review*. https://doi.org/10.1177/0894439319847672

O'Muircheartaigh, C., Gaskell, G., & Wright, D. B. (1995). Weighing anchors: Verbal and numeric labels for response scales. *Journal of Official Statistics*, *11*, 295–307.

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. Taylor & Francis.

Rhemtulla, M., Brosseau-Liard, P. E., & 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*(3), 354–373. https://doi.org/10.1037/a0029315

Schaeffer, N. C., & Presser, S. (2003). The science of asking questions. *Annual Review of Sociology*, *29*(1), 65–88. https://doi.org/10.1146/annurev.soc.29.110702.110112

Schlosser, S., & Höhne, J. K. (2018). ECSP – Embedded client side paradata. *Zenodo*. http://doi.org/10.5281/zenodo.1218941

Schlosser, S., & Höhne, J. K. (2020). ECSP – Embedded client side paradata. *Zenodo*. http://doi.org/10.5281/zenodo.3782592

Scholz, E., & Jutz, R. (2014). *ISSP 2012 Germany: Family and gender roles IV*. GESIS Report on the German Study. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-402638

Sturgis, P., Roberts, C., & Smith, P. (2014). Middle alternatives revisited: How the neither/nor response acts as a way of saying "I don't know"? *Sociological Methods & Research*, *43*(1), 15–38. https://doi.org/10.1177/0049124112452527

Wang, R., & Krosnick, J. A. (2019). Middle alternatives and measurement validity: A recommendation for survey researchers. *International Journal of Social Research Methodology*, *23*(2), 169–184. https://doi.org/10.1080/13645579.2019.1645384

## Appendix

English translations of the five survey questions (including answer options) on achievement motivation.

I like being in competition with other people.
It is satisfying when I achieve better results than other people.
I am always trying to perform better than other people.
I try harder when I am in competition with other people.
It is important for me to be the best at a task.

Answer options in the match condition: agree strongly, agree somewhat, agree moderately, agree hardly, agree not at all.

Answer options in the mismatch condition: agree strongly, agree somewhat, partly/partly, agree hardly, agree not at all.

The question order in the survey corresponds to the order displayed in the Appendix. The original German wordings of the questions and answer options are available from the first author on request.