

The SOM Institute's Notes on Survey Methodology - 2023:3

Investigating acquiescence response bias in a few Swedish questions

Freja Wessman

The SOM institute, University of Gothenburg
Freja.wessman@gu.se

ABSTRACT

This study investigated the prevalence of acquiescence response bias (ARB) in a Swedish context. Three versions of question formats were compared using an experimental approach; a Likert scale with agree/disagree format, a response option ranging from "Very good proposal" to "very bad proposal", and an item-specific response option. Surprisingly, item specific response options led to greater average positivity than the agree/disagree response option format. ARB was further assessed by comparing two questions with reversed order of response options, but inconsistent answers were not greater in any of the question formats. Finally the data quality of the three question formats were compared using concurrent validity. The item specific response option appeared to give slightly better data quality than agree/disagree and proposal response options. However, the lesser data quality in the agree/disagree and proposal formats seemed unlikely attributable to ARB. The results indicate that ARB may be less of an issue in a Swedish context.

INTRODUCTION

When answering questions in surveys, many forms of biases may influence the answers given by respondents. One such bias is acquiescence response bias (ARB). ARB is the bias appearing by respondents tendency to agree to questions regardless of content. This can cause respondents true opinions to not be reflected in the data. ARB is a well known phenomenon and has been documented over the past half century (Lelkes & Weiss, 2015). One reason for ARB is that respondents often treat surveys as conversational, and therefore wishes to be agreeable with the questions asked in surveys (Pasek & Krosnick, 2010). Another reason for ARB can be that the respondents sometimes satisfice when answering questions, which means that respondents who spend too little cognitive effort to elucidate an answer may default to agreeing to the questions (Krosnick, 1999; Tourangeau, 2018).

Both the design of the survey and characteristics of the respondents may increase the chance for ARB (Narayan & Krosnick, 1996). The survey design increase ARB when questions are asked in either an agree/disagree, true/false or yes/no-form (Krosnick, 1991). Individual characteristics known to increase ARB are less formal education, lower income, and living in a country with high level of corruption or autocracy (Harzing, 2006; Krosnick 1996; Meisenberg & Williams, 2008).

To test if ARB occurs in survey data, studies have investigated the tendency to score highly on agree/disagree response options, while also including contradictory statement which people should not agree to simultaneously (Meisenberg & Williams, 2008). The tendency to agree was measured by giving 0 to the extreme disagree response and then add 1 for each step towards agree, which meant a question with two response options had the value 0 and

1 while a question with five response options had the value 0 to 4. By adding all of these into a score the researchers could compare the acquiescence with individual characteristics. Further, Meisenberg and Williams (2008) analyzed a similar bias called extreme response styles, which meant the tendency to agree or disagree completely to statements. This was calculated by coding the two most extreme responses as 1, and all response items as 0, to see if certain characteristics influences the tendency to give an extreme response. The tendency of extreme responding was correlated with lower levels of education.

Other ways to estimate ARB is to compare agree/disagree response options to questions using item specific (IS) response option, to see if one of the two yield better data quality (Dykema et al., 2022). In IS questions the response alternatives include the construct being measured rather than just agreeing to a statement. Saris and Krosnick (2005) provided this example of an IS question: “How would you rate your health – excellent, very good, good, fair or bad?” whereas the agree/disagree question would be: “To what extent do you agree strongly or disagree strongly that your health is excellent? 1 agree completely, 2 agree somewhat, 3 neither agree nor disagree, 4 disagree somewhat, 5 disagree completely.”

Studies have also used eye tracking-technology to assess whether different ways of asking questions affect respondents’ cognitive processing (Höhne & Lenzner, 2018). The results indicated that item specific questions required more intense processing which may lead to more well-considered responses and less satisficing.

The aim of this study was to test ARB in a Swedish context, comparing three different question formats by analyzing the tendency to agree regardless of content and also assessing which form of response option yielded the best data quality.

METHOD

SAMPLE

To test the occurrence of ARB, an experiment was administered to a subsample of the Swedish Citizen Panel. A sample of 7,589 self-recruited panelists were invited to complete the study between December 5, 2022 and January 11, 2023. Reminders were sent the 14th and 19th of December to the respondents who had not yet completed the questionnaire. 4,444 answered more than 80 % of the questions, giving a participation rate of 59%. The sample was prestratified by sex (male, female), age (18–34, 35–49, 50–85 years), and education (low/middle education: less than 3 years of post-secondary education, high education: 3 or more years of post-secondary education).

PROCEDURE

Since the goal was to compare different formats of questioning, it was beneficial to keep the number of response options equal between the three groups, to ensure the difference was the actual wording rather than the number of response options. Previous research recommends using five fully verbally labeled response options over seven or eleven response options with only labeled endpoints for improved data quality of agree/disagree response options (Revilla et al., 2014). However, Menold (2020) measured an increase in data quality when using seven fully verbally labeled response options in comparison to five or seven response options with only endpoints labeled. Since the SOM Institute usually use five fully verbally labeled response options, five fully verbally response options were used in this experiment. Further, the verbalization of the middle alternative was consistent with the rating scale polarity in all three formats to increase reliability (Menold, 2021).

First, all respondents evaluated ten different political proposals where each respondent was randomly assigned to evaluate them using one of three question formats. Table 1 shows an example of how the questions appeared depending on to which group the respondent belonged. Respondents in the *Agree/Disagree* group evaluated the Likert scale using a Swedish version of agree/disagree. The second group named *Proposal* answered response options ranging from “Very good proposal” to “Very bad proposal”. These labels were chosen because they all appear in different forms in the annual SOM Institute surveys. Questions for the group *Agree/Disagree* and *Proposal* appeared in a grid-type format, with six questions on each side. Group *Agree/Disagree* and *Proposal* also had two the questions asked twice, but with reversed meaning. For example the question, “What is your opinion on the following proposal: Make the public sector smaller.” was also presented on a separate screen as, “What is your opinion on the following proposal: Make the public sector larger.” This was done to investigate if ARB occurred to a larger degree in one of the groups, where the respondent can agree to a contradictory statement. The third group, *Item specific* answered item specific question formats of the political proposals, where each response option was tailored to each individual question. The ten item specific questions appeared separately on one screen each. This group did not answer to contradictory statements as the response options utilized a bipolar scale, allowing respondents to answer to both sides of the proposal in each question.

Table 1. Response options for each group in a example question

| Agree/disagree | | | | | |
|---|---|--|---|--|---|
| Example question: | | | | | |
| "To what extent do you agree with the following proposal? Decrease the public sector " | Agree completely | Strongly agree | Partially agree | Agree a little | Do not agree at all |
| Proposal | | | | | |
| Example question: | | | | | |
| "What is your opinion of the following proposal: Decrease the public sector" | Very good proposal | Rather good proposal | Neither good nor bad proposal | Rather bad proposal | Very bad proposal |
| Item specific | | | | | |
| Example question: | | | | | |
| "Do you think the public sector should increase or decrease?" | The public sector should increase a lot | The public sector should increase a little | The public sector should neither increase nore decrease | The public sector should decrease a little | The public sector should decrease a lot |

ANALYSIS PLAN

Based on previous research, the response option using a Swedish translation of agree/disagree was predicted to yield the most ARB, followed by the proposal response option, with the item specific response option having the least amount of ARB. This was tested with OLS regression, assessing if one group had a higher tendency to agree to the political proposal than another. Another OLS tested the level of extreme response style in agreement which shows the tendency to agree fully to a statement. Agreeing answers to the

contradictory statement was also believed to show a higher level of contradiction for the *Agree/Disagree* group than for the *Proposal* group.

Finally, concurrent validity was used as a measure of data quality between the groups (Shaeffer et al., 2005). Concurrent validity estimates the amount of measurement error between a criterion question and a target question (the target question is the political proposal that had the manipulated question format). A more valid question format will yield less measurement error and a stronger association with the criteria that it should correlate with. The criterion questions were: 1. Income, 2. Left-right political orientation, 3. Trust in media, 4. Living in an urban or rural setting.

RESULTS

TENDENCIES TO AGREE

In contrast to the expected, respondents reported greater agreement when answering the IS format than when answering agree/disagree format ($b(4263) = -0.15$, $SE=0.00$, $p < .01$) or the proposal format ($b(4263) = -0.07$, $SE=0.0$, $p < .01$) (see Table 2). However, a reason for this result could be that the item specific questions measured both parts of each proposal, such as for example increasing or decreasing the public sector, while the other two groups only chose if it was a good idea to decrease the public sector or not. Therefore, the item specific format could receive different results and tendencies to agree than the other group due to the respondents being allowed to express a wider range of views, making it difficult to compare to the other groups.

Table 2. OLS regression predicting the amount of agreement to the political proposals depending on response option group

| | Political proposal question index |
|----------------------|-----------------------------------|
| Dummy Agree/disagree | -0.15*** |
| Dummy Proposal | -0.07*** |
| Constant | 0.53*** |
| Observations | 4,264 |
| R^2 | .31 |

* $p < .05$, ** $p < .01$, *** $p < .001$

Note. All questions were coded from 0 to 1. The highest agree response was coded to 1 (agree completely, very good proposal, or the most favorable option in the item specific question) while the lowest agree response (disagree completely, very bad proposal, or the least favorable option in the item specific question) was coded 0. 8 of the 10 questions were combined to make a total question index of the tendency to agree to the questions. Two questions were excluded from the index as they only had two response options in the *Item specific* group, making them difficult to compare between the different response options.

There was no significant difference to highly agree between the group *Agree/disagree* ($b(4261)=0.00$, $SE=0.00$, $p=.9$), and *Proposal* ($b(4261)=0.01$ $SE=0.00$, $p=.14$), and the reference group answering item specific questions (see Table 3). Further, it pointed in the direction of a slightly higher value for group *Proposal* which goes against the assumption that the agree/disagree response option would yield the highest level of agreement to the questions.

Table 3. OLS regression predicting the amount of extreme agreement responses to the political proposals depending on response option

| | Political proposal question index |
|----------------------|-----------------------------------|
| Dummy Agree/disagree | 0.001 |
| Dummy Proposal | 0.01 |
| Constant | 0.14*** |
| Observations | 4,264 |
| R^2 | .00 |

* $p < .05$, ** $p < .01$, *** $p < .001$

Note. All questions were coded from 1 for the highest agree response (agree completely, very good proposal, or the most favorable option in the item specific question) while the other alternatives were coded 0, this to see which groups had the most highly agree responses. 8 of the 10 questions were combined to make a total question index of the tendency to agree to the questions. Two questions were excluded from the index as they only had two scale steps in the *Item specific* group, making them more likely to have extreme responses than other response options and difficult to compare between the different response options.

CONTRADICTIONARY ANSWERS

If ARB occurs the respondent agrees no matter the content. Therefore, by reversing the same question and asking it twice on different pages, ARB can be assessed and compared between the *Agree/disagree* group and the *Proposal* group who have slightly different labels while both using a Likert scale. To make this clear the highly agreeable answers, which means to choose the most agreeable answer in the scale, were coded 1 and all other responses were coded as 0. As seen in Table 4, there were only 14 respondents in total who responded with the highest agreeable answer to contradictory questions. With less than one percent highly agreeing to both contradictory questions in either question format there was not any clear ARB measured. Therefore the data does not confirm the assumption that group *Agree/disagree* would have a higher level of ARB than the *Proposal* group based on contradictory answers.

Table 4. Amount of highly agreeable answers in 2 contradictory questions

| Groups | Contradictory question 1 (Increase/decrease public sector) | Contradictory question 2 (invest more/less in multicultural) |
|---|---|--|
| Agree/disagree | | |
| - Answered "Agree completely", both questions | 1 (0.07%) | 8 (0.58%) |
| - All other answers | 1,385 (99.03%) | 1,373 (99.42%) |
| Total | 1,386 (100%) | 1,381(100%) |

| | | |
|---|----------------|----------------|
| Proposal | | |
| - Answered “Very good proposal”, both questions | 2 (0.13%) | 3 (0.20%) |
| - All other answers | 1,502 (99.83%) | 1,499 (99.80%) |
| Total | 1,504 (100 %) | 1,502 (100%) |

Note. Highly agreeable responses here are respondents who has chosen “Agree completely” or “Very good proposal”, which is the highest value in the 5-step response option. For question 1 and 2 the exact same question was posed, only to reverse the meaning from increase to decrease or invest more to invest less.

CONCURRENT VALIDITY

The six models of OLS regressions in Table 5 showed different results for the groups depending on the political proposal, where the correlation to the criterion question sometimes increased and sometimes decreased for group *Agree/disagree* and group *Proposal* in comparison to group *Item specific*. In model d the correlation between the target and criterion variable increased for the *Agree/disagree* group ($b = 0.06, p = .03$), compared to the *Item specific* group, while it decreased in models a ($b = -0.06, p = .01$) and b ($b = -1.01, p = .06$). Further, in some questions it was insignificant which group the respondent belonged to for the concurrent validity between criterion and target question, with $p > 0.1$ in model b, c, e and f.

With such different results depending on the question it is difficult to say if the data quality increased in one specific response option. However, the item specific question format appeared to slightly strengthen the correlation between the target and criterion variable, giving better concurrent validity between concepts known to correlate.

Table 5. OLS regressions testing concurrent validity, comparing agree/disagree and proposal response options to item specific response options.

| | Political ideology (a) | Living rural or urban (b) | Living rural or urban (c) | Trust in media (d) | Trust in media (e) | Income (f) |
|-----------------------------|------------------------|---------------------------|---------------------------|--------------------|--------------------|------------|
| Q1: publicsector | 0.51*** | | | | | |
| agree/disagree | 0.07*** | 0.10* | 0.05* | -0.15*** | 0.01 | 0.01 |
| proposal | 0.02 | 0.04 | 0.02 | -0.11*** | 0.04** | -0.01 |
| agree/disagree*publicsector | -0.06* | | | | | |
| proposal*publicsector | -0.03 | | | | | |
| Q2:investrural | | 0.37*** | | | | |
| agree/disagree*ruralfunding | | -0.10 | | | | |
| proposal*ruralfunding | | -0.04 | | | | |
| Q3:carbontax | | | 0.22*** | | | |
| agree/disagree*carbontax | | | -0.06 | | | |
| proposal*carbontax | | | -0.02 | | | |
| Q4:statepress | | | | -0.56*** | | |
| agree/disagree*statepress | | | | 0.06* | | |
| proposal*statepress | | | | 0.04 | | |
| Q5:multiculture | | | | | -0.34*** | |
| agree/disagree*multiculture | | | | | -0.01 | |
| Proposal*multiculture | | | | | -0.03 | |
| Q6:benefits | | | | | | 0.13*** |
| agree/disagree*benefits | | | | | | -0.00 |
| Proposal*benefits | | | | | | 0.02 |
| constant | 0.27*** | 0.18*** | 0.43*** | 0.85*** | 0.6*** | 0.27*** |
| Observations | 4,451 | 4,343 | 4,348 | 4,443 | 4,446 | 4,115 |

$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note. a-f in the top horizontal row indicates each model and criterion variable used for each target question (political proposals Q1-Q6). Model a: Q1 proposal was: “Reduce the public sector” and the criterion variable was political ideology. Model b: Q2 proposal was “Increase government funding of rural areas in Sweden” and the criterion variable was living rural or urban. Model c: Q3 was “Increase the CO2 tax on petrol” and the criterion variable was living rural or urban. Model d: Q4 was “Reduce government press subsidies” and the criterion variable was trust in media. Model e: Q5 was “Invest more in a multicultural society” and the criterion variable was trust in media. Model f: Q6 was “Increase unemployment benefits” and the criterion variable was income.

CONCLUSION

The aim of this study was to investigate ARB in a Swedish context by testing three different response option formats in a randomized experiment, to see which response option was most at risk for increasing this type of bias. This was done to see how the SOM Institute could reduce measurement errors, increase data quality, and evaluate commonly used response options. In contrast to previous research, the results do not show any larger risk for ARB for the agree/disagree response options than good proposal/bad proposal or the item specific response options. Rather there was a larger general tendency to highly agree to a political proposal when answering item specific questions in comparison to the *Agree/disagree* and *Proposal* group. However, analyzing concurrent validity between concepts known to correlate, there was an indication that item specific response option could yield slightly better data quality, in line with previous research (Dykema et al., 2022). Since the questions in group *Agree/disagree* and group *Proposal* was presented in a grid type format while group *Item specific* received the questions one by one on separate pages, the effect of response option cannot be isolated alone in this study. Further, the response options in this study differed slightly between the three groups.

The IS questions had a bipolar scale where the agree/disagree and proposal questions were unipolar, which could lead to bigger differences between the response styles than if the questions had been kept more similar. More studies should be conducted investigating item specific response options compared to Likert scales and other commonly used response options to further develop ways to decrease acquiescence response bias and improve data quality in surveys.

REFERENCES

- Dykema, J., Schaeffer, N., Garbarski, D., Assad, N., & Blixt, S. (2022). Towards a reconsideration of the use of agree-disagree questions in measuring subjective evaluations. *Research in Social and Administrative Pharmacy*, 18(2), 2335-2344.
- Harzing, A. (2006). Response Styles in Cross-national Survey Research. *International Journal of Cross Cultural Management : CCM*, 6(2), 243-266.
- Höhne, J., & Lenzner, T. (2018). New insights on the cognitive processing of agree/disagree and item-specific questions. *Journal of Survey Statistics and Methodology*, 6(3), 401-417.
- Krosnick J. (1991). "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5:213-36
- Krosnick, J. (1999). Survey research. *Annual Review of Psychology*, 50(1), 537-567.
- Lelkes, Y., & Weiss, R. (2015). Much ado about acquiescence: The relative validity and reliability of construct-specific and agree-disagree questions. *Research & Politics*, 2(3), 2053168015604173
- Narayan, S., & Krosnick, J. A. (1996). Education moderates some response effects in attitude measurement. *Public Opinion Quarterly*, 60(1), 58-88.
- Meisenberg, G., & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education? *Personality and Individual Differences*, 44(7), 1539-1550.
- Menold, N. (2020). Rating-Scale Labeling in Online Surveys: An Experimental Comparison of Verbal and Numeric Rating Scales with Respect to Measurement Quality and Respondents' Cognitive Processes. *Sociological Methods & Research*, 49(1), 79-107.
- Menold, N. (2021). Response Bias and Reliability in Verbal Agreement Rating Scales: Does Polarity and Verbalization of the Middle Category Matter? *Social Science Computer Review*, 39(1), 130-147.
- Pasek, J., & Krosnick, J. (2010). Optimizing Survey Questionnaire Design in Political Science. In *The Oxford Handbook of American Elections and Political Behavior* (Oxford Handbooks of American Politics, p. The Oxford Handbook of American Elections and Political Behavior, 2010). Oxford University Press.

Revilla, M., Saris, W., & Krosnick, J. (2014). Choosing the Number of Categories in Agree–Disagree Scales. *Sociological Methods & Research*, 43(1), 73-97.

Shaeffer, E., Krosnick, J., Langer, G., & Merkle, D. (2005). Comparing the Quality of Data Obtained by Minimally Balanced and Fully Balanced Attitude Questions. *Public Opinion Quarterly*, 69(3), 417-428.

Tourangeau, R. (2018). The survey response process from a cognitive viewpoint. *Quality Assurance in Education*, 26(2), 169-181.



The SOM Institute is an academic organization located at the University of Gothenburg, Sweden. Since 1986 the SOM Institute conduct annual cross-sectional surveys among the Swedish population with a focus on Society, Opinion, and Media, as well as administering the web panel called the Swedish Citizen Panel. The annual surveys and the web panel both function as infrastructures, enabling researchers and public organizations to effectively collect research and opinion data in collaboration with researchers at the SOM Institute.

In order to strengthen contemporary research on Swedish society, as well as to contribute to international methodological development, the SOM Institute frequently publish notes on methodological research. These are made available for scholars and the public alike at <https://www.gu.se/en/som-institute/publications/research-on-survey-methodology>

For more information, please contact us at: info@som.gu.se