GESIS ᐧ P.O. Box 12 21 55 ᐧ 68072 Mannheim ᐧ GERMANY

Mannheim, 2023-03-07

Titel: "How to reduce Item Nonresponse in Face-to-Face Surveys? A Review and Evidence from the European Social Survey"

Dear Dr. Grönemann,

Thank you for submitting your manuscript "How to reduce Item Nonresponse in Face-to-Face Surveys? A Review and Evidence from the European Social Survey" to MDA.

It has been reviewed by 5 reviewers and a MDA editor. On the basis of their reports, we are pleased to inform you that the Editorial Board of MDA has decided to accept your manuscript for publication, conditional on you satisfactorily addressing the issues raised by the editor and the reviewers. It has been assigned the status of "*major revision*".

The advices from the reviewers varied from minor revision, though major revisions, to reject. Based on the advice of the MDA editor, we decided to ask for a major revision. All reviewers commented on the language and one very helpful reviewer even added an annotated file with suggestions for language changes in your rewrite. We have added this file. Also it was suggested to ask a native speaker to go over the final manuscript. Several reviewers commented that the manuscript was unnecessary complicated and long and suggested that you restructure it. Finally, it was suggested to incorporate the important Tables directly in the manuscript and not in a supplement. When rewriting the manuscript, we ask you to focus on this and take the detailed comments of the reviewers into account.

Please find the reports of the editor and the reviewers enclosed. We look forward to receiving your revision. When submitting, please include a separate detailed list of how you acted on the suggestions made by editor

and reviewers. Also note that a second round of evaluation of the manuscript by the reviewers might be necessary.

If you feel that some of the comments do not address important aspects of the manuscript or are uninformed either substantively or methodologically, please feel free to point that out to us.

Thank you for giving us the opportunity to consider your work. We hope that you will continue to do so in the future.

Sincerely,
Edith de Leeuw
Associate Editor

Report Reviewer A

This paper provides a nice review on item non-response, and an empirical study exploring strategies to reduce item non-response. Although the author acknowledges that 'don't know answers' for some respondents actually be the most applicable answers (i.e. "because they do not know about the content of the question or are unable to remember an event"), initially this is not mentioned, and also when discussing the four constructs related to the probability of item nonresponse, this issue is ignored. I think the author could put more emphasis on the difference between satisficing and privacy-originated sources of item nonresponse on the one hand, and a substantive don't know answers on the other. When a respondent really does not know the answer to a question, the answer with highest quality is a don't know answer. A clearer distinction can also be made for attitude versus behavioral items. I do not agree with the author's argumentation that in cases where DKs are genuine, this is "a product of the respondents' ability", since especially in attitude items, it may often happen that before the survey respondents have never thought about the issue being questioned. I would not say that is ability, and a don't know answer would in fact in that case be a genuine substantive answer. In addition, it is important to note that attitude items often have higher levels of item non-response than factual items on the same topic, reflecting opinion censoring, which is actually higher among higher educated respondents than lower educated respondents (Piekut, 2021). Therefore, I suggest that the type of item, attitudinal or factual, is also taken into account in the analysis of the ESS data.

Since length of the interview may vary depending on the applicability of questions, I am not sure what the mechanism here is: respondents for whom more questions are applicable will have longer duration and are different from those who have fewer applicable questions due to filter questions (i.e. in the ESS these are: respondents who indicate they rarely use a smartphone, are not eligible to vote, did not vote in the most recent election, do not feel close to a certain political party, do not consider themselves to belong to a religion or denomination, do not consider themselves to be part of a group that is discriminated against, etc). Since interview length cannot be disentangled from respondent characteristics, it is completely unclear what is causing the positive effects of interview length on item nonresponse. Again a clear distinction between attitude and factual items might provide some clarification.

The manuscript needs a thorough grammar (and spelling!) check as there are quite some grammatically incorrect sentences, like "Over time, respondents might lose the ability to concentrate and motivation".

References: Piekut, A. (2021). Survey nonresponse in attitudes towards immigration in Europe. *Journal of Ethnic and Migration Studies*, *47*(5), 1136-1161.

Report Reviewer C

- Relevant and important topic

- Contributes to the advance in the field

- Nice review of the literature in the first part.

- The second part (e.g. discussion) needs to be more developed, including more literature

- Be more careful in your tone. Use "seems", "I propose" or similar more often. I have added a few suggestions in the text, but it is necessary to revise the whole article in this respect.

- E.g. in the abstract: instead of "I extend", I suggest you use something like: "I propose the extension of

the model….."

- Concerning the style, I also especially suggest you read a few published articles by the journal mda to adapt your style accordingly

- I suggest assistance by an English proofreader, specifically a scientific proofreader (I have marked a few bits of texts yellow where I had doubts about the style, or made a suggestion for a change, but it is also in this respect necessary to revise the whole article).

- I have attached the text of the article with comments and suggestions.

Good luck!

Reviewer D

I appreciate this opportunity to review, "How to reduce Item Nonresponse in Face-to-Face Surveys? A Review and Evidence from the European Social Survey." The paper suggests some attributes of survey questionnaires, interview protocols and settings, and interviewers characteristics that may be associated with item nonresponse in ESS. While an important topic, the manuscript as currently written does not appear to be suitable to be a scholarly publication for the following reasons.

1. Structure of the paper

The motivation for this particular study is not clearly provided in Introduction. Rather, Introduction mixes in the findings from the study. Moreover, Introduction says, "it is important to understand the processes that can lead to missing data," the paper delivers only a small attributes related to the processes. The abstract states, "I extend the satisficing model," but the paper fails to deliver this extension.

1. Conceptualization of item nonresponse
    1. The author notes, "I have thus identified four constructs related to the probability of item nonresponse under which most research can be subsumed: cognitive ability, task difficulty, motivation and privacy concerns." I do not agree with this conceptualization, as it makes identifying the source difficult. Rather, it would be helpful to distinguish sources of item NR overall or within each construct the author proposes. For example, one can conceptualize respondents, questions, questionnaires, interview protocols/settings and interviewers as sources of the item NR and attributes of each source affecting item NR of specific types of questions as well as all questions in a questionnaire. For example, motivation can be a product of all sources. Lumping these sources into the constructs is likely to make coming up with strategies for reducing item NR extremely challenging.
    2. Privacy concern is hypothesized to come in in the editing process, but I argue it may also come during the comprehension stage. For example, to sexual behavior questions, respondents may react, "This is really private. Why are you asking this?", and refuse to answer before coming with answers.
2. Role of the item NR probability
    1. For the formula given on p.7, the author states, "I therefore suggest extending the formula by Krosnick". Provide the exact citation of Krosnick.
    2. In order for the formula to make sense, difficulty, ability, motivation and privacy need to be quantified in the probability format. How is this done? And why max function?

3. This probability is not used in the analysis at all, making its role uncertain for the paper.
3. Data and Methods
    1. I recommend ESS to be better introduced. Not all readers know about it.
    2. What is the showcard strategy in ESS? Provide the context of its protocols and use contexts.
    3. Why limit respondent characteristics to education and age?
    4. "Negative binomial regression with interviewer fixed effects" on p.15. Unless reading the supplemental materials, readers will not know why negative binomial model is used. Also, why interviewer effects are analyzed as fixed?
    5. "Standard errors are clustered by interviewer." on p.15 does not make sense.

"Missing data are deleted listwise." on p.15. What is the missing rate of each variable in the model? For example, some missing rates can be inferred from supplemental material -- duration is missing on 4450 cases, equating to almost 10% missing, which is not trivial. Any insight into this nontrivial missing rate? Combining all the variables in the model, how many cases were subject to the listwise deletion? How are the deleted cases similar or different from retained cases?

Reviewer E

I read the paper intitled 'How to reduce item non-response in face-to-face surveys? A review and evidence from the ESS' with a lot of interest since this is one of the issues that always arise when designing questions: will respondents collaborate and answer to the best of their abilities? The author made a good literature review tapping the most relevant issues that influence non-response behavior and present an adequate data analysis model.

(lines 35-40) Your novelty is the introduction of 'privacy concerns' in the model developed by Krosnick, so you should give more visibility to it. Therefore, my first suggestion is to make clearer and more evident the novelty you are bringing to the field. What in your study teach us that we didn't already know from the available literature?

Be careful with contradictions. For instance, in lines 49-51 (and 368-3689) you write: "Matching respondents' and interviewers' gender and age seem not to be efficient tools to reduce item nonresponse."; but in lines 234-236 it is said that "Vercruyssen, Wuyts and Loosveldt (2017) find less item nonresponse when matching age of interviewers and respondents and matching gender reduces item nonresponse for males and increases it for females"; and in lines 276-277, "Previous research suggests that respondents might be more willing to answer to socially similar interviewers."

One 'variable' that was not considered in the matching between interviewer and respondent is the ethnic/national origin. This is a very relevant issue, for instance for surveys on attitudes towards immigration, discrimination, racism, etc. And it has a strong link with privacy concerns in what relates to **social desirability**. I would suggest you introducing this topic in the literature review. For instance, in face of an interviewer with an accent that reveals his/her national origin, a native respondent may avoid giving her/his true opinions about immigration issues and opt for a non-response or for a politically correct answer (because it is also known that many people don't like to say 'I don't know').

Regarding the formula (lines 149-152) the explanation is the following (the underline in mine): "*The probability of a nonsubstantive answer by the respondent is given by the task difficulty divided by the product of ability and motivation or the respondents' privacy concerns, whichever is higher*".

Please explain this better; I don't understand what the underlined condition means.

 Lines 193-195: by 'changes in response categories' do you mean 'changes in response scales'? Routing and filtering are associated with higher non-response in PAPI, using CAPI mode may reduce routing and filtering problems to zero.

Lines 207-213. Regarding ability there is much more to say. The environment of the interview is important, but ability can (and must) be assessed before going into the field by a pilot study, in order to test the comprehension of the questions, the rates of item non-response, the impact of the order of the questions (contamination), among other aspect,

Lines 236-237: "Matching education has no effect (Vercruyssen, Wuyts and Loosveldt, 2017;…" not quite…what the authors say is that "the effects of education level (mis)matching could unfortunately not be tested." What had no effect was the education level of the interviewer.

In your results you find that "With every additional thirty minutes interview duration, the number of DK rises by about 21% and the number of refusals rises about 8% on average" (lines 336-337). Since we are talking about a specific questionnaire (ESS9) I suggest that you check what happens around the middle of the questionnaire (the average interview time is around 60 minutes). Maybe the nature of the questions has also an important impact on DK and refusals. To draw conclusions about the 30minutes it would be needed to test it with different questionnaires.

The discussion section would gain if the novelty of this study was stressed

Finally: the tendency of many surveys (due to high costs and low response rates) is to change from face-to-face to self-administrated modes (PAPI and CAWI). Consider discussing a bit this incoming shift and how your final recommendations could be adapted.

The paper needs a review from a native English speaker.

I hope my comments help the authors to improve the paper and I wish you all good luck.

Reviewer F

General Comments

This manuscript provides a review of the literature on item nonresponse in surveys and an empirical analysis of correlates of item nonresponse in the ESS. Both the literature review and the empirical study are adequate. The connection between the literature review and the empirical study is relatively weak but I do not see how this can be strengthened, because in the end the analysis must be restricted to variables available in the ESS. The authors could mention this in the transition from the review to the study.

I like the supplemental information and analysis setups in the additional materials.

Detail

Section 5.4: note that formula (1) is a multiplicative model which implies measurements on equivalent ratio scales. These are not available; hence the analysis is carried out as a regression model. Please mention this.

In addition to Figure 1 I would like to see the regression table in the text body, not in an appendix