Reviewer's report

Title: Applying an extended theoretical framework for data collection mode to health services research

Version: 1 Date: 9 March 2010

Reviewer: Caroline Roberts

Reviewer's report:

- Major Compulsory Revisions

I enjoyed reading this paper. It is a topic I am very familiar with and I believe it is particularly important to bring the literature relating to mode effects in quantitative data collection to the attention of researchers working in domains outside of the social sciences, where factors influencing data quality are perhaps better understood. For this reason, I am keen to recommend this paper for publication and I am advising that the paper be accepted after revision, but I do feel it needs more work than the minor revisions listed below.

The paper was generally very clearly-written and structured and the scope of the literature covered was broad. The model presented provides a concise framework for the reader to begin to grasp the complexity of the problem of mode effects and the various characteristics of data collection modes that can influence the quality of the data collected. I also liked the extension of the model to include additional features not addressed in its original form by Tourangeau et al.

Nevertheless, I was left feeling somewhat disappointed by my reading of the paper. Perhaps most importantly, I felt that it did not quite achieve what it set out to, in terms of highlighting the implications of the findings of research on mode effects for researchers working in the health domain. To address this, I think it would be nice to see an expanded version of the section on pages 21 and 22, perhaps incorporating and building on some of the comments raised at the bottom of page 20 about the specific features of research in a clinical context that might render some of what we know about mode effects less applicable. Is there any existing research on this? This seemed to me to be one of the most interesting aspects of the paper, but was never fully developed. If you were to discuss in a more systematic way the specific features of clinical and health-related research that make it different from social surveys, it would allow for a more systematic analysis of the literature - particularly in terms of highlighting where new research is needed. You mention the need for research in various health-related areas, and again I felt this could have been expanded on more - either in the health-related section, or in its own section (e.g. ‘Recommendations for future research’). Without building more on this unique contribution that the paper has the potential to make, I feel that it doesn’t offer much beyond what reviews available elsewhere are able to offer - except, of course, by bringing the literature to the attention of a new audience.
I was also concerned that the paper started to lose focus as it progressed. By page 23 and 24, where you start to discuss again the characteristics of modes that might influence mode choice, I started to question what the main point of the literature review was. Is it designed to help researchers in the health domain decide how to choose a mode of data collection? Is it about how to assess the quality of data collected in surveys? Is it trying to highlight the problems associated with comparing data from studies that have collected data in different modes? Or is it about the problems involved in using a mix of data collection modes in a single study? These are all important issues, and it is possible that you are attempting to address all of them here. The problem is that it isn’t always clear which one is the primary focus, and so the text becomes confusing to follow. I think it would help to really distinguish between these issues at the start of the paper, and try to make it clearer what the implications of the literature review are for each one (or for those you choose to focus on, if you want to select just a few).

This point about clarifying the focus is also important because at times new issues crop up but are never fully expanded on. For example, you mention possible advantages of switching modes in a single study on page 12, but you never really discuss or highlight the problems involved in mixing modes, though these are implied throughout. Similarly, you raise the problem of mode effects on survey participation (the fact that modes influences people decisions about whether to participate), but this is quite a separate issue from the effect on response quality (except where the mode acts as a ‘common cause’ of both nonresponse errors and measurement errors, which you mention at the top of page 13). It would be good to elaborate more on some of these points, because at present, I’m concerned that you run the risk of oversimplifying some important issues.

- Minor Essential Revisions

Page 5, 2nd para, lines 3 and 4: give references here
Page 8, 2nd para: Is this correct? My understanding is that Tourangeau et al. distinguish 3 features rather than 4 - there seems to be some overlap here.
Page 9, end of 2nd para: include a reference for satisficing
Page 9, 2nd line from the bottom: there’s a grammatical problem here
Page 10, end of para 1: check ‘an exhaustive listings’
Page 14, end of para 2: do you have any references relating to ‘pace’ that you can insert here?
Page 15, top para: similarly, a reference relating to multitasking here?
Page 15, 2nd para, 5 lines from bottom: should this be ‘pauses’?
Page 15, 2nd para, last sentence: do you mean the satisficing model? do you mean specifically in the health domain? There are several studies that have applied it in the social sciences (including some by myself and colleagues!).
Page 19, top of page: it’s not really clear here whether you’re talking about reliability across measures in a survey or across different surveys. It seems all three mode features could influence both reliability and validity.

- Discretionary Revisions

Page 6, last sentence before Discussion: is the ‘systematic method variance’ referred to here different from biases such as social desirability and acquiescence bias?

Page 6, third to last line: ‘greater efficiency’ - compared to what?

Page 7, 1st para: you describe differences between face-to-face and telephone questionnaires, but it doesn’t seem clear that the differences arise as result of the mode characteristics. The last sentence also seems a little vague - what exactly do you mean by ‘how the survey is appraised’? Can you expand on this? It’s not clear what is implied.

Page 7, 2nd para, 1st line: clarify what the ‘original distinction’ is.

Page 8, subheading: it’s not really clear at this point whether you are talking about the effect of mode on participation in surveys or on response quality.

Page 9, 2nd para: there is considerable evidence that different characteristics of the survey can influence legitimacy so I wonder whether it is more objective than you imply. Equally, there are individual differences influencing burden too - e.g. intelligence, topic-related knowledge, interest/ involvement in the survey topic, etc.

Page 9, top of para 3: which response quality indicators?

Page 11, last para: again I feel it’s worth mentioning that there are different ways of establishing legitimacy in surveys but that the mode may influence how it is perceived.

Page 12, end of para 2: How do leverage and salience interact with mode and what is the outcome? I feel it isn’t quite clear why this is important given the discussion up to this point focuses mainly on the effect of mode on response quality.

Page 12, half-way through last para: it’s not clear what ‘switching between modes’ means here - do you mean to follow up nonrespondents as in a sequential mixed mode design? This could equally have the effect of irritating sample members!

Page 17, end of para 2: does this mean the environment at the time of questionnaire completion?

Page 17, section on acquiescence: I’m not entirely convinced by the decision to treat acquiescence differently from other satisficing effects, although I take the point that it shares similarities with social desirability bias. Can you make a
stronger argument here?

Page 19, 1st para, last sentence: it’s not totally clear what this means.

Page 20, top of page: the important point here seems to be that ‘objective’ behavioural measures are in principle, objectively verifiable.

Page 20, para on respondent role: is it true that this research has mainly been on student samples? I thought there was quite a lot of research from population surveys?

Page 20/21: section on respondent characteristics: this seems a bit speculative in places. As mentioned, legitimacy depends on ‘objective’ things like survey topic, sponsor, etc., rather than respondent perceptions. Similarly, the likelihood of satisficing seems to depend partly of the design of the questionnaire (and as you’ve mentioned, the mode), so I’m not sure you can speculate about differences between different survey populations.

Page 21, 2nd para: self-reports of sexual activity also depends on respondent sex (men overreport, women underreport).

Page 22: last para: questionnaire construction only really becomes important when you are combining modes in a single study, but this hasn’t really been the focus of this paper.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests.