Theme 3

Measurement issues in web surveys: An overview of opportunities and challenges

By Lisa Calderwood, Institute of Education, l.calderwood@ioe.ac.uk

Synthesis paper for NCRM Web Survey Network opening conference

Introduction

One of the key challenges in conducting general population surveys on the web is the impact on data quality. There are concerns about whether web based data collection is able to deliver high quality survey data, particularly for complex surveys, when used in a single-mode context and about measurement equivalence when the web is used in a mixed-mode context. However, the web also offers many possibilities for enhancing survey measurement and the growth of new technologies provides opportunities for innovation in data collection. This paper gives an overview of measurement issues in web surveys, outlines the challenges of collecting complex data on the web and those relating to mixed-mode surveys and highlights the opportunities for using new technologies on web surveys.

Measurement issues in web surveys

The web is a remote self-completion data collection mode. This means that the effects of interviewers on survey measurement are eliminated. However, this is not necessarily a good thing; interviewers can have both positive and negative effects on measurement error. This depends largely on the skills and characteristics of the interviewer themselves, the nature of the questions and whether the interviewer is present, as in face-to-face surveys, or not, as in telephone surveys. For example, interviewers can provide motivation, clarification or additional help to respondents which can lead to improved measurement, particularly on complex questions. Similarly the presence of an interviewer can ensure that some types of questions are answered truthfully e.g. that respondents don't look up answers to knowledge questions and that the answers they give are their own i.e. that other people don't answer on their behalf or tell them what to say. Some of these benefits of interviewer-administered questionnaires can also accrue to non-remote self-

completion methods e.g. where a sub-section of a longer interview is carried out by self-completion, but with the interviewer on hand to provide assistance if needed. On the other hand, there are several types of questions where the presence of an interviewer can have a detrimental impact on measurement. For example, there is clear evidence that sensitive questions and questions which are susceptible to social desirability bias obtain more accurate answers when administered using a self-completion mode (e.g. Tourangeau et al, 2000) and some research has shown that web surveys provide more accurate answers than other self-completion modes (e.g. Kreuter et al., 2008). Although survey interviewers are trained to read questions exactly as they are written, in practice there may be some interviewer level variation in how they are asked and this can lead to non-systematic interviewer effects on measurement.

Another key feature of self-completion surveys is that questions are presented visually to respondents who read the questions themselves. Visual appearance, in term of layout and formatting, is therefore very important for web surveys. Web surveys offer many advantages in terms of visual presentation compared with paper self-completion questionnaires. For example, it is possible to embed images, graphics and videos in web surveys. There is established best practice in web survey design c.f. “Designing effective web surveys” (Couper, 2008) and most major survey agencies have standard web survey templates which incorporate this best practice. Nevertheless, there is lots of evidence that small changes in layout, design and question formatting can impact on measurement quality in web surveys (e.g. Couper et al, 2004; Dillman et al. 2009).

Web surveys also have all of the benefits of computer-assisted interviewing (CAI). In relation to survey measurement these include; automated routing (skip patterns) so questions are only asked to those for whom they are relevant, textfills (substitution) to vary the wording of questions and/or response options depending on prior responses and in-built range and consistency checks to carry out real-time data quality checks. These features of CAI surveys are particularly important for complex questionnaires. The computerised nature of web surveys also means that they can be interactive i.e. they can be programmed to respond to the behaviour of the respondent and provide feedback on how they are doing. This is often in written form e.g. error messages and progress bars, but can include e.g. using avatars to simulate the presence of an interviewer. Most online social surveys do not really harness the full potential of the web for adding value in this area e.g. through the use of audio, images, enhanced visual design, interactivity and animation. This is primarily because to date there is little evidence that these enhanced features can improve survey measurement.

In the context of single-mode internet panel survey in particular, the remote self-completion nature of web data collection may have a negative impact on data quality e.g. through respondent satisficing , straight-lining and other short-cutting

behaviours, particularly when they asked to complete questionnaires on a frequent basis.

Key questions for discussion:

- Can we, and should we, do more to harness the possibilities of the web to make our surveys more visually stimulating, interactive etc?
- How concerned should we be about the remote nature of web data collection leading to 'bad reporting behaviour' by respondents?

Collecting complex data using the web

In general, the use of the web as a primary or sole mode of data collection is still relatively rare in large-scale, complex surveys in the UK and internationally. Carrying out these types of surveys on the web presents distinct and complex measurement challenges.

Many questionnaires used on large-scale surveys are complex i.e. they make **extensive use of routing, textfills, and checks**. Although the computerised nature of web surveys facilitates this complexity, the remote nature of the web mode can pose challenges as interviewers are not available to guide respondents through difficult sections of the questionnaires. This usually requires some simplification of survey questions and in particular instructions and help screens. Less extensive consistency checking is generally carried out on web surveys than in interviewer administered modes, which may lead to poorer quality data. In addition, interviewers may also be better able than respondents to resolve inconsistencies where they do arise. In particular, checks can be problematic when they relate to answers given much earlier in the questionnaire, as it can be challenging for respondents to skip back to correct earlier answers whereas interviewers are generally better able to navigate within questionnaires.

For surveys which involve **multiple respondents per household**, adequately capturing the inter-related nature of these questionnaires can be challenging in a web context. There is also a concern that questionnaires may be completed, either accidentally or deliberately, by someone other than the intended respondent.

Many large-scale surveys include full **household composition grids**. This involves the enumeration of everyone in the household, the collection of identifying information about each person e.g. name, sex and DOB and the collection of relationships between household members. This is a complex task where the presence of an interviewer may be expected to improve data quality e.g. ensuring that sufficient identifying information is collected for each person, prompting for additional household members, answering queries about household definitions and relationships. In longitudinal surveys, household grids are often fed-forward from prior waves, which can lead to further complexity e.g. difficulties identifying prior

household members, especially if the grid is being completed by a different person, and differentiating returning household members from new people. It may be challenging to collect this information fully and accurately using the web. On Wave 5 Innovation Panel of *Understanding Society*: the UK Household Longitudinal Survey, the accuracy of household grids reported on the web was assessed by re-collecting them over the phone for a sub-set of respondents. Their findings will be available later this year.

On many CAI surveys, the benefits of the computerised context is harnessed to allow interviewers to carry-out **coding in real-time during the interview** e.g. using database look-ups and/or validation software for occupation and industry, medicines, countries, addresses etc. Although some types of look-ups are straightforward (e.g. addresses, countries), specialist coding (e.g. occupation, industry, medicines), is often not and interviewers are given special training on this. A recent ONS pilot test included a comparison of coding done by respondents and by interviewers (Portanti and Wilson, 2012). They found high rates of agreement (95-99%) for coding of workplace, country of birth and nationality but much lower agreement rates for SOC coding of occupation (68%) and SIC coding of industry (67%). There have also been a number of experiments in the US on complex coding e.g. on prescription medicines (Couper et al., 2012).

On most longitudinal surveys, extensive use is made of **dependent interviewing** i.e. feeding forward information from prior waves. This presents particular challenges for web surveys. For example, more extensive cleaning of text responses e.g. job title is required and situations where the respondent disputes the fed forward information can be more problematic to deal with on the web. In addition, it is possible that respondents may have more concerns about their answers from previous waves being fed forward in a web survey than on interviewer-administered surveys. Many longitudinal surveys also collect **event history data** about e.g. employment, housing, relationship histories. These usually use dependent interviewing, and often use visual calendars to input and/or display start/end dates and durations in particular states to respondents. It can be a very complex task to collect this event history information accurately, particularly for respondents who have made lots of transitions. In the context of a web survey, it may be challenging to collect this information accurately, and to design a calendar which provides a clear and helpful visual display and is easy to use. **Tina Glasner's presentation will provide evidence about the viability of using event-history calendars on the web.**

Key questions for discussion:

- Is it possible to deliver complex surveys on the web for the general population?
- Are there sub-groups of the population for whom this may not be feasible?

Collecting data using the web in a mixed-mode context

For most large-scale surveys in the UK and internationally on which the web has been used as a primary mode of data collection, this has been in combination with other modes. There are additional considerations when designing questionnaires for the web in the context of a mixed-mode survey. The equivalence of measurement between different modes, or the absence of it, is a key concern in this context. There is an extensive body of research about mode effects on survey measurement. There is a broad consensus in the literature about which types of questions are most susceptible to mode effects (see e.g. Roberts, 2007 for a review). Where surveys are designed as mixed-mode from the outset, a uni-mode approach to questionnaire design, which aims for equivalence in question wording and measurement between modes, can be adopted. However, many mixed-mode surveys are not designed in this way from the outset. When a secondary mode is introduced to an existing survey, uni-mode principles are usually applied to try to minimise mode differences. This can be done more successfully where there is scope for changing the wording of questions in the primary mode, but this is not always possible precisely because of concerns about wording changes introducing discontinuity in measurement.

However, even with good uni-mode questionnaire design from the outset, there is evidence that measurement differences between the web and interviewer-administered modes persist, and in some cases they may be relatively large. The extent of these differences is likely to depend on the nature of the survey questions. For some types of surveys e.g. attitude surveys, mode effects on measurement may be much larger than others e.g. surveys of non-sensitive behaviours. The importance of these mode effects on measurement is also context-specific. For example, repeat cross-sectional surveys which produce key government statistics will be very concerned about even small mode effects on measurement whereas ad-hoc surveys may be less concerned. It should also be noted that mode effects on measurement are also important for web-only surveys in order that their findings can be legitimately compared with other surveys.

In a mixed-mode context, mode effects on measurement may also be confounded by compositional differences between those responding in different modes, which can due to sampling and coverage issues as well as differential response by mode, and apparent mode effects are attenuated once these are taken into account. For example, a mixed-mode experiment using online and telephone data collection on the International Crime and Victimisation Survey showed that most mode difference in measurement did not remain significant after controlling for these other mode-related effects (Guzy and Leitgob, 2011). Differential mode choice may also attenuate the impact of mode on measurement e.g. if respondents who choose to complete the web survey are those who are most able to provide accurate answers without assistance and respondents who don't complete the web survey are those who most benefit from interviewer assistance.

Overall, it seems clear that not all mode differences in measurement are preventable. Where they are considered important, then the issue becomes what can be done to adjust for them. **Peter Lugtig's presentation will focus on adjustment methods for mode differences.**

Currently the most promising approach to this problem is to incorporate a randomly assigned single mode sub-sample in order that measurement error can be estimated and appropriate adjustments made at the analysis stage and/or when the data are combined. However, this is not always possible, and in practice data collected in different modes will often be combined and analysed without consideration being given to this issue.

Key questions for discussion:

- How should survey practitioners trade-off mode differences in measurement against the benefits of mixing modes in surveys?
- What level of measurement differences between modes is acceptable and how can this be assessed?

Web surveys and new technologies

In recent years, there has been rapid increase in the ownership of web-enabled portable devices i.e. Smartphones and tablets, and growing use of social media e.g. Facebook and Twitter. This means that increasingly respondents are likely to use these devices for completing web surveys. Although there is some research in this area, there is a relative lack of evidence about the optimal way to design web-surveys for these devices and the measurement implications of web-surveys being carried out on these devices. There is also increasing interest in conducting web surveys using embedded applications, though social survey practice and research in this area is at a relatively early-stage. In practice, complex probability surveys which are carried out on the web do not tend to be optimised for these devices, and Smartphones in particular are unlikely to lead to high-quality data for these types of surveys due to small screen and keyboard sizes. **Mick Couper's presentation will cover carrying out web surveys on mobile devices.**

Increasingly, these devices mean that it is possible to connect to the internet anytime and anywhere and people's lives are increasingly being documented electronically through their use of social media. The cutting edge of survey practice internationally is concerned with leveraging these new technologies to improve measurement. This ubiquitous connectivity provides opportunities for respondents to both report their behaviour, feelings, attitudes, etc in real-time, rather than retrospectively during a survey interview, and for these things to be measured directly in real-time. For example, GPS can be used to track the whereabouts of respondents. Social media data generated by respondents themselves can also be gathered and analysed and

these technologies also facilitate the collection of different types of digital data e.g. photos, videos, audio-recordings. In principle, this could lead to improved survey measurement through reduced reliance on respondent self-report and recall. Indeed, these technological changes are arguably leading to a paradigm shift in the way that data is captured and the beginning of a 'post-survey data capture era' (Link, 2011).

Although there are undoubtedly ways in which survey measurement can be enhanced in this way, it seems unlikely that these technological changes will obviate the need for large-scale general population surveys in the near future. However, the growing use of these technologies may imply that we should re-think the 'traditional' model of surveys as a periodic one-off data collection exercise based largely on self-report and recall and consider other approaches to meeting our information requirements e.g. shorter, more frequent data collection exercises, supplemented by direct measurement and/or social media data.

Key questions for discussion:

- How can we harness the advantages of new technologies without compromising data quality?
- Are we really about to enter a new 'post-survey data capture era'? Is the 'traditional' survey really an endangered species?

References

Couper, M.P. (2008). *Designing Effective Web Surveys*. New York: Cambridge University Press.

Couper, M.P, Tourangeau, R., Conrad, F.G., and Crawford, S. (2004). What they see is what we get: Response options for web surveys. *Social Science Computer Review,* 24(2), 227-245.

Couper, M. P., Zhang, C., Conrad, F., and Tourangeau, R. (2012). Database lookup in Web Surveys. Presentation to 6th International Internet Survey Methodology Workshop.

Dillman, D., Smyth, J., and Christian, L. (2009). *Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method (3rd edition)*. Hoboken, NJ: Wiley

Guzy, N. and Leitgob, H. (2011). Mode effects in Online and Telephone Based Victimisation Surveys. Presentation to the Stockholm Criminology Symposium, 2012.

Kreuter, F., Presser, S. and Tourangeau, R. (2008). Social Desirability Bias in CATI, IVR and Web Surveys: The Effects of Mode and Question Sensitivity. *Public Opinion Quarterly,* 72 (5), 847-865.

Link, M.W. (2011). Evolving Survey Research: New technologies and the next steps forward. Webinar for the American Association of Public Opinion Research.

Portanti, M. and Wilson, L. (2012). Design, Implementation and Testing of LFS Questionnaire Features in an Online Mode – Results from the 2010/11 Internet Pilots. *Social Survey Methodology Bulletin.* Office for National Statistics, 70 3/12.

Roberts, C. (2007). Mixing modes of data collection in surveys: A methodological review. Economic and Social Research Council – National Centre for Research Methods. NRCM Methods Review Papers NRCM/008.

Tourangeau, R., Rips, L., and Rasinski, K. (2000). *The Psychology of Survey Response.* Cambridge: Cambridge University Press.