ESRC National Centre for Research Methods Briefing Paper

Mixing modes of data collection in surveys: A methodological review

Caroline Roberts

Senior Research Fellow

Centre for Comparative Social Surveys, City University, London

MARCH 2007

ESRC National Centre for Research Methods

NCRM Methods Review Papers

NCRM/008

Abstract

In this paper I provide a review of the literature related to mixing modes of data collection in surveys. The paper is structured in the following way: I begin with an overview of the range of mode choices available to survey researchers and their advantages and disadvantages with respect to a range of criteria, including their impact on data quality (so-called 'mode effects'). Increasingly, survey designers are exploiting the potential offered by using a combination of data collection modes, either to offset the weaknesses of a particular mode with the strengths of another (Dillman, 2000) or, for example, to try to reduce the overall costs of fieldwork. The paper describes the major challenges involved in mixing modes, focusing in particular on the problem of mode effects on measurement error – or how people respond to survey questions. Modes of data collection appear to affect respondents' answers by influencing either a) the amount of effort needed to answer the question; or b) the respondents' willingness to answer questions honestly (Holbrook et al., 2003). The first type of influence can lead to a range of response errors referred to as 'satisficing effects' (Krosnick, 1991), while the second type of influence can lead to a tendency to give socially desirable answers. Mode effects have implications for the comparability of data collected in different modes; understanding their causes is, therefore, an important step in developing ways of reducing the negative impact of mixing modes on data quality. After describing some mechanisms through which modes result in measurement error, I discuss what, if anything, researchers can do to tackle the problem.

Key words: data quality; mixed mode; survey error; mode effects; satisficing; social desirability bias

Acknowledgements

This review was conducted as part of a Short Term Project funded by the ESRC's National Centre for Research Methods on 'Developing mixed mode methods in comparative social surveys'. I wish to thank Roger Jowell, Eric Harrison and Gillian Eva for their invaluable comments on earlier drafts of the paper. The paper draws on some excellent existing reviews of the literature in this area, including De Leeuw (2005), Dillman (2000), Tourangeau, Rips and Rasinski (2000) and Holbrook, Green and Krosnick (2003), to acknowledge and credit just a few notable influences.

Contents

1 Modes of data collection in surveys – Options for researchers

Survey researchers can choose from a wide range of methods of collecting information from sample members. The most commonly-used methods or 'modes' of data collection include (1) *face-to-face interviewing*, in which an interviewer typically visits a respondent in their home or at their place of work and administers the questionnaire in person (either using paper-and-pencil personal interviewing (*PAPI*) or computer-assisted personal interviewing (*CAPI*) in which the questionnaire has been programmed into a laptop or handheld computer); (2) *telephone interviewing (or computer-assisted telephone interviewing – CATI)*, in which respondents participate in a survey either via a fixed-line telephone or their mobile phone; and (3) *self-administered questionnaires (SAQs)*, in which respondents are invited to complete either a paper questionnaire, often as part of a postal survey; or a questionnaire hosted on the Internet (sometimes referred to as web-based computer-assisted self-interviewing or web-CASI). Advances in information and communication technology have expanded this range of options, such that most of the principal modes now offer a number of variant forms, including *computer-assisted self-interviewing* (CASI) in the context of a face-to-face interview, *audio-CASI* (or A-CASI), in which the questions are pre-recorded and played back to respondents who enter their data into the computer; *touchtone data entry* (TDE) - a form of telephone interviewing, in which respondents enter their answers using the keys on their handset, and *web-cam interviewing* using Voice over Internet Protocol (VOIP). Choosing one mode over another typically involves an assessment of the strengths and weaknesses of each with respect to a range of different factors.

Data collection modes vary along a number of dimensions making them more or less suitable to the needs of particular surveys. Firstly, modes vary in the extent to which they provide access to different survey populations. For example, for surveys of nationally-representative samples, face-to-face interviews usually offer the most dependable method of accessing sample members (depending on the availability of comprehensive lists of addresses), because in theory, interviewers can visit any household in the country. This is true for the UK, but it varies from one country to another, depending on its size and geography. In the same way, postal surveys can also provide access to the majority of households, but self-administered modes will not always be suitable for everybody because they depend on a minimum level of literacy. Telephone surveys of nationally representative samples can also be carried out effectively in the UK using Random Digit Dialling (RDD), because most households have telephones, but the effectiveness of RDD for this purpose is being reduced by the rising number of people abandoning their fixed-line telephones in favour of mobile phones only (resulting in under-coverage in the achieved sample). At present, the use of web surveys of random samples tends to be restricted to special populations, such as students in schools and universities and employees of particular sorts of organisation known to have Internet access. However, nationally representative Internet panels based on random samples have been established (e.g. in The Netherlands and the United States), and this is likely to become an increasingly popular method of collecting survey data in other countries in the future.

Secondly, the choice of mode is guided by the extent to which each has differential administrative and resource burdens (Czaja and Blair, 2005). These include the financial costs of implementing the survey in each mode and the time taken to complete fieldwork. The four principal modes can be ranked in terms of their relative

costs, starting with face-to-face interviewing as the most expensive option (mainly due to the interviewers' travel expenses). Telephone interviews generally offer a more economical solution (especially as call costs have decreased over time). Groves and Kahn (1979), for example, estimated the cost of the telephone survey in their comparison study to be less than half that of the face-to-face survey. A similar estimate of the cost differential between the two interviewer-administered modes is provided by Czaja and Blair (2005). Postal survey methods offer one of the lowest cost options, but the relatively low administrative costs are often offset by the comparatively long fieldwork periods required to ensure questionnaires are completed and returned – particularly where multiple reminders are sent out, as advocated by Dillman (1978). Meanwhile, the development of the Internet as a tool for data collection has revolutionised the speed with which survey fieldwork can be carried out, though there may be significant costs associated with programming, as well as with software development and server support. Nevertheless, once the initial set up costs have been met, there is effectively no limit on sample size, as there are with other data collection modes (though of course there may be considerable limits in terms of representativeness).

Thirdly, the four principal modes vary in the extent to which they are suited to the administration of questionnaires of different types (Czaja and Blair, 2005). Questionnaires differ in length and complexity, for instance in terms of the routing and skip patterns that guide respondents to applicable questions. Longer, more complicated instruments are less well-suited to self-completion modes because of the opportunity for such routing errors. Computer-assisted interviewing has enabled survey questionnaires and questions to become considerably more complex than they were in the past (Smith, 1987). Likewise, web-based questionnaires can be programmed with sophisticated interactive designs to ensure only those questions that are applicable to the respondent (based on responses to previous questions) are presented. Long questionnaires have generally been avoided in telephone interview surveys, with some survey organisations restricting interview length in order to minimise the burden on respondents and to prevent break-offs mid-interview. However, the empirical evidence to support the negative impact of interview length on cooperation in telephone surveys is mixed (Bogen, 1996; Collins et al., 1988). There is considerably more evidence to suggest that lengthy self-administered questionnaires can be off-putting to would-be respondents (Dillman, 1978; 2000) and the ease of abandoning a web survey makes questionnaire length a particularly important variable to control in Internet mode (Heerwegh and Loosveldt, forthcoming). These considerations have meant that face-to-face interviews have traditionally been relied upon for administering long and complex questionnaires, although the evidence suggests that respondents may be less concerned about the length of a questionnaire, if the survey topic is of particular interest to them (Groves, 2004).

2 Opting to mix modes

The researcher's aim when designing a survey is to select the best possible data collection strategy to suit the particular requirements of the research question (including the survey topic and the population of interest), while maximising the quality of the data collected (de Leeuw, 2005). In reality, however, the process of survey design involves a series of compromises between the researcher's ideal and the costs of implementing it, so decisions about modes tend to be based more on

administrative and resource factors (Czaja and Blair, 2005). For example, researchers are restricted in their choices by the available budget, the time period in which the data are required and the existing infrastructure available for conducting surveys in different modes.

Increasingly, decisions about survey modes are also having to be based on the so-called 'survey-taking climate' (Lyberg and Dean, 1992) – i.e. the social context in which the research is being undertaken (de Leeuw, 2005). It is widely accepted that survey response rates generally have been declining almost universally (e.g. de Leeuw and de Heer, 2002), a trend that has been attributed to two main causes. Firstly, non-contact rates have been rising, partly because of the growing number of barriers between researchers and sample members encountered using traditional modes (Groves and Couper, 1998). For example, more and more people in urban areas are living in so-called 'gated communities' or using entry-phone systems, making it difficult for interviewers to gain access to households in their sample. The use of telephone answering machines, voice mail, caller ID systems and 'Do Not Call' registers has similarly introduced impediments to contact in telephone surveys. Secondly, refusal rates also appear to have risen due to a growing reluctance on behalf of the public to take part in surveys, a trend attributed to 'survey fatigue' – either a real or imagined increase in the number of requests to participate in surveys. At the same time, the cost of conducting surveys has gone up, partly through attempts to mitigate the potentially detrimental effect of under-coverage and nonresponse on survey quality; the number of person hours required to achieve a successful interview has risen (with interviewers having to make repeated visits/calls to households before making contact), and increasingly generous incentives are becoming necessary to motivate sample members to participate. While the impact of these factors has affected the different modes to varying degrees, the fact remains that traditional single mode data collection strategies appear in many ways to no longer be fit for purpose (Dillman, 2000). This combination of factors has led to the growing attraction of mixing modes of data collection (de Leeuw, 2005).

When considering mode choices and the motivations for mixing modes, it is helpful to distinguish between different phases of the survey process, which are often characterised by different forms of interaction with sample members (de Leeuw, 2005; Balden 2004, cited in de Leeuw, 2005). Each phase may entail a different means of communication, resulting in multiple modes being combined in order to implement the survey. For example, at the 'pre-contact phase', respondents may be contacted with an advanced letter. Contact with sample members to arrange an appointment for interview may be conducted by telephone, while the interview itself is carried out in person. It is also not uncommon for face-to-face interviews to incorporate a self-completion component or for a supplementary self-administered questionnaire to be left with the respondent to be collected later. Some time after the interview is complete it may be followed-up by a short telephone interview for the purposes of quality control, and so on. De Leeuw refers to these different ways of combining modes – for both communication and data collection purposes – as 'mixed-mode survey systems' (2005; p.237). Mixing modes of data collection has implications for the amount of error affecting the quality of the survey and this varies depending on the motivations for mixing modes and the type of mixed mode system adopted. The next section considers the different sources of survey error associated with mode.

3 Mode effects

Setting aside the practical considerations that govern decisions about the suitability of different modes for particular survey needs, data collection methodologies can also be evaluated according to a range of factors concerned with the quality of the data each produces. Data quality can be affected by a range of different sources of bias in surveys, which together are referred to as the 'total survey error' (Groves, 1989; Weisberg, 2005). Total survey error consists of both random and systematic errors, including sampling errors (which arise because estimates are based on a sample rather than a full census of the population) and a range of other types of non-sampling errors: including coverage error, non-response error, and measurement error. The choice of data collection mode influences the extent to which the data are affected by each type of non-sampling error. These different types of survey error constitute the three main types of 'mode effect' researchers interested in mixing modes should be aware of.

3.1 Coverage error

Coverage error (also know as sampling frame bias) occurs when not all members of a target population have an equal chance of being selected in the survey sample. This type of error tends to be lower for surveys that are able to make use of more comprehensive lists for sampling purposes. For all data collection modes, the availability of comprehensive sampling frames depends on the population of interest and on the intended mode of contact. For single mode national population surveys, coverage error is generally lower for face-to-face and postal modes that draw samples from lists of addresses, electoral rolls or population registers (although considerable variation exists between countries in terms of the availability of suitable lists). The level of coverage error in RDD surveys depends upon the proportion of households in the population with fixed telephone lines, so there are national differences due to variation in the extent of fixed-line telephone penetration across countries, as well as due to variation in the proportion of 'mobile-only' households. Similarly, absence of coverage error in random samples for surveys conducted via the Internet depends on each member of the population having *access* to the Internet. For this reason, under-coverage of the population represents the main barrier to switching to web surveys of nationally representative samples for population surveys.

3.2 Non-response error

Nonresponse error refers to bias in the sample responding to the survey, which results from differential response rates across different subgroups of the population. Response rates vary with mode of contact and mode of data collection. In this respect, face-to-face surveys have long been viewed as the 'gold standard' for their effectiveness at securing high levels of participation. This has been attributed partly to sample members' perceptions of the survey's legitimacy, which tends to be higher than for other modes and partly to the fact that interviewers on the doorstep appear to be more effective than either telephone calls, advance or accompanying letters or email invitations at persuading would-be respondents to take part. Response rates tend to be considerably lower across each of the other modes, increasing the likelihood of response bias (the tendency for certain subgroups to be more likely to respond than others). Thus, data collection modes not only vary in their effectiveness

at achieving high levels of participation but also because they are more or less likely to attract different members of the population to participate. Face-to-face interviewing appears to be one of the least problematic modes in this respect, with a generally equal cooperation rate across groups (Czaja and Blair, 2005). Telephone surveys, on the other hand, not only achieve considerably lower response rates, but also higher levels of response bias. According to Holbrook et al.'s (2003) review of seven studies comparing the demographic composition of responding samples to face-to-face surveys[1] with those participating in RDD telephone surveys found considerable agreement on a range of observed demographic differences. For the most part, the telephone samples had fewer respondents with low education, fewer respondents on low incomes, fewer older respondents, and fewer minority respondents than the face-to-face samples (Holbrook et al., 2003; p.94-95). Postal and web-based surveys, on the other hand, tend to favour better educated, more literate (including computer literate) members of the population. The likelihood of survey participation is also intrinsically linked to the survey topic and the level of interest in the topic among members of the target population (Groves, Singer and Corning, 2000; Groves, Presser and Dipko, 2004). This means that in the absence of extrinsic motivators for participating in a survey (e.g. a persuasive interviewer), self-administered modes tend also to favour those with a particular interest in the survey topic.

3.3 Measurement error

Measurement error refers to bias in the actual responses recorded in the survey. It can be attributed to two main sources: the 'questions' and the 'actors' involved in the survey process (Groves, 1979) – that is, they typically result from either the design of the questionnaire and the particular questions being asked, from the respondents themselves, or – in the case of face-to-face and telephone interviewing – from the interviewer. Mode of data collection can affect the quality of the measurements by determining the way in which questions are asked or presented and by influencing the cognition and behaviour of the respondent and interviewer. For example, one of the best-documented types of mode-related measurement error is the tendency for respondents to modify the true answer to certain types of survey question in order to present themselves in a more favourable light (Dillman, 2000) – an effect that is more common in interviewer-assisted surveys than in self-administered surveys (see section 8). Other types of mode-related measurement errors result from the medium in which the question stimulus is presented to respondents – whether visually (as in self-administered questionnaires), aurally (as in telephone interviews) or both (as in face-to-face interviews). The main types of mode-related measurement errors are described in further detail later in the paper.

Mode effects of all types are problematic because they increase the total amount of non-sampling errors affecting the overall quality of the survey. They are especially problematic in surveys that use a combination of modes, however, because of their effect on the comparability of the data. The next section considers different types of mixed mode survey design and their implications for survey error.

[1] (of national area probability samples of the US population)

4 Types of mixed mode design

One of the main motivations for using a combination of modes – whether during different phases of the survey, or specifically for the purposes of data collection – is that it offers the possibility of off-setting the weaknesses of one approach with the strengths of another. The fact that modes vary with respect to factors such as the cost and speed of fieldwork, their suitability for administering different types of questionnaire and their impact on data quality means that, in principle, using a mix of modes allows the researcher to minimise both the costs and errors associated with using any given single-mode approach. Mixed-mode designs are, therefore, becoming increasingly popular as a means to tackling the problem or threat of coverage, nonresponse and measurement errors, as well as a way of reducing survey costs. For example, an additional mode can help provide access to a group of respondents that would otherwise be hard or impossible to contact in the principal survey mode. Similarly, using an alternative mode to administer certain types of questions may help to reduce the likelihood of certain forms of measurement error. Furthermore, combining modes 'sequentially', whereby one starts data collection with the most economical mode, and follows up non-respondents with increasingly 'expensive' modes appears to offer advantages with respect to both minimising costs and increasing participation (e.g. Hochstim, 1967; Mooney, Giesbrecht and Shettle, 1993; Voogt and Saris, 2005). In each case, the rationale for implementing each type of mixed mode system (de Leeuw, 2005), is to reduce a particular form of survey error.

In longitudinal, panel or cross-national surveys –there may be additional motivations for mixing modes. For example, different modes may be better suited to the various stages of a longitudinal survey, depending on the type of data to be collected at different time points. Similarly, panel surveys may also benefit from the use of a range of modes for different purposes. For surveys conducted in more than one country there are further considerations. Because multi-nation surveys depend for their reliability on a kind of 'principle of equivalence' (Jowell, 1998), which applies to all aspects of the survey process, the most ambitious multinational projects tend to require all participating countries to employ the same mode of data collection. Yet national variations in survey practice – including the experience of using different modes, the available infrastructure for conducting surveys in different modes, and the level of coverage offered by different modes cross-nationally – means that insisting on the same methods in all countries may not always be the best way of ensuring that *equivalent* methods are used. Variation in survey response rates across countries, which in turn have implications for the validity of cross-national comparisons, and differences in fieldwork costs lend further weight to the arguments in favour of adopting multimode data collection designs in comparative surveys – whether using some of the designs described, or by simply allowing participating countries to adopt their own preferred method of data collection.

Dillman (2000; pp219-222) identifies five different situations in which researchers might be motivated to mix modes of data collection and discusses the error implications of each. They include using different modes to (1) collect the same data from different members of a sample; (2) to collect data at different stages of a panel survey; (3) to administer different parts of the questionnaire within a single data collection period; (4) to collect data from different populations (as in a cross-national survey); and (5) for communicating with respondents at different phases of a survey process (i.e. making contact in one mode to request participation in another). Of these, neither (3) nor (5) appear to present any particular problems in terms of their

error consequences. However, while the remaining mixed mode designs may help to reduce survey costs *and* coverage, nonresponse and specific forms of measurement error, they entail the risk of differences in the measures obtained across modes. De Leeuw (2005) expands on this analysis by identifying further effects on survey quality when modes are mixed at the response phase of a survey, including confounds between effects on measurement and response biases in the case of design (1) and time effects in the case of design (2) and the *introduction* of coverage and nonresponse errors in the case of design (3), resulting in incomparability both across modes and across samples (de Leeuw, 2005; p.238). In each situation, the researcher must decide whether it is better to try to reduce one form of error at the risk of increasing a different form – notably, different types of measurement error (de Leeuw, 2005). In the next section, I review what is currently known about mode effects on measurement error, their causes and their implications for the decision over whether or not to mix modes.

5 Characteristics of modes that influence data quality

While the use of a mix of modes may offer solutions to problems of coverage and nonresponse – and even help to reduce fieldwork costs – when modes are combined, their unique measurement properties may mean that differences observed in the data can be attributed to how the data have been collected rather than to real differences in the survey population. In particular, the fact that mode affects how respondents answer questions presents a significant barrier to adopting multi-mode designs. When considering the option of mixing modes in surveys, therefore, it is helpful to have an understanding of the causes of different types of mode effect on measurement error. Based on a number of contributions to the field (notably by Groves, 1979; Groves and Kahn, 1979; Schwarz, Strack, Hippler and Bishop, 1991; de Leeuw, 1992; 2005; Tourangeau, Rips and Rasinski, 2000; Dillman, 2000; and Holbrook, Green and Krosnick, 2003), it is possible to draw up a list of the characteristics of modes of data collection that help to explain differences in respondents' answers across modes. This list is presented in table 1 in the appendix, with a brief explanation of the significance of each characteristic.

A number of attempts have been made to classify these different characteristics. For example, de Leeuw (1992; 2005) distinguishes between three classes of factors (see figure 1 in the appendix): media-related factors, which are "concerned with the social conventions and customs associated with the media utilized in survey methods" (de Leeuw, 2005; 205); factors influencing information transmission (such as whether the mode uses mainly visual or auditory communication); and interviewer effects. By contrast, Tourangeau, Rips and Rasinski (2000; see also Tourangeau and Smith, 1996; 1998) focus on the psychological effects that the different characteristics of data collection methods have on respondents during the response process (see figure 2). They identify three features of modes in particular (whether or not the questionnaire is administered by an interviewer or by the respondent himself, whether or not the data collection process is computer-assisted and whether or not the question and response stimuli are presented aurally or visually) as being important because of how they relate to psychological variables assumed to mediate the effect of mode on response quality. These psychological variables include "the sense of impersonality [or privacy] the method fosters, the cognitive burdens it imposes, and the legitimacy it seems to confer" (2000; 305).

Crucially, the model shown in figure 2 highlights the fact that certain psychological variables which co-vary with the mode of data collection have the capacity to influence how respondents decide their answers to survey questions (Schwarz et al., 1991). Holbrook, Green and Krosnick's (2003; 81) study of differences between telephone and face-to-face interviewing emphasised two main types of influence in particular: (1) the extent to which the mode makes respondents feel comfortable to answer openly and honestly to questions that may be of a sensitive or personal nature and (2) the influence of the mode on the respondent's likelihood of exerting the required cognitive effort to answer the survey questions carefully (see also Jäckle, Roberts and Lynn, 2006; Roberts, Jäckle and Lynn, 2007). This more parsimonious explanation of how differences between modes lead to differences in response has been used to classify the mode characteristics shown in table 1 and forms the focus of the next section.

6 How respondents answer survey questions

In order to understand how differences between modes of data collection influence people's answers to survey questions, it is helpful to consider the cognitive processes involved in survey response and the types of response errors that can occur. A number of contributors have developed models of survey response (e.g. Cannell, Miller and Oksenberg, 1981; Tourangeau, 1984; Tourangeau, 1987), distinguishing between the different stages of cognitive processing respondents go through when presented with a survey question. Most recently, these different approaches have been consolidated into a four-component model that forms the basis of Tourangeau, Rips and Rasinski's (2000) book 'The Psychology of Survey Response'. In summary, the four components include "comprehension of the item, retrieval of relevant information, use of that information to make required judgements and selection and reporting of an answer" (Tourangeau et al., 2000; p.7). Each component encompasses a range of different cognitive processes. For example, in order to comprehend what a survey question is asking, the respondent must interpret the grammatical structure or 'syntax' of the question, the semantics of the question (i.e. its meaning) and its 'pragmatics' or its intended use (p.25). In order to retrieve the relevant information needed to answer the question, the respondent must search the relevant memory system in which the information has previously been encoded and activate information held in long-term memory so that it can be used in other cognitive processes necessary for formulating a response (p.77). Both the judgement and response selection components of survey response similarly comprise multiple processes.

The nature of the processes required to arrive at a survey response will depend on the precise nature of the question being asked and on how it is asked. For example, factual questions (e.g. asking about respondent's sex, age, occupation) typically elicit a fairly automatic response requiring minimal processing, whereas more complex questions, particularly those requiring computations (e.g. number of visits in the past 12 months to a General Practitioner; number of hours spent watching television each week) involve more extensive processing in order to retrieve the relevant information from long-term memory and carry out the necessary calculation to arrive at the answer. Equally, the cognitive processes involved will depend on the mode in which the question is asked. Self-completion questions require the respondent to read the question first in order to comprehend what is being asked, while interviewer-administered surveys require respondents to listen and attend to the question being

asked, which again involves different processes depending on whether the interview is conducted face to face (allowing the use of visual cues, including response aids like showcards) or by telephone (which is restricted to audio-only communication).

The model is based on the assumption that respondents are sufficiently motivated and able to engage in the cognitive work required to answer survey questions carefully and honestly. Yet even under optimal conditions, difficulties can arise at each stage of the process and in all modes, resulting in errors in the data. Sometimes these errors can be attributed to problems with the questionnaire, such as questions that are worded ambiguously or are unnecessarily complex. At other times, the answer respondents formulate in response to a question does not easily map onto the response categories provided. Research into difficulties of this kind has resulted in an extensive literature providing guidelines on how best to design survey questions (e.g. Krosnick and Fabrigar, forthcoming; Oppenheim, 1992; Sudman and Bradburn, 1982). In addition, errors can be attributed to respondents, who are either unable or unwilling to exert the requisite effort to formulate an optimal response to the survey question (Krosnick, 1991) or feel reluctant to answer certain types of survey question openly and honestly. Errors of the first sort are known as 'satisficing' effects (Krosnick, 1991), while errors that arise because respondents are either unwilling to disclose personal information about themselves or wish to present themselves in a favourable light are referred to as social desirability biases (e.g. see DeMaio, 1984 for a review). The following sections deal with these two different sources of response error and describe how each is influenced by characteristics of the data collection mode.

7 Respondent satisficing

Completing each of the four stages of cognitive processing systematically might be termed the optimal response strategy (Krosnick, 1991). According to Krosnick's theory of survey satisficing, however, not all respondents are sufficiently motivated or able to carefully execute each of the four components of processing well enough to provide optimal responses throughout the questionnaire. Rather, under certain conditions, respondents will either complete each of the stages less thoroughly or skip one or more of the stages of processing altogether. Carrying out the cognitive processing necessary for responding optimally but doing so less thoroughly is referred to as a 'weak' form of satisficing. The respondent chooses a "merely satisfactory" response rather than an optimal one (Krosnick, 1991; 215). By comparison, missing out or shortcutting some of stages of processing altogether is referred to as 'strong satisficing', in which respondents select answers almost arbitrarily, looking for cues in the question to help them choose a reasonable-sounding answer to avoid having to think carefully about their response. The degree of effort expended by the respondent in formulating responses to survey questions can be described as varying along a continuum, anchored at the extreme points by optimising at one end, where respondents are able and willing to engage in the necessary cognitive processing to arrive at an optimal response and by strong satisficing at the other end, where respondents miss out crucial stages in the response process, selecting answers instead seemingly at random (Krosnick, 1991; 1999).

Krosnick (1991) classifies a range of frequently-observed response strategies according to whether they constitute the weaker form of satisficing or the strong form. Weak forms of satisficing include acquiescence (e.g. Knowles and Condon, 1999), in which respondents agree with assertions made by the interviewer (because this

requires less cognitive effort than generating reasons to disagree with a statement) and response order effects (Schuman and Presser, 1981; Krosnick and Alwin, 1987), in which respondents select the 'most accessible' response option from the list provided, to give an acceptable answer to the question being asked. Examples of strong satisficing identified by Krosnick include strategies such as selecting the no-opinion or 'Don't Know' response option; 'non-differentiation', in which objects to be rated on the same scale are rated on the same scale point; 'endorsing the status quo', in which respondents simply agree that 'keeping things the same' is better than changing policy, to avoid making the requisite effort to assess the impact of a possible change in policy; or simply picking any of the response options at random – also referred to in the literature as 'mental coin flipping' (Krosnick, Narayan and Smith, 1996). Other response strategies have also been investigated as possible indicators of satisficing, including selecting the middle response category and extremeness, a preference for selecting answers from the end points of a scale (Holbrook, Cho and Johnson, 2006).

Krosnick (1991) identifies three factors that contribute to the likelihood of adopting a satisficing response strategy: task difficulty, respondent ability and respondent motivation. Respondent ability refers to the respondent's capabilities with respect to executing the cognitive processes required to formulate responses to survey questions and is influenced by factors such as individual differences in cognitive skills (including intelligence), level of education, and knowledge of the survey topic. Also relevant is the respondent's own level of interest in – or 'involvement' in the survey topic (i.e. how important it is to them). Motivation can also be influenced by individual characteristics of the respondent such as 'need for cognition' (Krosnick, 1991; Cacioppo and Petty, 1982) – the fact that some respondents simply enjoy engaging in cognitive tasks more than others – and situational factors such as how far into the interview the respondent is (motivation typically wanes as the respondent progresses through the questionnaire). Task difficulty refers to the cognitive burden placed on the respondent during survey participation. This will vary by survey as a function of the topic and the nature of the questions being asked, as well as over the course of an individual respondent's participation in a survey; certain questions may also be more complex or difficult to answer than others (e.g. because they are harder to interpret or make particular demands during the retrieval, judgement or response selection stages of processing).

Crucially, as can be seen from table 1, both respondent motivation and task difficulty appear to vary as a function of the mode of data collection. For example, interviewer-administered surveys offer considerable advantages over self-administered surveys, because interviewers are able to engage respondents in the survey task and to facilitate the response process, for example, by answering respondents' queries or detecting comprehension problems and responding as appropriate (Holbrook et al., 2003). Face-to-face interviewers can make use of a range of visual cues including nonverbal means of communication to encourage the respondent and visual aids (e.g. showcards) to help alleviate the cognitive burden on the respondent's working memory. Nonverbal communication also helps to slow down the pace of the interview, reducing task difficulty associated with answering survey questions under time pressure.

Based on these considerations, research investigating mode effects on satisficing predicts greater satisficing in telephone interviews compared with interviews conducted face-to-face (e.g. Holbrook, Green and Krosnick, 2003; Jordan, Marcus and Reeder, 1980; Krosnick, Narayan and Smith, 1996; Jäckle, Roberts and Lynn,

2006). However, the predictions are less clear cut with respect to self-completion questionnaires. On the one hand, self-completion surveys eliminate time pressure altogether, as respondents are able to control how and when they complete the questionnaire. On this basis, satisficing may be less likely in SAQs than in survey interviews (e.g. Fricker et al., 2005). On the other hand, for the reasons already described, SAQs may be associated with a greater likelihood of satisficing because overall the cognitive burden on respondents is greater and motivation to participate is typically lower.

In fact, the evidence suggests that variation in satisficing effects by mode depends on additional factors such as the length of the questionnaire and the types of questions asked. For example, with respect to the former, Holbrook and her colleagues (2003) found evidence of more satisficing effects in telephone interviews (including selecting 'Don't Know' response options, non-differentiation and acquiescence) where the questionnaire was long (over an hour), whereas Jäckle, Roberts and Lynn (2006) found no difference between the two modes in the level of satisficing where the interviews were relatively short (around 15 minutes). Fricker and his colleagues (2005) found more don't know responses to items testing scientific knowledge among telephone respondents than among web respondents, which they attributed to the greater time pressure on response in the telephone mode. Web respondents on the other hand gave less differentiated responses than did the telephone respondents to a series of attitude measures all using the same scale; an effect that may have been mediated by the visual layout of the items on the computer screen. However, with the exception of response order effects – identified as a weak form of satisficing by Krosnick (1991) – relatively few mode comparison studies have explicitly examined differences in the extent of respondent satisficing (Holbrook et al., 2003).

Response order effects are defined by Krosnick and Alwin (1987; p. 202) as "changes in answers to closed-ended survey questions produced by varying the order in which response options are presented." They take two forms, each characterised by respondents showing a preference for items either at the start or the end of a list of options. The form the response order effect takes is determined by the mode in which response categories are presented to respondents (Schwarz et al., 1991; Krosnick, 1999). *Primacy* effects – the selection of items at or near the start of the list – mainly occur when response options are presented visually, as in self-completion modes and face-to-face interviews using showcards. By contrast, *recency* effects – the selection of items at or near the end of a list – occur when the response options are read out to respondents, as in telephone interviews, or face-to-face interviews without showcards. In each case, the effects are attributed to satisficing, resulting from the cognitive burden associated with processing multiple items in a list (Krosnick and Alwin, 1987; Krosnick, 1999). Consistent with this theory are findings confirming increased likelihood of response order effects under conditions which foster satisficing – e.g. where the list of items is long, where questions are difficult to comprehend; among respondents with lower levels of education and so on (e.g. Holbrook et al., in press).

As with other forms of satisficing, the extent to which the data collection mode affects the likelihood of response order effects depends in part on the type of question. For example, questions with categorical response options appear to be particularly susceptible to the effect as the psychological processing of unordered lists operates somewhat differently to the processing of ordered rating scales (Krosnick, 1999). However, despite the large number of studies that have investigated primacy and recency effects, the number of strictly comparable studies of mode differences

appears to be somewhat limited, making it difficult to draw conclusions about whether this type of response bias is more prevalent in one mode over another. Part of the problem is that the kind of questions vulnerable to primacy effects – e.g. questions with long lists of unordered categories – are rarely used in telephone surveys precisely because of how demanding they are. As a form of satisficing, however, it seems reasonable to conclude that mode characteristics that increase task difficulty and reduce respondent motivation will make surveys particularly susceptible to this form of measurement error. As Schwarz and his colleagues (1991) argue, the effect of mode on the direction of response order effects is undisputed, and the combined effect of primacy and recency in surveys using a mix of visual and aural modes is likely to have a substantial effect on the quality of the estimates.

8 Social desirability bias

In the same way that respondents will not always be motivated or able to exert the necessary cognitive effort to provide optimal responses to survey questions, they will not always feel able to report their true answers to questions openly and honestly. In these situations, respondents are assumed to adapt their true answers to survey questions either in order to portray themselves in a more favourable light or to give answers they think the research team will want to hear (Sudman and Bradburn, 1974; Bradburn et al., 1978; Dillman, 2000). As noted, this tendency is referred to as social desirability bias. It long been recognised in the psychometric literature as a problem affecting the validity of personality and various clinical inventories (Crowne and Marlowe, 1960; Barger, 2002), and has since been documented in a large number of studies by survey methodologists (e.g. see DeMaio, 1984). Questions perceived to be intrusive or which ask about 'threatening topics' (Bradburn, Sudman and Wansink, 2004) appear to be especially likely to elicit socially desirable reporting (Fowler, 1995).

Evidence of the occurrence of social desirability bias comes from a range of sources, including studies that have validated the accuracy of survey estimates by consulting external records (Tourangeau et al., 2000). For example, studies of this kind have recorded over-reports of church attendance in the United States (e.g. Hadaway, Marler and Chaves, 1993) over-reports of voting behaviour (e.g. Traugott and Katosh, 1979; Silver, Anderson and Abramson, 1986; Karp and Brockington, 2005); and under-reports of declaring bankruptcy (Locander, Sudman and Bradburn, 1976). Similarly, Tourangeau and Smith (1998) have compared the number of opposite sex partners reported by men and women in surveys and shown that men tend to over-estimate the number of partners they have had, while women tend to under-estimate the number of partners (if respondents are reporting truthfully, the total number of opposite sex partners reported by both sexes should be equal). Studies of how people respond to sensitive questions in different reporting situations have similarly informed our understanding of social desirability bias. Notably, the more respondents feel assured of the anonymity of the reporting situation, the more honest their responses tend to be (e.g. Paulhus, 1984). This has been demonstrated in so-called 'randomised response' studies in which interviewers are unaware to which of the randomly-assigned questions respondents are providing an answer (Warner, 1965). Similarly, where respondents believe that the researcher can discover their true response via other means – such as in 'bogus pipeline' studies (e.g. Tourangeau, Smith & Rasinski, 1997; see Roese and Jamieson, 1993 for a review) – findings suggest respondents are more likely to report their answers truthfully (Holbrook et al., 2003).

Mode comparison studies comparing responses to sensitive questions across different modes of data collection have also been particularly informative in relation to social desirability bias. One of the most consistent findings of studies of this kind is that socially-desirable responding is significantly more likely with interviewer-administered modes of data collection, compared with self-administered modes (Sudman and Bradburn, 1974; Dillman, 2000; Tourangeau and Smith, 1998). Tourangeau et al. (2000) reviewed seven studies comparing self-reports of drug use in surveys conducted in different modes. For each estimate obtained in the studies, they calculated the ratio of drug use reported in self-administered surveys to the corresponding estimates obtained in interviewer-administered surveys. Across 63 different comparisons, they found that 57 showed higher levels of reporting of drug use among respondents participating in self-administered modes (Tourangeau et al., 2000; pp.295-96). Jobe et al.'s (1997) study of sexual behaviour similarly showed that respondents to self-administered questionnaires not only reported more sex partners, but also more sexually transmitted diseases and greater use of condoms than respondents interviewed face-to-face.

The occurrence of social desirability bias in interviewer surveys is not restricted to self-reports of 'sensitive behaviours' either. Krysan et al. (1994) found respondents to a postal survey were more likely to express negative attitudes on 'racial integration and affirmative action' compared with respondents in face-to-face interviews. Similarly, people were found to be more likely to report their health favourably in telephone surveys than they were in mail surveys (Fowler, Roman and Xiao Di, 1998). These accumulated evidence led Tourangeau and Smith (1998) to conclude that "self-administration may not eliminate misreporting due to social desirability, but it may be the single most effective means for minimising this source of error" (p.449), and in keeping with this, it is now common practice to use self-administered questionnaires to administer sensitive questions either for entire surveys of topics likely to elicit socially desirable responses, or as a self-completion component within a face-to-face survey.

The robustness of these mode effects suggests that the principal cause of social desirability bias is the level of privacy or perceived anonymity of the reporting situation (Tourangeau et al., 2000). As Weisberg (2005) states,

> *"The lower the cost to the respondent for disclosing something of a private nature, the less social desirability effects should be present; survey modes that provide high anonymity should therefore lead to greater reports of sensitive behaviour"* (pp. 290-291).

Consistent with this viewpoint are findings from mode studies that have examined the effect of computerised methods of questionnaire administration, which either eliminate the need for interviewers altogether or enhance the respondent's sense of distance from the researcher (e.g. Lucas et al., 1977; Erdman, Klein and Greist, 1983; Locke et al., 1992; Tourangeau and Smith, 1998; Wright, Aquilino and Supple, 2000; Hewitt, 2002). These studies provide some evidence to suggest that computerised self-administered modes may be effective at reducing socially desirable reporting, although not necessarily more so than self-administration per se (see Jobe et al., 1997), and perhaps not equally so across all groups of respondents (Wright, Aquilino and Supple, 1998; Hewitt, 2002).

When the presence of social desirability bias in face-to-face surveys is compared with that found in telephone surveys, the results are less clear-cut. Early contributors (e.g.

Hochstim, 1967; Sudman and Bradburn, 1974; Aquilino, 1994) hypothesised that the greater social and physical distance between the interviewer and respondent in telephone interviews compared with those conducted in-person, would make social desirability bias more likely to compromise data quality in the latter, compared with the former, and the findings of some studies support this. However, more recently a consensus in the literature is emerging, suggesting that actually telephone interview data are more affected by social desirability bias than data obtained from in-person interviews. For example, Smith's (1984) review of 12 studies found no clear evidence to support the idea that telephone interviews decrease socially desirable reporting; De Leeuw and van der Zouwen's (1988) meta-analysis of 31 studies found a slight tendency for *more* socially desirable reporting in telephone interviews; and more recently, Holbrook et al. (2003) found consistent evidence of greater social desirability bias (for items with empirically-established social desirability connotations) across three national surveys involving face-to-face and telephone comparisons. In our own research on the European Social Survey, we also found socially desirable responses were more likely in telephone mode (see Jäckle, Roberts and Lynn, 2006).

These findings suggest that factors other than privacy, anonymity and confidentiality may influence the likelihood of socially desirable reporting. Notably, in telephone interviews, there are fewer opportunities for interviewers to convey to respondents the legitimacy of the survey request (Groves, 1989), a factor that may affect respondents' willingness to answer truthfully when asked about sensitive topics. Groves' (1989) suggestion that different cognitive scripts govern telephone interactions compared with face-to-face ones is also consistent with this idea – the 'solicitation' script invoked by a request to participate in a telephone survey might explain respondents' unwillingness to divulge personal information to strangers on the telephone, as compared with the 'visit from a guest' script that is likely to enhance trust between face-to-face respondents and interviewers. Related to this is the suggestion by Holbrook and her colleagues (2003) that face-to-face interaction facilitates the build-up of rapport between interviewer and respondent, leaving the latter less likely to be influenced by social desirability concerns.

In fact, these alternative accounts reflect wider conceptual ambiguity surrounding the problem of social desirability in surveys (DeMaio, 1984). DeMaio's review of the literature revealed disagreement surrounding a number of key issues, including (1) how to conceptualise social desirability – as a characteristic of particular survey questions or a personality trait (so-called 'need for social approval'); and (2) what determines what is socially desirable (social norms governing the acceptability of certain types of behaviour or shared values among different groups about what is desirable). There is also disagreement about the cognitive mechanisms underlying the bias and the extent to which they are under the conscious control of the respondent (e.g. Paulus, 1984; Holtgraves, 2004). The commonly-accepted view is that respondents are motivated to execute the response process systematically, but that they edit their true response to the survey question in order to avoid embarrassment (Tourangeau and Smith, 1998; Tourangeau, Rips and Rasinski, 2000). In fact, social desirability concerns may trigger other mechanisms, including self-deception, biased retrieval or even shortcutting – a suggestion consistent with the finding that social desirability is more common in telephone interviews, which are typically conducted at a faster pace and therefore limit opportunities for response editing (Holtgraves, 2004). An alternative explanation is that effects that look like social desirability bias – e.g. over-reports of voter turnout – may in fact result from a combination of recall errors

and nonresponse bias and that biased reporting – whether conscious or unconscious – may not be as prevalent as is assumed (Kronick, 1999; p.546). Resolving these outstanding ambiguities represents an important challenge for future research into social desirability and for developing methods of minimising its negative effects on the quality of data collected using different survey modes.

9 Reducing the likelihood of mode effects

The costs and benefits of mixed mode data collection will vary, depending on the reasons for its introduction and the type of mixed mode design adopted. More specifically, multimode surveys offer opportunities to reduce fieldwork costs and to tackle specific sources of error, but they carry with them the risk of introducing other forms of error. These can be a complex mix of measurement errors that hinder cross-mode comparisons, confounded with differential coverage and non-response errors, which make it difficult to identify the causes of observed effects in the data.

Researchers must assess the potential contribution of mixed modes to reducing total survey error and try to develop strategies for dealing with the negative implications of more complex survey designs. In this paper, I have focused on the issue of on measurement error, as this perhaps represents the most difficult problem when researchers consider using a mix of data collection modes in a single survey. For the most part, research on mode effects has tended to focus on identifying problems rather than on developing solutions. However, three possible approaches to reducing the likelihood of mode effects present themselves.

The first possibility is to try to develop data collection instruments that are insensitive to the effects of the administration mode. According to Dillman (2000; p.224), one of the biggest causes of apparent mode effects is the tendency for questions to be constructed differently for different types of questionnaire. For example, questions designed for self-administration or for face-to-face surveys using showcards often need to be adapted considerably for telephone interviews, to make them suitable for aural administration. As a result, respondents in each mode may be asked questions in quite different formats resulting in question form effects that *appear* to be mode effects because they vary by mode. To combat this, Dillman advocates what he calls the ‘unimode’ approach to questionnaire construction, whereby questions are designed to be suitable for administration in all the modes to be used in the multimode survey. He outlines nine guiding principles in this approach, aimed at designing questionnaires so that they provide a ‘common mental stimulus, regardless of survey mode’ (Dillman, 2000; pp.232-240). For example, they include reducing the number of response categories to make the questions suitable for both visual and aural administration; reversing the order of response categories for half the sample in order to minimise the impact of response order effects and using descriptive labels of scale points that will work in a similar way in both aural and visual modes.

Dillman’s design principles provide an important battery of tools for researchers seeking to reduce the likelihood of mode differences in mixed mode surveys and the unimode approach is particularly appealing to any survey designer developing a new survey from scratch. It can also provide useful guidelines for research teams considering a switch in data collection strategy on continuous or repeated surveys that generate time series data. However, in these situations there are often constraints on whether or how the existing data collection instruments may be adapted and adopting

a unimode approach to questionnaire construction may not be possible because of the need to preserve the continuity of measurements (which may be disrupted by any changes to the design of the questions). A further concern is that constructing questions so that they are suitable for all modes will result in question formats or methods of administration that are less optimal for some modes than they are for others (Weisberg, 2005). For example, Dillman advises against the use of unfolding or branching question structures, because they can be burdensome in SAQs (p.238). However, questions forms of this kind can be particularly helpful in telephone interviews as a means of breaking down otherwise long and complex items that are hard to deliver without the use of visual aids.

A second option for researchers seeking to address the problem of mode effects on measurement error in the context of a time series is to try to develop measures that are *equivalent* to those obtained by the existing mode and to develop a range of other techniques to mitigate the anticipated administration effects. Understanding the causes of mode effects represents an important first step in this direction. For example, realising that fast-paced long interviews conducted by telephone may be more likely to induce respondent satisficing than the same survey interview conducted face-to-face enables researchers to develop methods of minimising its likelihood – e.g. by instructing interviewers to talk more slowly and encouraging respondents to take their time while answering in order to reduce the overall difficulty of the task.

Rather than highlighting the importance of a single instrument suitable for all modes, this approach places more emphasis on enhancing quality across all the modes administered in the survey. In practice, this means trying to minimise the likelihood of measurement errors across all the modes being used to collect data in the survey. Optimising the design of surveys for different modes in this way has been shown to enhance data quality overall, and by minimising the measurement error associated with each mode individually, it can also help to reduce the likelihood of problematic inter-mode differences (de Leeuw, 1992). The problem, however, is that any continuous, longitudinal or repeated cross-sectional surveys seeking to incorporate new modes into their data collection designs should do so only after careful assessment of the likelihood of any mode effects and their impact on the data. This means carrying out controlled experiments in which as many differences as possible in the implementation of the survey in each of the different modes are experimentally controlled for.

A third alternative for survey designers seeking to develop equivalent measures across modes is to apply correction factors or weights to the data to correct for the differential measurement error across modes (as well as for differential non-response and selection biases). Such an approach provides a solution to both the problem of how to assess the likelihood of mode effects in the data, as well as of how to deal with them. Saris and Gallhofer (forthcoming), for example, show that the requirements for equivalence typically used in surveys using different samples (namely, comparative surveys) are too strict and that equivalence should only be assessed after first correcting for differential measurement error across modes. Saris (2007) has demonstrated the effectiveness of such an approach in the context of differential measurement error across countries in the European Social Survey, but the same approach could provide a suitable solution for mixed mode surveys longer term. There are drawbacks, however, including the added complexity for analysts of multimode data and the experimental research required for developing the correction factors to start with. However, given the complexity of errors involved in mixed

mode designs, this approach allows survey designers to be resigned to the different measurement properties of different modes of data collection and to acknowledge that the quest for survey items that work in the same way across all modes may be something of an illusory goal.

Groves (1987) identifies two approaches to tackling the problem of survey error. One involves attempts to mitigate it, while the other involves attempts to quantify the extent to which the survey estimates are affected. Understanding the causes of errors provides some insight into how best to minimise their likelihood, but one of the greatest challenges to researchers seeking to mix modes is how to assess the extent of the problem to start with.

It is not uncommon for mode comparison studies to produce evidence of mode effects – usually defined as statistically significant differences in responses to questions administered in different modes. However, it is not always easy to evaluate whether the differences between modes 'matter' in practice. Tests of significance provide only a somewhat limited indicator of the effect of mode on the conclusions drawn by analysts of multimode data (Biemer, 1988; de Leeuw and van der Zouwen, 1988). According to Biemer (1988) what is also required is an assessment of the magnitude of the mode effect, the direction of the effect (i.e. whether one mode appears to provide more 'accurate' data than the other) and of whether or not the effect can be explained by other factors (notably, other differences in the implementation of the survey that have not been controlled for). Regrettably, there appears to be little agreement in the literature as to how best to assess the severity of mode effects and many mode comparison studies have suffered from confounds in experimental designs that have rendered it either difficult or impossible to draw robust conclusions from them (see Holbrook et al., 2003 for a review). This perhaps reflects the tension between the motivation to start using mixed mode designs (and to retrospectively assess whether or not there are mode effects on data quality) and the stringent methodological requirements of the studies needed to properly evaluate the presence of mode effects (Groves and Kahn, 1979; Groves, 1989; Weisberg, 2005). Holbrook and her colleagues (2003) identify 8 criteria that mode comparison studies must meet in order to be able to draw meaningful conclusions from the findings. In practice, however, most multimode survey designs would fail to meet these criteria, making it difficult for analysts to disentangle and evaluate the extent of any mode effects.

10 Summary and conclusion

This paper presents a review of the literature relating to mixing modes of data collection in surveys. It provides an overview of the different modes available to researchers designing surveys and the different factors influencing decisions about which modes to use. It describes some of the motivations for mixing modes of data collection, and the possible advantages of adopting a mixed mode approach. In particular, mixing modes of data collection offers survey designers the possibility of compensating for some of the weaknesses associated with each mode – either to reduce particular forms of survey error or to tackle other challenges affecting survey research nowadays, such as the rise in refusal rates and the costs involved in dealing with them. The paper also describes the major disadvantages of combining modes – particularly at the data collection stage of a survey. Principally, the fact that people tend to give different answers to survey questions depending on the mode in which

they are asked presents a major challenge to researchers seeking to mix modes, because of the implications it has for data comparability.

Understanding the causes of mode effects on measurement error, however, provides researchers with information with which to develop methods of minimising the likelihood of mode effects. Mode characteristics appear to influence respondents' answers to survey questions in two main ways (Holbrook, Green and Krosnick, 2003; Jäckle, Roberts and Lynn, 2006): firstly, by influencing the likelihood that the respondent will exert the required cognitive effort to answer the survey questions carefully (Holbrook et al., 2003; 81); and secondly, by influencing the extent to which respondents feel comfortable enough to answer openly and honestly to questions that may be of a sensitive or personal nature. If mode characteristics influence the respondent's execution of the response process in these ways, then the resulting quality of responses is likely to be affected, manifesting itself either as satisficing effects or as social desirability bias (Jäckle, Roberts and Lynn, 2006). Based on this understanding, researchers can develop a range of strategies aimed at mitigating mode effects, including following careful guidelines for the development of mode insensitive questionnaires, as in Dillman's (2000) unimode construction method.

As de Leeuw (2005; p.235) argues, "In mixed-mode designs there is an explicit trade-off between costs and errors," where the aim is to minimise the total survey error (including mode effects on data quality) within the administrative and resource constraints affecting affordability. Building an understanding of the different forms of survey error – and in particular, of how each is influenced by different data collection modes – is an essential step in developing new and innovative approaches to research design to meet the challenges of the contemporary survey climate.

11 References

Aquilino, W. S. (1994). Interview mode effects in surveys of drug and alcohol use: A field experiment. *Public Opinion Quarterly, 58*(2), 210-240.

Aquilino, W. S., & Lo Sciuto, L. A. (1990). Effects of Interview Mode on Self-Reported Drug Use. *Public Opinion Quarterly, 54*(3), 362-395.

Balden, W. (2004). *Multi-mode data collection: Benefits and downsides.* Paper presented at the 2004 Conference of the Great Lakes, Chapter of the Marketing Research Association, Cancun, Mexico.

Biemer, P. P. (1988). Measuring data quality. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II & J. Waksberg (Eds.), *Telephone survey methodology*. New York: John Wiley and Sons.

Bradburn, N. D., Sudman, S., & Wasink, B. (2004). *Asking questions: The definitive guide to questionnaire design for market research, opinion polls, and social and health questionnaires*. New York: John Wiley and Sons.

Bradburn, N. M., Sudman, S., Blair, E., & Stocking, C. (1978). Question threat and response bias. *Public Opinion Quarterly, 42*(2), 221-234.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116-131.

Cannell, C., Miller, P., & Oksenberg, L. (1981). Research on interviewing techniques. In S. Leinhardt (Ed.), *Sociological Methodology 1981* (pp. 389-437). San Francisco: Jossey-Bass.

Czaja, R., & Blair, J. (2004). *Designing surveys: A guide to decisions and procedures*. Pine Forge: Sage Publications.

de Leeuw, E. (1992). *Data quality in mail, telephone, and face-to-face surveys.* Amsterdam: TT Publications.

de Leeuw, E. (2005). To mix or not to mix? Data collection modes in surveys. *Journal of Official Statistics, 21*(2), 1-23.

de Leeuw, E., & van der Zouwen, J. (1988). Data quality in telephone and face-to-face surveys: A comparative analysis. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 283-299). New York: Wiley.

DeMaio, T. J. (1984). Social desirability and survey measurement: A review. In C. F. Turner & E. Martin (Eds.), *Surveying subjective phenomena* (Vol. 2, pp. 257-282). New York: Russell Sage Foundation.

Dillman, D. A. (1978). *Mail and telephone surveys: The total design method*. New York: Wiley-Interscience.

Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method* (2nd ed.). New York: John Wiley Co.

Dillman, D. A. (2007). *Mail and internet surveys: The tailored design method. 2007 update with new internet, visual and mixed mode guide*. Hoboken, NJ: John Wiley and Sons.

Erdman, H., Klein, M. H., & Greist, J. H. (1983). The reliability of a computer interview for drug use/ abuse information. *Behavior Research Methods and Instrumentation, 15*, 66-68.

Fowler, J., F.J. (1995). *Improving survey questions: Design and evaluation* (Vol. 38). Thousand Oaks, CA: Sage Publications.

Fowler, J., F.J., Roman, A. M., & Xiao Di, Z. (1998). Mode effects in a survey of medicare prostrate surgery patients. *Public Opinion Quarterly, 62*(1), 29-46.

Fricker, S., Galesic, M., Tourangeau, R., & Yan, T. (2005). An experimental comparison of web and telephone surveys. *Public Opinion Quarterly, 69*(3), 370-392.

Groves, R. M. (1979). Actors and Questions in Telephone and Personal Interview Surveys. *Public Opinion Quarterly, 43*(2), 190-205.

Groves, R. M. (1987). Research on survey data quality. *Public Opinion Quarterly, Winter*(50th Anniversary Issue), S156-S172.

Groves, R. M. (1989). *Survey errors and survey costs*. New York: Wiley.

Groves, R. M., & Couper, M. (1998). *Nonresponse in household interview surveys*. New York: John Wiley and Sons.

Groves, R. M., & Kahn, R. L. (1979). *Surveys by telephone: A national comparison with personal interviews*. New York, NY: Academic Press.

Groves, R. M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly, 68*(1), 2-31.

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-Salience Theory of survey participation: Description and an illustration. *Public Opinion Quarterly*, 299-308.

Hadaway, C. K., Marler, P. L., & Chaves, M. (1993). What the polls don't show: A closer look at U.S. church attendance. *American Sociological Review, Vol. 58*(No. 6), 741-752.

Hewitt, M. (2002). Attitudes towards interview mode and comparability of reporting sexual behaviour by personal interview and audio computer-assisted self-interviewing. *Sociological Methods and Research, 31*(1), 3-26.

Hochstim, J. R. (1967). A critical comparison of three strategies of collecting data from households. *Journal of the American Statistical Association, 62*, 976-989.

Holbrook, A. L., Cho, Y. I., & Johnson, T. P. (2006). *Extreme response style: Style or substance.* Paper presented at the Annual meeting of the Americal Association for Public Opinion Research, Montreal, Canada.

Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone vs. Face-to-Face Interviewing of National Probability Samples With Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias. *Public Opinion Quarterly, 67*, 79-125.

Holbrook, A. L., Krosnick, J. A., Moore, D., & Tourangeau, R. (In press). Response order effects in dichotomous categorical questions presented orally: The impact of question and respondent attributes. *Public Opinion Quarterly*.

Holtgraves, T. (2004). Social desirability and self-reports: Testing models of socially desirable responding. *Personality and Social Psychology Bulletin, 30*(2), 161-172.

Jäckle, A., Roberts, C. E., & Lynn, P. (2006). *Telephone versus face-to-face interviewing: Mode effects on data quality and likely causes (Report on phase II of the ESS-Gallup Mixed Mode Methodology Project)*. Colchester: University of Essex.

Jobe, J. B., Pratt, W. F., Tourangeau, R., Baldwin, A. K., & Rasinski, K. A. (1997). Effects of interview mode on sensitive quesitons in a fertility survey. In L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz & D. Trewin (Eds.), *Survey Measurement and Process Quality*. New York: John Wiley & Sons, Inc.

Jordan, L. A., Marcus, A. C., & Reeder, L. G. (1980). Response styles in telephone and household interviewing: A field experiment. *Public Opinion Quarterly, 44*, 210-222.

Jowell, R. (1998). How comparative is comparative research? *Amercian Behavioural Scientist, 42*(2), 168-177.
Karp, J. A., & Brockington, D. (2005). Social desirability and response validity: A comparative analysis of overreporting voter turnout in five countries. *The Journal of Politics, 67*(3), 825-840.
Knowles, E. S., & Condon, C. A. (1999). Why people say "yes": A dual-process theory of acquiescence. *Journal of Personality and Applied Psychology, 77*(2), 379-386.
Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*, 213-236.
Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50*, 537-567.
Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response order effects in survey measurement. *Public Opinion Quarterly, 51*, 201-219.
Krosnick, J. A., & Fabrigar, L. R. (Forthcoming). *The handbook of questionnaire design*. New York: Oxford University Press.
Krosnick, J. A., Narayan, S., & Smith, W. R. (1996). Satisficing in surveys: Initial evidence. In M. T. Braverman & J. K. Slater (Eds.), *Advances in Survey Research* (Vol. 70, pp. 29-44). San Francisco: Jossey-Bass Publishers.
Krysan, M., Schuman, H., Scott, L. J., & Beatty, P. (1994). Response rates and response content in mail versus face-to-face surveys. *Public Opinion Quarterly, 58*(3), 381-399.
Locander, W. B., & Burton, J. P. (1976). The effect of question form on gathering income data by telephone. *Journal of Marketing Research, Vol. 13*(No. 2, May), 189-192.
Locke, S. E., Kowaloff, H. B., Hoff, R. G., Safran, C., Popovsky, M. A., Cotton, D. J., et al. (1992). Computer-based interview for screening blood-donors for risk of HIV transmission. *Journal of the American Medical Association, 29*(1301-5).
Lucas, R. W., Mullen, P. J., Luna, C. B. X., & McInroy, D. C. (1977). Psychiatrist and computer interrogators of patients with alcohol-related illnesses: A comparison. *British Journal of Psychiatry, 131*, 160-167.
Lyberg, L., & Dean, P. (1992). *Methods for reducing nonresponse rates: A review.* Paper presented at the Annual meeting of the American Association for Public Opinion Research, St. Petersburg, Florida.
Mooney, G., Giesbrecht, L., & Shettle, C. (1993). *To pay or not to pay: That is the question.* Paper presented at the Meeting of the American Association for Public Opinion Research, St. Charles, IL.
Narayan, S., & Krosnick, J. A. (1996). Education moderates some response effects in attitude measurement. *Public Opinion Quarterly, Vol. 60*, 58-88.
Oppenheim, A. N. (1992). *Questionnaire design, interviewing and attitude measurement* (New ed.). London: Pinter Publishers.
Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*, 598-609.
Peytchev, A., Couper, M. P., McCabe, S. E., & Crawford, S. D. (2006). Web survey design: Paging versus scrolling. *Public Opinion Quarterly, 70*(4), 596-607.
Roberts, C. E., Jäckle, A., & Lynn, P. (2007). Causes of mode effects: Separating out interviewer and stimulus effects in comparisons of face-to-face and telephone surveys. *Proceedings of the Social Research Methods Section of the American Statistical Association*.
Roese, N. J., & Jamieson, D. W. (1993). Twenty years of bogus pipeline research: A critical review and meta-analysis. *Psychological Bulletin*(114), 363-375.

Saris, W. E., & Gallhofer, I. N. (2007). *Design, evaluation and analysis of questionnaires for survey research*. New York: Wiley.
Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments on question form, wording, and context*. New York: Academic.
Schwarz, N., Strack, F., Hippler, H. J., & Bishop, G. (1991). The impact of administration mode on response effects in survey measurement. *Applied Cognitive Psychology, 5*(3), 193-212.
Silver, B., D., Anderson, B., A., & Abramson, P., R. (1986). Who overreports voting? *American Political Science Review, Vol. 80*(No. 2 (June)), 613-624.
Singer, E., Hippler, H. J., & Schwarz, N. (1992). Confidentiality assurances in surveys: Reassurance or threat? *International Journal of Public Opinion Research, 4*(3), 256-268.
Smith, T. W. (1984). *A comparison of telephone and personal interviewing*. Chicago: National Opinion Research Center.
Smith, T. W. (1987). The art of asking questions, 1936-1985. *Public Opinion Quarterly, Vol.51*(Part 2: Supplement: 50th Anniversary Issue), S95-S108.
Sudman, S., & Bradburn, N. M. (1974). *Response effects in surveys: A review and synthesis*. Chicago: Aldine Publishing.
Sudman, S., & Bradburn, N. M. (1982). *Asking questions: A practical guide to questionnaire design*. San Francisco: Jossey-Bass.
Tourangeau, R. (1984). Cognitive science and survey methods. In T. B. Jabine, M. Straf, J. Tanur & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp. 73-100). Washington, DC: National Academy Press.
Tourangeau, R. (1987). Attitude measurement: A cognitive perspective. In H. J. Hippler, N. Schwarz & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 149-162). New York: Springer-Verlag.
Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly, 60*(2), 275-304.
Tourangeau, R., & Smith, T. W. (1998). Collecting sensitive information with different modes of data collection. In M. P. Couper, R. P. Baker, J. Bethlehem, C. Z. F. Clark, J. Martin, W. L. Nicholls II & J. M. O'Reilly (Eds.), *Computer assisted survey information collection*. New York: John Wiley & Sons, Inc.
Tourangeau, R., Smith, T. W., & Rasinski, K. A. (1997). Motivation to report sensitive behaviours on surveys: Evidence from a bogus pipeline experiment. *Journal of Applied Social Psychology, 27*(3), 209-222.
Voogt, R. J. J., & Saris, W. E. (2005). Mixed mode designs: Finding the balance between nonresponse bias and mode effects. *Journal of Official Statistics, 21*(3), 367-387.
Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the Amercian Statistical Association, 60*(63-69).
Weisberg, H. F. (2005). *The total survey error approach: A guide to the new science of survey research*. Chicago: University of Chicago Press.
Wright, D. L., Aquilino, W. S., & Supple, A. J. (1998). A comparison of computer-assisted and paper-and-pencil self-administered questionnaires in a survey on smoking, alcohol and drug use. *Public Opinion Quarterly, 62*, 331-353.

12 Appendix

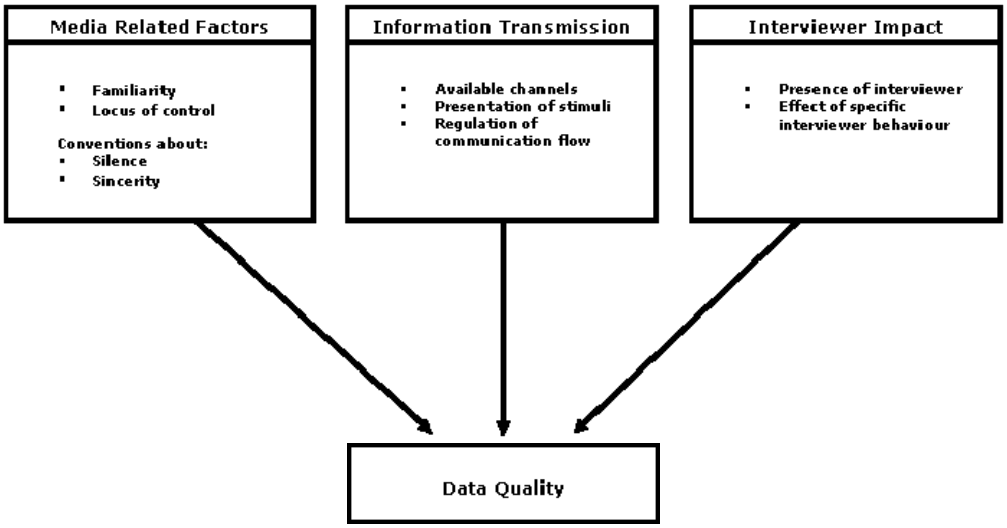

Fig. 1 de Leeuw's model of mode characteristics influencing data quality. (Source: de Leeuw, 1992; p.19)

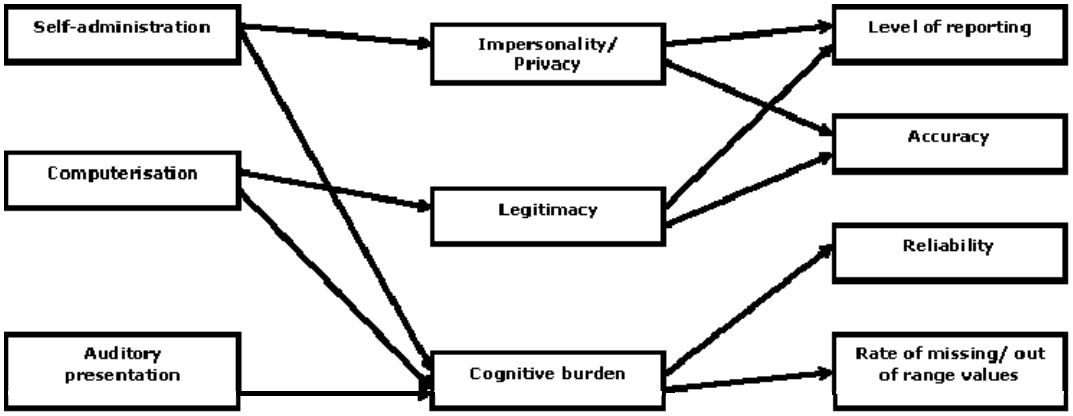

Fig. 2 Tourangeau, Rips and Rasinski's (2000) model of psychological variables mediating mode effects on data quality. (Source: Tourangeau, Rips and Rasinski, 2000; p. 306)

Table 1 – Characteristics of modes that influence data quality

Factors influencing the degree of effort required to answer questions		
1.	***Sensory channel in which the questions are presented***	One of the main distinctions between modes is whether or not the question stimulus is presented visually – as in traditional paper-and-pencil self-administered questionnaires (SAQs) and web surveys; aurally, as in telephone interviews and audio-SAQs such as Touchtone Data Entry; or using a mix of both, as in face-to-face interviews that make use of visual aids such as showcards, or as in Audio-CASI. This primary difference between the modes has implications for each of the factors listed below and is a key determinant of the degree of cognitive burden placed on the respondent.
2.	***Response medium***	Also relevant is the medium in which respondents are required to respond. Self-completion modes place the onus on the respondent to accurately record their answers to the questions, while interviewer-administered modes give interviewers that responsibility. Computer-assisted modes have helped to enhance the accuracy of survey data in both types of survey.
3.	***Locus of control***	Closely related to the question and response medium is the degree of control the respondent has over how the questions are administered. This is relevant to two factors in particular: the temporal order in which the questions are presented and the amount of time available to complete the questionnaire (see below). In interviewer-administered surveys, the respondent has almost no control while in self-administered modes the respondent has the maximum amount of control possible over how the questionnaire is completed.
4.	***Temporal order in which questions are presented***	In self-administered modes, respondents can choose to answer the questions either sequentially or in parallel, leaving questions out to answer others later in the questionnaire, then returning to previous questions they left unanswered. In interviewer-administered modes, the interviewer asks respondents the questions in the order in which they appear on the questionnaire.
5.	***Time pressure/ pace***	Self-completion respondents can usually take as much time as they need to fill in a questionnaire, taking breaks as required. In face-to-face and telephone surveys the interviewer typically dictates the pace of the interview, and particularly in telephone surveys, respondents may feel under particular pressure to keep up. Telephone interviews are sometimes designed not to exceed a specific length, but even where the same questionnaire is administered both by telephone and face-to-face, the former tends to be completed more quickly than the latter (Holbrook et al., 2003; Jäckle et al., 2006).
6.	***Conversational norms***	Related to the above, conventions about the acceptability of pauses in face-to-face and telephone interaction differ due to the possibility of filling gaps in conversation using non-verbal forms of communication when the interview is conducted in person. In telephone interactions, long silences can feel awkward; as a result, telephone interviews are likely to be conducted at a faster pace than interviews conducted face-to-face (de Leeuw, 1992; Groves and Kahn, 1979).
7.	***Familiarity with mode***	As new technologies expand the range of available mode options, respondents to surveys are likely to be less familiar with the modes they are invited to participate in. Familiarity can influence data quality by altering the cognitive demands of the survey. Most people are very familiar with the requirements of filling in a questionnaire on paper, but not all will have experience of using say, TDE, telephone A-CASI or the Internet. In countries where some modes are more commonly-used than others, even traditional modes that are widely-used elsewhere – such as face-to-face and telephone interviews – may feel unfamiliar to respondents and consequently more likely to

		influence the response process.
8.	***Use of computers***	Perhaps the most important developments in data collection technology has been the introduction of computer-assisted modes. For interviewer-administered surveys, computers have generally been shown to enhance data quality, by improving the accuracy with which answers are recorded, and ensuring that routing instructions are correctly followed. However, for respondents using CASI, either as part of a face-to-face interview, or over the Internet, the extent to which the computer facilitates the response process (and so enhances data quality) will depend of the respondent's experience of using computers.
9.	***Layout and format of the questionnaire***	An important aspect of self-administered modes of data collection which is frequently overlooked is the layout and visual design of the questionnaire (e.g. Dillman 2000; 2007). The development of web surveys has highlighted the importance of these aspects of survey design for all kinds of self-completion questionnaire (paper SAQs as well as web-based) and the underlined the potential impact of poor design on data quality and thereby, for data comparability in multi-mode survey designs. The format of the questionnaire in web mode (i.e. whether a scrolling method is used or a page-by-page interactive design) is also relevant here because of its potential impact on response quality (Peytchev et al., 2006).
10.	***Presence of external distraction***	Modes vary in the extent to which sources of external distraction are likely to influence respondents during questionnaire completion. The presence of an interviewer in a face-to-face survey may help to minimise the extent to which respondents can be distracted (though this is not always the case), as it may also to some extent in telephone interviews. In self-completion modes, however, the possibility of respondents' attention being divided is considerably more likely – a factor related to the degree of control respondents have over questionnaire administration.
11.	***Opportunities for multi-tasking***	Related to the presence of external distraction during questionnaire completion are the opportunities each mode provides for multi-tasking – or engaging in other activities whilst answering survey questions. Self-completion modes typically require greater concentration on behalf of respondents, but as noted, SAQ respondents have more freedom to do other things while they are taking part in a survey. Web surveys perhaps offer even greater opportunities for multitasking, as respondents are often quite adept at switching between windows to work on different tasks (de Leeuw, 2005). Mobile, hands-free and cordless telephones similarly make it easier for respondents to attempt other tasks whilst answering the interviewer's questions, although anecdotal evidence supports the conclusion that even face-to-face interviewers are not always able to prevent respondents from doing something else while taking part the survey (Krosnick, Narayan and Smith, 1996).
12.	***Presence of an interviewer***	A finding that emerges from each of the above is that the presence of an interviewer in face-to-face and telephone surveys helps to reduce the cognitive demands of the survey for respondents. Interviewers can positively influence the cognitive demands of the survey both directly – for example, by providing respondents with additional explanations about the meaning of particular survey questions and the requirements of the survey task (Schwarz et al., 1991) – and indirectly, by providing verbal and nonverbal reinforcement for positive response behaviour (Holbrook et al., 2003).
13.	***Channels of communication between interviewers***	Face-to-face interviews benefit from the availability of non-verbal means of communication, including body language, facial expressions and gestures and paralinguistic cues (such as clothing and other visual identifiers of status), all of which serve

	and respondents	to enrich the interaction between the interviewer and respondent (e.g. Groves and Kahn, 1979; Groves, 1979). By contrast, telephone interactions rely solely on audio communication, so they are not able to benefit from the visual channel, which may help to encourage and motivate respondents and to increase the sense of rapport between the respondent and interviewer (see below). Similarly, nonverbal signals can be used to fill silences in conversation and thus slow the pace of the interaction (see e.g. Groves and Kahn, 1979; Schwartz et al., 1991; Holbrook et al., 2003).
Factors influencing respondents' willingness to self-disclose		
1.	***Mental models of the survey procedure***	The way in which respondents approach surveys is partly influenced by their existing mental models about what is required. Telephone and face-to-face interviews, for example, tend to elicit different cognitive scripts governing how the interaction should proceed (Groves, 1989). According to Groves, in a face-to-face interview, the respondent might treat the interviewer in a similar way to how they would a guest, so the interaction may be governed more by norms about politeness, which in turn may lead the respondent to feel obligated to answer the questions as honestly and accurately as possible. By contrast, telephone interviews are more likely to elicit scripts similar to other cold-call attempts made by telephone, so the interaction is more likely to resemble a conversation with a stranger making a solicitation attempt than a household guest (Groves, 1989; p.510).
2.	***Perceived impersonality***	Modes vary according to how impersonal respondents perceive the data collection situation to be. The social distance between the respondent and the interviewer and/or researcher appears to be an important factor in this. Accordingly, in-person interviews are likely to be perceived as more personal than either telephone interviews or any self-completion method.
3.	***Perceived anonymity/ privacy of the reporting situation***	Related to the perceived impersonality of the interview are the respondents' perceptions of how anonymous the reporting situation is and how much privacy the mode affords them to answer questions openly and honestly. Self-completion modes offer the most privacy (and computer-administration may enhance this – Tourangeau et al., 2000). Respondents may perceive telephone interviews to offer more privacy than face-to-face interviews given the greater social distance between actors in the former compared to the latter.
4.	***Perceived confidentiality***	Perceptions of anonymity appear to be closely tied to respondents' concerns about the legitimacy of the survey and the confidentiality with which their answers are kept (Singer, Hippler and Schwarz, 1992). Modes vary in the extent to which they provide researchers with opportunities to dissipate respondents' concerns about data confidentiality. An interviewer is likely to be better able to reassure respondents about data confidentiality than the covering letter of a postal survey or an email invitation to participate in a web survey, particularly as they will be able to immediately respond to any questions the respondent has. In-person interviewers may be better able to reassure respondents on these issues than an interviewer on the telephone (Holbrook et al., 2003).
5.	***Perceived legitimacy of the survey***	Related to 3 and 4 are respondents' perceptions of the legitimacy of the survey. In addition to the points made in relation to perceived confidentiality, face-to-face interviewers may do a better job at conveying survey legitimacy (Groves, 1979), as they are able to present respondents with appropriate forms of identification (Holbrook et al., 2003), as well as with additional background material about the survey (such as leaflets) to

		reassure respondents of the sincerity of the research aims (de Leeuw, 1992; 2005).
6.	***Interviewer/ respondents rapport***	While self-completion surveys offer respondents greater privacy to report their attitudes and behaviours openly and honestly, respondents' willingness to self-disclose may also be enhanced by forming a close and trusting bond with the interviewer. As noted previously, factors such as the availability of non-verbal cues in face-to-face interactions serves to strengthen the rapport between interviewers and respondents and with it, respondents' willingness to answer openly (Groves and Kahn, 1979; Holbrook et al., 2003).

Sources: Groves, 1979; Groves and Kahn, 1979; Groves, 1989; Schwarz, Strack, Hippler and Bishop, 1991; de Leeuw, 1992; 2005; Tourangeau, Rips and Rasinski, 2000; Dillman, 2000; and Holbrook, Green and Krosnick, 2003