Survey Research Methods (2008)
Vol.2 , No.1 , pp. 21-32
ISSN 1864-3361
http://www.surveymethods.org

The Influence of the Design of Web Survey Questionnaires on the Quality of Responses

Stéphane Ganassali
I.R.E.G.E. - University of Savoie

The first objective of this article is to propose a conceptual framework of the effects of on-line questionnaire design on the quality of collected responses. Secondly, we present the results of an experiment where different protocols have been tested and compared in a randomised design using the basis of several quality indexes.
Starting from some previous categorizations, and from the main factors identified in the literature, we first propose an initial global framework of the questionnaire and question characteristics in a web survey, divided into five groups of factors. Our framework was built to follow the response process successive stages of the contact between the respondent and the questionnaire itself.
Then, because it has been studied in the survey methodology literature in a very restricted way, the concept of 'response quality' is discussed and extended with some more 'qualitative' criteria that could be helpful for researchers and practitioners, in order to obtain a deeper assessment of the survey output.
As an experiment, on the basis of the factors chosen as major characteristics of the questionnaire design, eight versions of a questionnaire related to young people's consumption patterns were created. The links to these on-line questionnaires were sent in November 2005 to a target of 10,000 young people. The article finally presents the results of our study and discusses the conclusions. Very interesting results come to light; especially regarding the influence of length, interaction and question wording dimensions on response quality. We discuss the effects of Web-questionnaire design characteristics on the quality of data.
Keywords: Web surveys, questionnaires, response quality

Introduction

Web-based surveys have been substantially developing for the last ten years. The Esomar association (2004) estimates that in the United States, more than one third of market research is now conducted through on-line surveys. An international professional panel run by a leading survey software editor on more than 7,000 institutions indicates that in 2006, 32% of them implemented on-line surveys, through an internal or an external network. Simultaneously to this practitioner concern, academic research was gradually becoming interested in the topic and was producing numerous contributions, in order to better understand these new methods of data collection. Logically, the first papers focused on the description of the various technological devices (Galan and Vernette 2000), with a view to pointing out the opportunities and the drawbacks of these new protocols (Ilieva, Baron and Healey 2002; Couper 2002). Internet surveys have been compared to other self-administered methods or to telephone protocols (Roster et al. 2004), mainly on the response rate criteria (Schmidt et al. 2005) and more recently on response quality (Fricker et al. 2005).

It is now established that web-based surveys are inexpensive, with a short response time and that they can achieve satisfying response rates compared to questionnaires delivered by 'classical' mail. Additionally, the nature and the quality of responses are not inevitably affected (Tuten, Urban and Bosnjak 2002). Some authors even suggest that on-line surveys provide more complete information than traditional mail surveys do (Ilieva, Baron and Healey 2002). They can also avoid some data quality problems such as social desirability bias (Fricker et al. 2005) or survey 'satisficing' patterns, (Skitka and Sargis 2005). For researchers, Internet surveys can also facilitate the use of embedded experiments (Mc Fadden et al. 2005).

After several years of experience, we consider that web surveys are especially well adapted to internal surveys (staff evaluation or social satisfaction), to access panels and more generally to a well identified target population, particularly in a Business-to-Business context (Roster et al. 2004), for customer satisfaction surveys for example. As far as Business to Consumer surveys are concerned, the medium coverage could still be a methodological difficulty. Even if this problem is now gradually decreasing, it could still be dissuasive in many cases. Because of the inability to identify all on-line users, web-based surveys do not provide generalisable results, due to self-selection, non-random and non-probabilistic sampling (Andrews, Nonnecke and

Contact information: Stéphane Ganassali, I.R.E.G.E. - University of Savoie, 4, chemin de Bellevue - BP 80439 - 74944 Annecy-le-Vieux Cedex - France, +33 450 09 24 00, email: sgana@univ-savoie.fr

Preece 2003). But comparing data from online and telephone (Schillewaert and Meulemeester 2005) or face-to-face protocols (Duffy et al. 2005), some experiments showed that the nature of the responses can be similar (for interests, attitudes or voting intentions for example) or sometimes different (knowledge or behaviour patterns). For some other authors (Roster et al. 2004), web surveys may be equally, if not more, accurate than phone surveys in predicting behaviours.

In academic research, a movement has been progressively established with a view to defining the circumstances in which we could obtain the *best response quality* in a Web survey. Our paper is part of this research trend. In fact, in recent publications, numerous experiments on the topic provide us with some very promising results. However, we think that past research have three major limitations, which we will mainly address in this article:

- no conceptual framework is really available to give an exhaustive description of the topic: a lot of experimental studies are available but most of them are often limited in scope and do not take all the specific aspects of a web survey into account,
- some important features specific to web surveys (illustration and especially interaction) have seldom been studied in the past,
- response quality has received much less research attention than response rate, and with a rather restricted view.

Then, the objectives of our research are:

1. to propose a general conceptual model in order to study the effects of on-line questionnaire design on the quality of collected responses,
2. to report experimental results to test how four major questionnaire features would influence a wider range of response quality indexes.

The decision to participate in a survey and the response process

Before proposing a theoretical framework of the determinants of response quality in a web-survey, it is necessary to describe the decision process implemented when a person is asked to participate in a survey and also the components of the response process itself. As far as the *decision process* is concerned, from a psychological point of view, several authors (Groves, Cialdini and Couper 1992) concentrated on the works made on "compliance warrants". The "compliance with requests" approach insists on six principles that influence the decision to perform a requested activity, such as the active participation in a survey: reciprocation, consistency ("desire to be consistent within attitudes, beliefs, words and deeds"), social validation (how similar others are acting), authority, scarcity and liking. In addition to these theoretical concepts, Groves, Cialdini and Couper (1992) introduced some practical knowledge drawn from professional survey interviewers' contributions. Two important components are outlined by experienced interviewers within the techniques they use to obtain good participation: tailoring (Groves and Couper 1998) and maintaining interaction. On the basis of the classical literature on persuasion and attitude change, Bosnjak and Tuten (2001) established the importance of motivation, opportunity and ability in the message information process. On the basis of these psychological theories, the authors reviewed the factors that may influence the participation in a survey. They are basically divided into four categories: the *societal-level factors*, the *characteristics of the sample person*, the attributes of the *interviewer* and finally, the attributes of the *survey design*. The performance of these attributes is affected by the sample person individual characteristics, on the basis of the "leverage-salience theory" proposed by Groves, Singer and Corning (2000).

The components of the *response process* itself were fully described by Tourangeau, Rips and Rasinski (2000). The process is made of four stages: comprehension, retrieval, judgment and finally response. For them, the presentation of the questionnaire is one of the most important variables that may affect the response process especially at both the comprehension and reporting (response) stages. They conclude their chapter on the comprehension component by giving some practical advice for survey designers, that are generally "consistent with the evidence and theoretical analyses" presented by the experts (Bradburn and Sudman 1979, Converse and Presser 1986). This advice covers various aspects of the questionnaire design, and focuses on the importance of simplicity of syntax.

Most of the authors say (Couper, Traugott and Lamias 2001, Dillman 2000) – and we also think – that as far as *web surveys* are concerned, a deeper investigation is needed on the relationships between the questionnaire characteristics and the response patterns. This is probably because they are in fact the only factors that can be really manipulated when implementing a web survey and also, because electronic surveys offer a wide range of design possibilities that can have a great influence on the quality of the collected data (Couper 2002). Moreover, it is known that within self-administrated surveys, in the absence of an interviewer, the respondent tends to seek information from the instrument itself: the verbal and visual elements of the questionnaire (Schwarz 1996).

The questionnaire characteristics affecting the response patterns

Within the literature dedicated to survey methodology, many contributions provide tips and hints on how to write a 'good' questionnaire in order to get 'good' responses. Despite the voluminous mass of relevant research data concerning response effects, few theoretical frameworks are available to structure this knowledge. Based on the Bradburn and Sudman (1979) proposals, Dijkstra and Van der Zouwen (1982) first designed a general model of the survey interview. It is divided into three sets of variables influencing the response patterns: the characteristics of the *questions* themselves, the *interviewer* variables and the *respondent* variables. More specifically, the question factors are split into two groups: the *formal characteristics* and the *content-related* ones. Structural-task characteristics (such as method of administration, instructions and so on) are described as

moderating variables that would condition the relationships between the basic factors and the response patterns.

On the basis of previous research, our framework is built to consistently follow the successive stages of contact between the respondent and the questionnaire itself. First, the person quickly sees the length of the questionnaire (or the number of screens) that represents the level of effort required to answer. Secondly, the respondent will get an impression of the conviviality of the form according to the balance between texts and illustrations. The third step would be the reading of the questions themselves and the assessment of the wording. Then the respondent is supposed to apprehend the interactive components of the survey-interview (tailoring) that could be crucial for web-surveys in the absence of an interviewer. Finally, the person would successively look at the response formats and would know exactly what kind of task and what kind of data is expected: ticks, numbers, texts, etc.

General structure and length of the questionnaire

From the beginning of the history of research on survey methodology and design, this first category of questionnaire characteristics has been very frequently studied and more specifically the length of the questionnaire. Common sense first suggests that a long questionnaire will obtain a lower response rate than a short one. Some contributions recommend an optimal length ranging between 15 and 30 questions for self-administrated questionnaires, even if it can be empirically considered as too brief for substantial market and academic research. Much research focuses on the effect of the length of the questionnaire on the return or response rates. In many general contributions on survey methodology (Foddy 1993 for example), one recommends a concise drafting. A too long questionnaire would produce on the respondent an effect of 'weariness'. A tendency to reproduce systematic answers (and thus to reduce their variability) is also reported at the end of long questionnaires (Herzog and Bachman 1981). As a matter of fact, the literature about either traditional mail surveys or Internet-based surveys provides mixed results. For traditional surveys, Dillmann's Total Design Method (1983) stated that a mail questionnaire must be perceived as easier and faster to complete and more visually appealing and interesting, to obtain higher response rates. However, a complete quantitative review concludes that the questionnaire length is almost uncorrelated with the response rate, there seems to be "a negative but very weak relation" between the variables (Yu and Cooper 1983). Then, if we consider the findings of three recent reviews made specifically on Web-based surveys, the results seem to be contrasted (Galesic 2002). On the one hand, statistically, the questionnaire length is not particularly associated to the response rate (Cook, Heath and Thompson 2000). On the other hand, researchers and practitioners stress the length of the questionnaire as the largest problem for high drop-out rates (Lozar Manfreda and Vehovar 2002).

Apart from the number of questions, the length of the questionnaire can also be perceived by the respondent on the basis of *the number of screens*, for example, the distinction between one and multiple-page design has been frequently discussed (Couper, Traugott and Lamias 2001; Lozar Manfreda, Batagelj and Vehovar 2002; Reips 2002; Ganassali and Moscarola 2004). One of the conclusions was that a one-page design resulted sometimes in higher item non-response or in more non-substantive answers. More generally, it is accepted that different questionnaire structures can lead to different response patterns.

Obviously, the length of the questionnaire is linked to the *required effort* perceived by the target audience of the survey. To help the respondent to estimate his/her position in the completion process (and to indicate how far they are from the end), some authors advise to use a point of completion (POC) indicator (Dillmann, Tortora and Bowker 1998). It seems that a POC indicator would reduce dropouts later in the survey (Healey, Macpherson and Kuijten 2005) but if the questionnaire is very long, it may not be effective in reducing break-offs (Conrad, Couper and Tourangeau 2003).

Internet-based surveys offer the opportunity to track more precisely the respondent behaviour during the interview session. Previously used by some researchers, the *log files* provide factual information about the on-line response process: how the person navigates from one screen to another, how many pages are seen and above all, where he or she quits. In our study, with the information available on the 'SphinxOnLine' survey server, we had the opportunity to analyse the survey log files in order to measure the drop-out rates and the 'drop-out points'.

Intensity of illustration

The Internet has introduced new opportunities for the illustration of questionnaires. In surveys, we can distinguish verbal information from visual information. This visual information can be displayed on three different levels: questions on which images play a major part (such as brand recognition questions for example), images as supplements with the text (i.e. embellishments, illustrations) and incidental images (i.e. background). The most problematic situation seems to be the second because the target audience might not know whether the images are designed as task or style elements (Couper, Tourangeau and Kenyon 2004). Opinions are divided on the impact of illustrations on the quality of the responses. On the one hand, pictures may enhance the *attractiveness of the questionnaire* and may make it more convivial for the respondent; on the other hand, these visual features make the questionnaire more difficult to access or complete, which could reduce the response rate (Deutskens et al. 2004). Couper, Tourangeau and Kenyon (2004) found little support for a positive effect of illustrations on respondents' enjoyment or reduction of perceived burden. However, exploring presentational influences, Ganassali and Moscarola (2004) have measured increased responses when relevant visual clues are presented in web interviews. More investigation is needed to test the effects of these various types of illustration on response quality.

Question wording

The length of the questions, the grammatical syntax and the level of language have been frequently studied in the survey methodology literature. In this area, we find a lot of general contributions that provide advice on "how to write good questions". For example, some authors traditionally suggest that the length of the questions should not exceed 20 words (Payne 1951) or that short questions would reduce the probability of respondents misunderstanding (Molenaar 1982). Some experimental research have tried to measure the effects of these wording features on the quality of the responses and they do not seem to strongly support this hypothesis (Bogen 1996).

As far as the wording itself is concerned, the methodological guidelines agree to recommend a 'simple syntax'. A complex wording would also lead to a higher probability of misunderstanding and consequently to a lower response quality (Foddy 1993). However, in the literature, the concept of 'complex syntax' is seldom defined or measured. Belson (1981), after studying more than 2,000 questionnaires, built a framework of the so-called 'difficult' questions where the more frequent ones are questions with special clauses, negative questions, questions including conjunctions, or multiple-timed questions. From another point of view, Brennan and Holdershaw (1999) demonstrated that the length, form and cue tone of an open-ended question have a significant effect on both the length and tone of the generated responses. On these aspects again, more detailed research is needed to define and evaluate the effects of the grammatical syntax of the questions, on the quality of the generated responses.

Interactivity

For most authors (see Stewart and Pavlou 2002), the particularity that differentiates Internet from other traditional media is its "potential for interactivity". Interaction is considered as one of the most relevant opportunities of the web-based interviews (Conrad, Couper and Tourangeau 2003). Nowadays, most of the Internet surveys are managed by an integrated server application. This means that – if it has been programmed – the server is able to react to the person's received answers. Usually, the answers are sent to the server at the end of each screen (or page), some applications even send the data at the end of each question. What could be the server's reactions? What do we mean by 'interaction' within a web survey?

Basically, the questionnaire can be adapted to the respondents' profile, presenting only the adapted questions. The server can provide some automatic jumps or display conditions. But, it can also provide some additional information (time elapsed, level of completion, details, explanations etc.) if apparently needed or asked by the respondent. The survey designer can also develop some conditional instructions or sentences, based on the previous answers or in the absence of any answer. For Best and Krueger (2004) or Conrad et al. (2005), researchers possess several interactive options for promoting more accurate survey data or optimising the quality of responses, such as progress indicators, missing data messages, answer feedbacks, continuation procedures or even social presence.

Some sophisticated softwares (see for example `www.sphinxonline.com`) allow the survey designer to easily develop some various and complex scenarios. This is to create and maintain a kind of interaction with the respondent, taking into account multiple combinations of his/her previous answers to ask him/her very personalised questions. The ultimate objective would be to simulate a so-called "tailored" interview: a very important component outlined by experienced interviewers to obtain good participation (Groves and Couper 1998).

A few recent works have studied the impact of automatic interactive questionnaires on the quality of the collected data. But we must say that the interactive features that have been analysed are (on average) not very rich and that few complex situations have been experimented. Nevertheless, the first results (Ganassali and Moscarola 2004) give some interesting perspectives on the effect of these features, and our research will try to go further in the investigation.

Response format

The impact of the response format on the quality of the collected data in self-administrated surveys has been frequently studied. It has been demonstrated (Reips 2002, Christian and Dillman 2004) that the graphical presentation of the response formats do influence the answers to Internet surveys. For example, Smith (1995) recognized that when respondents were given a larger space for an open-ended question, they wrote longer answers. As far as closed questions are concerned, some experiments (Heerwegh and Loosveldt 2002), showed that radio-buttons can slightly be preferred to drop-boxes, because of their positive impact on response rates in a Web context. In addition to that, the choice of the response format (radio-buttons vs. drop-boxes again) may clearly lead to different response distributions (Couper et al. 2004). More precisely, some other experiments (Tourangeau, Couper and Conrad 2004) demonstrated that even the spacing of the response options affects the selection of the answers.

The quality of the responses within a web-based survey: an extended definition

The different sources of errors in surveys are well identified. A reference work on the topic (Groves 1989) distinguishes four major types of error sources: coverage, sampling, non-response and measurement. Our research is mainly focused on the two last mentioned sources. For web studies, error can be enhanced by the respondent's attitude towards the survey (motivation, comprehension, distortion etc.) and of course, by the questionnaire itself: design, wording, illustration etc. We will also give special attention to non-response error that could be critical within the context of on-line studies, especially with the problem of the 'drop-outs'.

But before that, the concept of 'quality of responses' needs to be discussed and enlarged, probably because it has been studied in a quite restricted way in past survey methodology literature. Response quality has received much less research attention than response rates (Schmidt et al. 2005). Too often, researchers make the confusion between quality of 'data' and quality of 'responses'. The quality of data is only considered from a methodological point of view, as if collecting the data itself was the ultimate achievement of the survey. When you are running a survey, you do not need only 'data' but you need also good-quality responses to the questions you asked.

The most common indicators of data quality are linked to the non-response error, and the most frequent output that are measured are the non-response and the completion rates. We think that the notion of quality of *responses* could be extended with some more 'qualitative' criteria that could be helpful for researchers and practitioners, in order to obtain a deeper assessment of the survey output. In a review, apart from response rate or speed, Tuten, Urban and Bosnjak (2002) have identified four dimensions of response quality: item omission, response error, completeness of answer and equivalence of response (between modes of data collection). For Schonlau, Fricker and Elliot (2002), data quality can be judged using several criteria such as unit and item non-response, completeness of response (particularly for open-ended questions), honesty of response and transcription error rate. Then, another main purpose of our article is to suggest an *extension of the measurement of the quality of the responses*.

The response rate

Among all the criteria that have been studied as indexes of quality, the *response rate* is the most frequent (Jackob and Zerback 2006). Recently (Kaplowitz, Hadlock and Levine 2004), it has been confirmed that web surveys can achieve comparable response rates to a questionnaire delivered by 'classical' mail. Two important meta-analyses are available to get a complete overview of the factors that could influence response rates within web-based surveys. This research studied respectively 68 (Cook, Heath and Thompson 2000) and 102 (Lozar Manfreda and Vehovar 2002) papers. One of their global conclusions is that between all the different characteristics of a study, the number, the persistence and the personalisation of the contacts are the dominant factors affecting response rates in web surveys.

The 'drop-out' rate

The *drop-out rate* represents the frequency of the respondents who started the survey and finally did not end it. An exit after viewing only the first screen of the questionnaire is considered as a drop-out as well. The technological opportunities offered by web surveys allow us to track and to identify those persons who quit the survey, thanks to the log files created by the web-survey server system. Few articles focus on this very interesting indicator of the efficiency of the study process. The drop-out rate can be a substantial problem within some Internet surveys and can reach a frequency of 15-20% (Healey, Macpherson and Kuijten 2005). Some authors (Knapp and Heidingsfelder 1999) showed that higher quit rates are produced when using open-ended questions or questions arranged in tables.

The filling-up rate

Completeness was described as one of the main components of response quality (Goetz, Tyler and Cook 1984; Tuten, Urban and Bosnjak 2002). In prior research, item non-response is frequently used to assess the quality of the data (Schmidt et al. 2005; Roster et al. 2004 for example). This variable can be measured by the number of "no opinions" or "don't knows" (Fricker et al. 2005). More globally, indicating the proportion of the completed questions on the overall number of the questions in the survey, the *"filling-up" rate* is seldom taken into account as a possible assessment of the quality of the collected data. Cobanoglu, Warde and Moreo (2001) used an index (called "completeness") to compare the response quality from mail, fax and e-mail surveys. In a research that was relatively similar to ours, Deutskens et al. (2004) established that the number of "don't knows" and semi-complete answers were slightly higher in a long and visual version of a given questionnaire.

The abundance of responses: response length and number of responses

Taking into account the length of the responses to open-ended questions (Sproull 1986; MacElroy, Micucki and McDowell 2002; Ganassali and Moscarola 2004; Schmidt et al. 2005; Deutskens, de Ruyter and Wetzels 2006) or less frequently the number of items quoted in the multiple-choice questions (Healey, MacPherson and Kuijten 2005), we are able to evaluate the levels of involvement and consented effort made by the respondent. From an early research, some results suggested that in e-mail surveys, responses for open-ended questions were longer than in traditional mail studies (Sproull 1986). Some authors even propose to quote (by doing a content analysis) the number of themes evoked in the open-ended responses (Healey, MacPherson and Kuijten 2005). These new quality indexes (that we call 'abundance') were seldom analysed in the past and we think it can be crucial to incorporate it in our model and in our experiment.

The variety or differentiation of responses

In the lists of scale questions (very often used in the surveys, such as satisfaction surveys), the respondents sometimes tend to choose more or less the same points and only pick a very narrow range of responses from all the possibilities. This behaviour pattern, which is called "use of subscales" or *"non-differentiation"* has been described in some previous research (Ray and Muller 2004; Fricker et al. 2005). It shows a lack of interest and a weak level of effort for answering. Variety in the answers would match a high degree of involvement in the survey. This concept was studied in a

Table 1: Justifications of the questionnaire assessments by the respondents (content analysis)

Topics of the comments	n	percent
Response formats	91	16.6%
Wording	89	16.2%
General structure	53	9.7%
Length	52	9.5%
Illustration	30	5.5%
Interaction	19	3.5%
Total	549	

survey run within an access panel (Gritz 2004), and also in an experimental comparison of web and telephone surveys (Fricker et al. 2005).

The satisfaction of the respondent

The respondent's satisfaction could also be a predictor of the questionnaire's ability to maximise response quantity and quality. Of course, it has nothing to do with the real quality of the collected data, but we find it could be interesting to see how quality measures and satisfaction are linked together. At the end of the survey which we describe in section 3 below, we asked the respondents to rate the questionnaire with a mark from 0 (very poor) to 10 (excellent). Then, with an open question, we asked them to justify their assessment, by explaining why they gave such a mark. From the content analysis made conjointly by 3 experts (on a sample of 550 answers), we identified the six most frequent criteria quoted by the respondents to illustrate their satisfaction with the questionnaire design characteristics. 47% of the sample mentioned at least one of the formal characteristics of the questionnaire. Among these answers, the wording and the response formats came first, then, the length and the general structure. Finally, illustration and interactivity were evoked. These results show that our model is quite coherent with the comments spontaneously given by the respondents of our Internet survey.

After reviewing the questionnaire characteristics affecting the response patterns and the indexes of response quality, we developed a general *theoretical framework*. Starting from some previous categorisations (Grémy 1987; Delamater 1982; Molenaar 1982), and using the main factors identified the literature (Couper, Traugott and Lamias 2001), we can propose a first global framework of the questionnaire characteristics in a web survey, divided into five groups of factors. Grémy (1987) proposed a first categorization including three groups of factors: the responses modalities, the wording of the questions and the questions context (order, instructions and so on).

Methodology: eight versions of the same questionnaire to test the effects on response quality

On the basis of the factors that we chose as major characteristics of the questionnaire design, we created *eight versions of a questionnaire related to young people's consumption patterns*. We did not design a complete experimentation because it was quite demanding and most of all, we feared to have too few answers for each version. For this reason, we first selected only two levels for each factor in order to simplify the experiment.

The links to those eight on-line questionnaires (see table 2) were sent in November 2005 to a target of *11,200 young people* composed of students of the University of Savoie and external private contacts (friends and relatives) given by the students involved in this project. They are aged from 18 to 25, 58% of them were female and 30% are workers. Of course, these demographics are stable among the eight randomly extracted sub-samples.

Two weeks after launching the survey, we had received 1,935 answers, representing a *global response rate of 17.28%*. One single follow-up was sent right after the majority of respondents reacted to the initial e-mailing and it helped in maximising the response rate (Dillman 2000). No real incentive was used but within the text of the e-mail, we insisted on the fact that this survey was part of an important educational and research project and we presume that it had a very strong positive impact on the participation in the survey.

Implementation of the questionnaire variables

Length: we developed two versions of the questionnaire, the short one had 20 questions and the long one had 42 questions. The length of the survey was not mentioned to sample members prior to completion, but only on the first screen of the questionnaire (via page number indicator).

Illustration: the plain version had no illustration, the illustrated one included 21 photographs.

Wording: we wrote a direct-wording version of the questionnaire and another one with a more complex style. To be more objective, we ran the Flesh analysis (available in Microsoft Word©) that indicates an approximate reading grade level. The direct version obtained 60/100 and the sophisticated one 40/100 (0 = very difficult reading, 100 = very easy reading).

Interaction: we designed a first copy without any interaction in it and a second one including some repetitions of previous answers and also some kinds of 'forced answering': for example if the open-ended question about advertising was not completed, another request for an answer was presented.

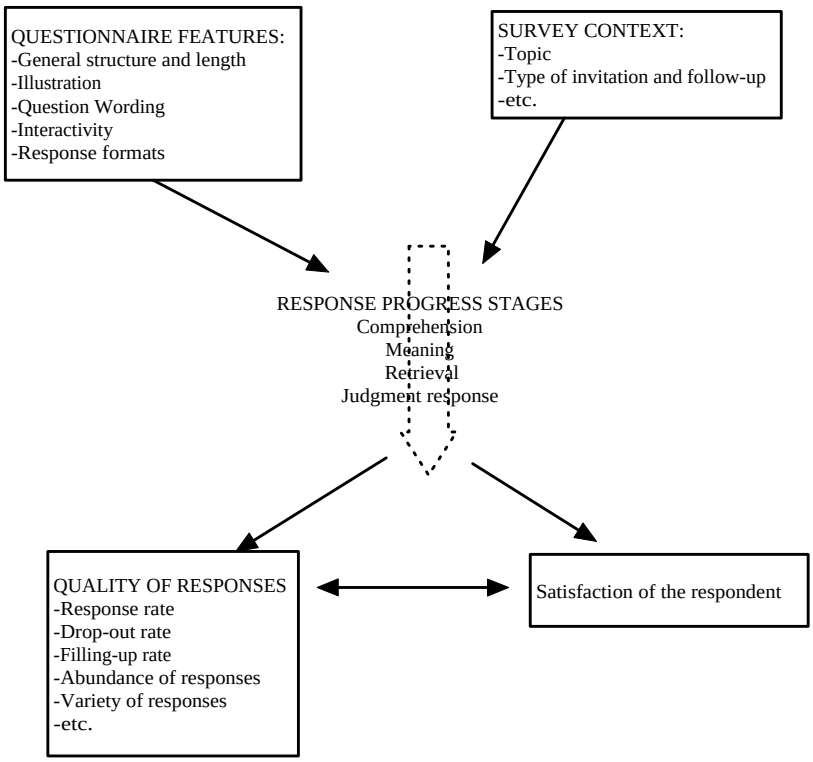

Figure 1. Conceptual framework of the impact of questionnaire features on quality of responses

Table 2: The eight versions of the survey questionnaire and urls where they can be seen

N^o	Length	Illustration	Wording	Interaction	url
1	Long	Illustrated	Direct	Interactive	http://ate-j165.univ-savoie.fr/young/young/q1.htm
2	Short	Illustrated	Direct	Interactive	http://ate-j165.univ-savoie.fr/young/young/q2.htm
3	Long	Plain	Complex	Interactive	http://ate-j165.univ-savoie.fr/young/young/q3.htm
4	Short	Illustrated	Complex	Interactive	http://ate-j165.univ-savoie.fr/young/young/q4.htm
5	Short	Plain	Complex	Linear	http://ate-j165.univ-savoie.fr/young/young/q5.htm
6	Long	Plain	Complex	Linear	http://ate-j165.univ-savoie.fr/young/young/q6.htm
7	Short	Illustrated	Direct	Linear	http://ate-j165.univ-savoie.fr/young/young/q7.htm
8	Long	Plain	Direct	Linear	http://ate-j165.univ-savoie.fr/young/young/q8.htm

As described before, we came up with eight versions of the survey (still available on-line) mixing the questionnaire design features as follows:

Measurement of response quality indexes

The response rate was easy to compute but additionally we had the opportunity to track the log files on the server with a view to obtaining a measurement of the *drop-out rates*. We decided not to study the response rate because many contributions are now available on this topic. Some meta-analyses have already demonstrated that the invitation variables (persistence and personalisation of contacts) are probably the most crucial factors.

According to the log files registered on the server, we had a global drop-out rate of 27%. Apart from the frequency of the drop-outs, we found it important to visualise the specific screens where the respondents usually quit. The next figure shows that drop-outs are more frequent (close to 50%) on the second screen of the questionnaires, from both short and long versions.

It would mean that most of the respondents made a quick decision in the first seconds, whether to participate or not in the survey after a quick overview of the two first screens. We can see from the graph that the drop-outs are not so frequent in the last screens. Then, if the length of the questionnaire

has a positive effect on abandon, it is the announced length and its perception more than the 'real' length that was endured along the answering process.

The *filling-up rate* was directly calculated by the software we used to process the examination of the data (Sphinx Survey©) but we needed to create some new variables in order to quantify the *abundance of the responses*: we designed an overall measure of the length of the responses to the open-ended questions, combining all the words obtained in all these questions. We call this new variable a 'verbose'. Then the *variety of the responses* was assessed by the number of different points used as answers within the list of ten common scale questions. This number is then ranged from 1 to 4. The *satisfaction of the respondent* was basically measured with a final question where the person was asked to give a mark to the questionnaire, from 0 (very poor) to 10 (excellent).

Hypotheses

According to this theoretical background, our individual hypotheses are the following:

1. a short questionnaire would produce better-quality responses, especially less drop-outs and a higher filling-up rate,
2. a direct wording would also lead to a better retention and a higher filling-up rate,
3. illustration is expected to facilitate abundance of responses,
4. interaction would produce longer and more varied responses.

Then, our overall hypothesis is that Questionnaire n^{o}2 (Short-Illustrated-Direct-Interactive) would obtain the best global quality of responses. We think that a restricted length, a direct wording, an illustrated layout and an interactive survey process would lead to a lower drop-out rate, a higher filling-up rate, more abundant responses to open-ended questions, more various answers to scale questions and also, a better respondent satisfaction.

Summary of results

Because the dependant variables are numeric and the independent variables are categorical, and because we decided not to design a complete experimentation, we often used the principal effects analysis of variance to process our analysis with Statistica.

The influence of length and wording on the drop-out rate

Categorical regression was used to test the effects of 3 variables on the drop-outs. We could not use the 'interaction' variable because of technical reasons: the log files were not available with this kind of questionnaire.

Unsurprisingly, we can see from the next table that you can reach a high drop-out rate of 34% with a long and complex questionnaire ($p < 0.1\%$), whereas a short and direct version would reduce the drop-outs to 21%, which is really a substantial gain.

Table 3: Determinants of the drop-out rate

Effect	-2 log-Likelihood of reduced model	χ^2	Deg. of Freedom	Signif.
Constant	36.86	0.00	0	
Length	*42.35*	*5.49*	*1*	*0.01**
Wording	*41.99*	*5.12*	*1*	*0.02**
Illustration	37.89	1.02	1	0.31
Interaction	Not available			

Table 4: Effects of length and wording on the drop-out rate

Drop-out rates/ Length & wording	Direct	Complex	Total
Long	26%	*34%*	29%
Short	21%	27%	23%
Total	24%	31%	27%

The impact of interaction on the filling-up rate

The overall filling-up rate was very high in our study, reaching an average of 97%. It was not coherent to compare the filling-up rates from long versus short questionnaires: if only one question was skipped in the short form, the filling-up rate was 94,7%, if one question was skipped in the long one, it was 97,6%. The rate was 'mechanically' dependent of the number of proposed questions. So, we had to neutralise the length characteristic in this set of analyses. Interactive questionnaires seem to clearly facilitate a high filling-up rate. We must mention too that the 'wording' factor was close to being significant ($p = 0.08$) and that the direct versions would obtain higher filling-up rates.

Influences of length and interaction on the response abundance

From the last table, it is interesting to see that the long and interactive questionnaire obtained the longer textual responses (average length 78 words) while the linear forms produced shorter texts, approximately 60 words, that is 25% less ($p < 0.1\%$).

Table 5: Determinants of the filling-up rate

	Sum of Squares	Deg. of Freedom	Mean Square	F	*p*
Interaction	*93.47*	*1*	*93.47*	*3.99*	*0.04**
Wording	72.09	1	72.09	3.08	0.08
Illustration	58.24	1	58.24	2.49	0.11

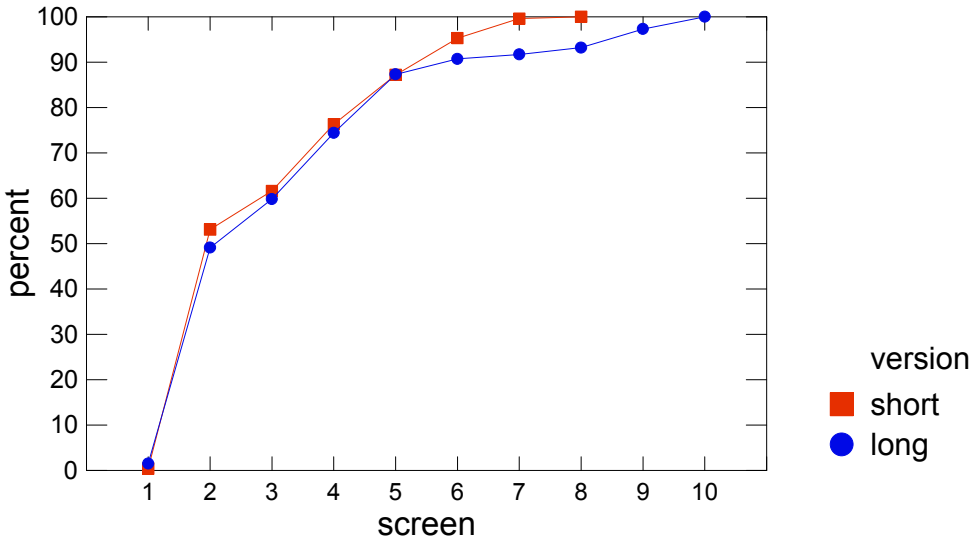

Figure 2. Cumulative drop-out percentages for short and long version questionnaires

Table 6: Determinants of the filling-up rate

	Sum of Squares	Deg. of Freedom	Mean Square	F	*p*
Length	*19562*	*1*	*19562*	*8.22*	*0.00**
Interaction	*18763*	*1*	*18763*	*7.89*	*0.00**
Wording	609	1	609	0.25	0.61
Illustration	1366	1	1366	0.57	0.44

Table 7: Effects of length and interaction on the response abundance

Abundance of responses/ Length & interaction	Interactive	Linear	Total
Short	64	62	63
Long	*78*	61	71
Total	72	62	67

Table 8: Determinants of the variety of responses

	Sum of Squares	Deg. of Freedom	Mean Square	F	*p*
Length	0.05	1	0.05	0.13	0.71
Interaction	0.77	1	0.77	1.88	0.17
Wording	0.29	1	0.29	0.72	0.39
Illustration	0.00	1	0.00	0.01	0.98

Table 9: Determinants of the satisfaction of the respondent

	Sum of Squares	Deg. of Freedom	Mean Square	F	*p*
Length	*94.23*	*1*	*94.23*	*31.52*	*0.00**
Interaction	0.01	1	0.01	0.00	0.95
Wording	0.15	1	0.15	0.05	0.82
Illustration	0.14	1	0.14	0.05	0.83

Variety of responses

No one design characteristic seems to have an impact on the variety of responses. Maybe it was too difficult to have high variance because there were 'only' ten common scale questions and four possible choices in the scale. This could be clearly a limitation of our study.

Satisfaction of the respondent

It was surprising to discover that the long questionnaire generates a significantly higher respondent satisfaction than the short one: 6,75 versus 6,10. The other questionnaire components are absolutely not significant.

Discussion

As we can see in table 10, our hypotheses are only partially confirmed. As a first step, the summary of our results first suggests that *a perceived short and direct-style wording of on-line questionnaires could significantly reduce the drop-out rate* (from 34% to 21% in our study). Concerning the length of the survey, the respondents had only the possibility to visualise the total number of screens, because we did not use any POC indicators that was recommended by Dillman, Tortora and Bowker (1998) that - maybe - could have min-

Table 10: Summary of results

	Short	Illustrated	Direct	Interactive
Drop-out rate	+		+	(not tested)
Filling-up rate			+	+
Abundance	-			+
Variety				
Satisfaction	-			

imised the drop-outs again.

Apparently, this perception is critical in the very first screens of the questionnaire because the analysis of the log files on the server indicated that *drop-outs are notably more frequent on the second screen of the questionnaires* (50%), from both short and long versions. Then, we can say that the decision to quit the survey is influenced by perceived length and by style of wording, on the very first pages of the form.

However, once a respondent is persuaded to continue beyond the first two screens, *an interactive questionnaire would have very positive effects on the quality of the collected data*. Interaction in the survey process would generate a higher filling-up rate and richer responses (defined in our research by longer answers to open-ended questions). Obviously, those effects would depend on the nature of the interactivity. Interactive features that were introduced in our experiment were likely to enhance the perception of the importance and seriousness of the survey by our respondents. Other kinds of interactivity could produce different effects.

In this experiment, the illustration of the questionnaire had no impact on the quality of the collected data. We have to mention that the pictures we used were only simple 'illustrations' of the questions. They were neither response modalities nor illustrations of the response proposals, as mentioned and tested in some past research (Ganassali and Moscarola 2004). It means that the pictures were used at the lower level of illustration. More research is needed to test the impact of the other types of illustrations that can be implemented in a web-based survey.

Finally, for these motivated and involved respondents, *the length of the survey does not seem to be an obstacle any longer*. In contrast, our results suggest that they may produce more abundant responses within a long survey and may feel significantly more satisfied in the end.

Conclusion: limits and implications for future research

A few limits naturally exist in our research. From a theoretical point of view, basically, we believe that one of the most important determinants of response quality in web surveys is the contacting procedure. It would have been very interesting to analyse how invitation aspects would combine with the tested questionnaire features, in a more global theoretical model.

As far as methodology is concerned, first of all, we were not able to run a more complete experimentation plan, that could have been useful for a richer analysis and also to study interaction effects between the independent variables. Secondly, the questionnaire characteristics had only two levels. It is probably not enough to study for instance how the length of the questionnaire could precisely affect the drop-out patterns. We could expect the relationships between the questionnaire features and the data quality indexes to be non-linear but this could be studied with at least three or four levels for each independent variable. But of course, these two last improvements would have resulted in a very complex and heavy experimentation plan, with many different versions of the on-line questionnaire.

For the same reasons, it was not feasible to test all the questionnaire design characteristics that came out of our literature review. Above all, we regret that the questionnaire's general structure (logic and progression) was not taken into account in our experiment. It is seldom studied in the literature but we discovered in our study that it was quite frequently pointed out by the respondents as a justification of their evaluation of the questionnaire.

The second methodological limit would be the typicality of the sample that we used for the research. It is only composed of young people from 18 to 25, which could represent a major restriction for the generalisation of the results we observed.

As a conclusion, we think that the three most promising results of our study are:

- first the *influence of interaction*,
- secondly, the appraisal of *drop-outs patterns*,
- and thirdly, the operational improvement of the concept of *response quality*.

For web surveys, it was obvious that the interactive possibilities needed to be more deeply described and analysed. In our experiment, we only tested a small part of the available technical options. Missing data messages, on-line help for respondents, confirmation and validation of the previous answers are new technical possibilities offered by the on-line survey process that could be implemented in order to test their potential impact on response quality.

As far as the drop-outs are concerned, we think our study provides an interesting understanding of this specific non-response pattern. Our results show that it is probably possible to reduce substantially the losses with shorter and more direct questionnaires.

Finally, our extension of the implementation of response quality is a claim to an evolution towards a more enlarged measurement of this concept. In the specific context of surveys, we think we should reconsider the information coming from respondents, with a more pragmatic and useful point of view, as "answers-to-questions" instead of only "data".

References

Andrews, D., Nonnecke, B., & Preece, J. (2003). Electronic survey methodology: a case study in reaching hard to involve internet users. *International Journal of Human-Computer Interaction*, *16*(2), 185-210.

Belson, W. A. (1981). *The design and understanding of survey questions*. Aldershot: Gower.

Best, S. J., & Krueger, B. S. (2004). *Internet data collection, quantitative applications in the social sciences*. Thousand Oaks: Sage Publications.

Bogen, K. (1996). *The effect of questionnaire length on response rates - a review of the literature.* (Proceedings of the Survey Research Methods Section, Alexandria: American Statistical Association, pp. 1020-1025)

Bosnjak, M., & Tuten, T. L. (2001). Classifying response behaviours in web-based surveys. *Journal of Computer-Mediated Communication*, *6*(3). (Online)

Bradburn, N. M., & Sudman, S. (1979). *Improving interview method and questionnaire design*. San Francisco: Jossey-Bass.

Brennan, M., & Holdershaw, J. (1999). The effect of question tone and form on responses to open-ended questions: Further data. *Marketing Bulletin*, *10*, 57-64.

Christian, L. M., & Dillman, D. A. (2004). The influence of symbolic and graphical language manipulations on answers to self-administered questionnaires. *Public Opinion Quarterly*, *68*, 57-80.

Cobanoglu, C., Warde, B., & Moreo, P. J. (2001). A comparison of mail, fax and web-based survey methods. *International Journal of Market Research*, *43*(4), 441-452.

Conrad, F. G., Couper, M. P., & Tourangeau, R. (2003). *Interactive features in web surveys.* (Joint Meetings of the American Statistical Association, Program on the Funding Opportunity in Survey Research Seminar. Washington D.C.: Federal Committee on Statistical Methodology)

Conrad, F. G., Couper, M. P., Tourangeau, R., & Galesic, M. (2005). *Interactive feedback can improve quality of responses in web surveys.* (60^{th} Annual Conference of the American Association for Public Opinion Research, Miami Beach.)

Converse, J. M., & Presser, S. (1986). *Survey questions: Handcrafting the standardized questionnaire. quantitative applications in the social sciences*. Thousand Oaks: Sage Publications. (Sage University Papers 07-063)

Cook, C., Heath, F., & Thompson, R. L. (2000). A meta-analysis of response rates in web-or internet-based surveys. *Educational and Psychological Measurement*, *60*, 821-836.

Couper, M. P. (2002). *New technologies and survey data collection: Challenges and opportunities.* (International Conference on Improving Surveys, Copenhagen, AAPOR)

Couper, M. P., Tourangeau, R., Conrad, F. G., & Crawford, S. D. (2004). What they see is what we get - response options for web surveys. *Social Science Computer Review*, *11*(1), 111-127.

Couper, M. P., Traugott, M. W., & Lamias, M. J. (2001). Web survey design and administration. *Public Opinion Quarterly*, *65*, 230-253.

Delamater, J. (1982). Response-effects of question content. In W. Dijkstra & J. Van der Zouwen (Eds.), *Response behavior in the survey interview* (p. 13-48). London: Academic Press.

Deutskens, E., de Ruyter, K., & Wetzels, M. (2006). *An assessment of equivalence between online and mail surveys in service research.* (Journal of Service Research (forthcoming))

Deutskens, E., de Ruyter, K., Wetzels, M., & Oosterveld, P. (2004). Response rate and response quality of internet-based surveys: An experimental study. *Marketing Letters*, *15*(1), 21-36.

Dijkstra, W., & Van der Zouwen, J. (1982). *Response behaviour in the survey-interview*. London: Academic Press.

Dillman, D. A. (1983). Mail and other self-adminisrated questionnaires. In P. H. Rossi, J. D. Wright, & A. B. Anderson (Eds.), *Handbook of survey research* (p. 359-377). San Diego: Academic Press.

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

Dillman, D. A., Tortora, R. D., & Bowker, D. (1998). *Principles for constructing web surveys*. Washington: Pullman.

Duffy, B., Smith, K., Terhanian, G., & Bremer, J. (2005). Comparing data from online and face-to-face surveys. *International Journal of Market Research*, *47*(6), 615-639.

Foddy, W. (1993). *Constructing questions for interviews and questionnaires*. Cambridge: Cambridge University Press.

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

Galan, J.-P., & Vernette, E. (2000). Vers une quatrième génération: les études de marché on-line. *Décisions Marketing*, *19*, 39-52.

Galesic, M. (2002). *Effects of questionnaire length on response rates: Review of findings and guidelines for future research.* (5^{th} German Online Research Conference, Hohenheim)

Ganassali, S., & Moscarola, J. (2004). Protocoles d'enquête et efficacité des sondages par internet. *Décisions Marketing*, *33*, 63-75.

Goetz, E. G., Tyler, T. R., & Lomax Cook, F. (1984). Promised incentives in media research: A look at data quality, sample representativeness, and response rate. *Journal of Marketing Research*, *21*(2), 148-154.

Göritz, A. (2004). The impact of material incentives on response quantity, response quality, sample composition, survey outcome, and cost in online access panels. *International Journal of Market Research*, *46*(3), 327-345.

Grémy, J.-P. (1987). Les expériences françaises sur la formulation des questions d'enquête. Résultats d'un premier inventaire. *Revue Française de Sociologie*, *28*(4), 567-599.

Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly*, *56*, 475-495.

Groves, R. M., & Couper, M. P. (1998). *Non-response in household interview surveys*. New York: Wiley & Sons.

Groves, R. M., Singer, E., & Corning, A. D. (2000). Leverage-salience theory of survey participation. *Public Opinion Quarterly*, *64*, 299-308.

Healey, B., MacPherson, T., & Kuijten, B. (2005). An empirical evaluation of three web survey design principles. *Marketing Bulletin*, *16*, 1-9. (Research Note 2)

Heerwegh, D., & Loosveldt, G. (2002). An evaluation of the effect of response formats on data quality in web surveys. *Social Science Computer Review*, *20*(4), 471-485.

Herzog, A. R., & Bachman, J. G. (1981). Effects of questionnaire length on response quality. *Public Opinion Quarterly*, *45*, 549-559.

Ilieva, J., Baron, S., & Healey, N. M. (2002). Online surveys in marketing research: Pros and cons. *International Journal of Market Research*, *44*(3), 361-376.

Jackob, N., & Zerback, T. (2006). *Improving quality by lowering non-response. a guideline for online surveys.* (W.A.P.O.R. Seminar on Quality Criteria in Survey Research, Cadenabbia, Italy)

Kaplowitz, M. D., Hadlock, T. D., & Levine, R. (2004). A comparison of web and mail survey response rates. *Public Opinion Quarterly*, *68*(1), 94-101.

Knapp, F., & Heidingsfelder, M. (1999). *Drop-out-analyse: Wirkungen des untersuchungsdesigns.* U.-D. Reips et al., Current Internet Science - Trends Techniques, Results. (Proceedings of the 3^{rd} German Online Research conference, Zrich)

Lozar Manfreda, K., Batagelj, Z., & Vehovar, V. (2002). Design of

web survey questionnaires: Three basic experiments. *Journal of Computer Mediated Communication*, *7*(3). (Online)

Lozar Manfreda, K., & Vehovar, V. (2002). *Survey design features influencing response rates in web surveys.* (International Conference on Improving Surveys, University of Copenhagen, Denmark)

MacElroy, B., Micucki, J., & McDowell, P. (2002). A comparison of quality in open-end responses and responses rates between web-based and paper and pencil survey modes. *International Journal of On-line Research.* (Online)

McFadden, D. L., Bemmaor, A. C., Caro, F. G., Dominitz, J., Jun, B., Lewbel, A., et al. (2005). Statistical analysis of choice experiments and surveys. *Marketing Letters*, *16*(3), 183-196.

Molenaar, N. J. (1982). Response-effects of formal characteristics of questions. In W. Dijkstra & J. Van der Zouwen (Eds.), *Response behaviour in the survey-interview* (p. 49-89). London: Academic Press.

Payne, S. (1951). *The art of asking questions*. New Jersey: Princeton University.

Ray, D., & Muller, C. (2004). Des limites de l'échelle 1-10: Caractérisation des sous-échelles utilisées par les répondants. In P. Ardilly (Ed.), *Echantillonnage et méthodes d'enquête.* Paris: Dunod.

Reips, U. D. (2002). Context effects in web surveys. In B. Batinic, U.-D. Reips, & M. Bosnjak (Eds.), *Online social sciences* (p. 69-80). Göttingen: Hogrefe & Huber Publishers.

Roster, C. A., Rogers, R. D., Albaum, G., & Klein, D. (2004). A comparison of response characteristics from web and telephone surveys. *International Journal of Market Research*, *46*(3), 359-373.

Schillewaert, N., & Meulemeester, P. (2005). Comparing response distributions of offline and online data collection methods. *International Journal of Market Research*, *47*(2), 163-178.

Schmidt, J. B., Calantone, R. J., Griffin, A., & Montoya-Weiss, M. M. (2005). Do certified mail third-wave follow-ups really boost response rates and quality? *Marketing Letters*, *16*(2), 129-141.

Schonlau, M., Fricker, R. D. J., & Elliot, M. N. (2002). *Conducting research surveys via e-mail and the web*. Santa Monica: Rand.

Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods and the logic of conversation*. Mahwah: Lawrence Erlbaum Associates.

Skitka, L. J., & Sargis, E. G. (2005). Social psychological research and the internet: the promise and the peril of a new methodological frontier. In Y. Amichai-Hamburger (Ed.), *The social net: the social psychology of the internet* (p. 1-26). Oxford: Oxford University Press.

Smith, T. W. (1995). *Little things matter: a sample of how differences in questionnaire format can affect survey responses.* (GSS Methodological Report n^{o}78. Chicago: National Opinion Research Center)

Sproull, L. (1986). Using electronic mail for data collection in organisational research. *Academy of Management Journal*, *29*, 159-169.

Stewart, D. W., & Pavlou, P. A. (2002). From consumer response to active consumer: measuring the effectiveness of interactive media. *Journal of the Academy of Marketing Science*, *30*(4), 376-396.

Tourangeau, R., Couper, M. P., & Conrad, F. (2004). Spacing, position and order: Interpretive heuristics for visual features of survey questions. *Public Opinion Quarterly*, *68*(3), 368-393.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.

Tuten, T. L., Urban, D. J., & Bosnjak, M. (2002). Internet surveys and data quality: a review. In B. Batinic, U.-D. Reips, & M. Bosnjak (Eds.), *Online social sciences* (p. 7-26). Göttingen: Hogrefe & Huber Publishers.

Yu, J., & Cooper, H. (1983). A quantitative review of research design effects on response rates to questionnaires. *Journal of Marketing Research*, *20*, 36-44.