Internet Surveys: Current & Future

Among all the existing survey modes, Internet surveys have attracted more and more attention from social science researchers. There are many advantages of Internet surveys. They provide much more efficiencies compared to traditional survey modes, such as the nearly complete elimination of paper, postage, mailout, and data entry costs, as well as the reduced time for survey implementation (e.g., Dillman, 2000). They also have the potential for dramatically reducing the close correspondence between sample size and survey costs (e.g., Dillman, 2000; Einhart, 2003). In addition, surveys via Internet make international studies easier than before (e.g., Coomber, 1997; Dillman, 2000). And they have greater appeal to certain target groups, and can get access to hard to reach populations on sensitive topics (e.g., Coomber, 1997; Dillman, 2000; Kypri & Gallagher, 2003; Smith & Leigh, 1997).

Internet surveys, however, have several disadvantages. First, Internet access and use is not as pervasive as telephone. According to Pew Internet & American Life (August, 2003), the growth of Internet usage in the United States is from just under half of American adults in 2000 to about 59% of adults at the end of 2002. These statistics show that Internet penetration in the United States has been and continues to be uneven among those in various racial and ethnic groups, those of various ages, those with different levels of education and income, as well as those in different geographical regions. The unevenly
distribution of Internet across the country brings sampling problems to survey research. Second, doing Internet surveys requires more skills than filling out paper-and-pencil questionnaires. People get used to the latter one might meet difficulties when facing computer logic, which differs from the traditional questionnaire logic. Third, some technical problems might hinder the effect of Internet surveys. For example, complicated design will make it impossible for some web users to receive the questionnaire and respond to it. Another possible case would be that certain kind of design will appear significantly dissimilar on the participant’s computer to the designer’s, due to different types of computer operation system and screen configuration. Lastly, the security and confidentiality issues associated with electronic technologies lift a concern of trust that can not be ignored (Dillman, 2000).

Currently, there are two main forms of Internet surveys: email and web surveys. Because web surveys has far more flexibility and power, and technology is becoming available that may allow questionnaires with web-like design qualities to be embedded within email messages (Dillman, 2000), in this paper, I examine web surveys specifically. Therefore, Internet surveys and web surveys are exchangeable in the following.

A Review of Literature on Internet Surveys

After doing a systematic literature review on Internet surveys, I have seen two research trends in this area. One is from the study of the generalizability to more detailed technical and methodological issues on improving response rates and data quality. The
other is from dubious and hesitant attitudes to cautiously optimistic ones among researchers.

(1) Generalizability

Sampling

For the time being, Internet surveys cannot get a probability sample as mail surveys or telephone surveys do. Therefore, many researchers doubt the inference of Internet survey results to general populations. For instance, Dillman (2002) believes that the early use of Internet surveys is limited to populations with high rates of computer use. Best, Krueger, and Hubbard (2001) also state that current Internet sampling techniques only permit diverse, not representative, samples. Braithwaite, Emery, Lusignan, and Sutton (2003) question the external validity of Internet surveys, specifically how to obtain a representative sample and adequate response rate.

While researchers are still cautious on this point, they seem to shift to appreciate the use of Internet surveys in certain populations and certain ways. Concluded from a large Internet survey done among international students, Sills and Song (2002) proposed that Internet, as a delivery method for survey research, could be very useful for studying special “technologically savvy” populations for reasons of design flexibility, geographic reach, anonymity, and minimized interviewer error, which are apparently drawbacks of conventional delivery methods. In a study aimed to measure citizen perceptions of community policing, Ballard and Prine (2002) compared Internet and mail survey responses. The results showed that methods of data collection did not significantly affect
answers given. Those who tended to respond by Internet did not differ substantially from those inclined to complete and return mail surveys. Comparably, a Germany research done by Bandilla and Bosnjak (2003) recruited a random sample of Internet users (collected via an online survey) and a random sample of the whole population (collected via a traditional written survey). They found that for high Internet penetration groups, the data gathered via the web were basically identical to those obtained in a traditional self-administered mode. Best et al. (2001) also addressed positively that the psychological mechanisms underlying survey topics did not differ between Internet users and the population in their study of people’s common political decisions in a small city.

In general, Internet surveys will continue to increase among certain population subgroups, especially among the young and individuals with higher education. This mode has the potential to become a practical and valuable resource for social scientists. However, it is too early for “general population surveys” employing a random sample of Internet users.

**Response rates**

Response rates is another important issue when considering generalizability of Internet surveys. One Internet survey research by Kypri and Gallagher (2003) obtained a high response rate of 85% among university students. This again reflects the above summary that Internet mode can be very attractive for researchers to do surveys among specific groups.

However, response rates are not always as high as 85%. Prior research has proved that prepaid, instead of promised, incentives consistently exert the largest effect on
response rates in mail surveys (Church, 1993). Does this have the same positive effect on response rates in Internet surveys?

Dillman (2000) tends to believe the power of the prepayment and expects to see prepaid incentives to increase web survey response rates. However, till now, prepaid incentives in Internet surveys do not seem to work as effectively as in traditional survey modes. One study demonstrates that mailed incentives increase the number of mail responses but do not affect the number of web response (Schonlau, Asch, & Du, 2003). A recent experiment then employed online services (e.g., PayPal) to send cash equivalents via Internet (Bosnjak & Tuten, 2003). This study indicated that prepaid incentives in Internet surveys have no advantages in terms of the willingness to participate, actual completion rates, and the share of incomplete response patterns when compared with postpaid incentives. Postpaid incentives present no advantages over no-incentives either. However, compared to no incentives, prize draws significantly increase the willingness to participate, the number of participants starting the survey, and actual completion rates, and also tend to reduce the various incomplete participation patterns. The authors concluded that prize draws are important strategies that might be beneficial to response rates in Internet surveys. Another study on Internet surveys illustrates that the size and type of incentives appear relatively unimportant (Kypri & Gallagher, 2003). However, this may be due to the absence of a no-incentive condition.

The other way to increase response rates is to decrease the burden of the survey. Generally speaking, participants perceive more burden towards web surveys because of the constraint on time, space and equipment (Crawford, Couper, & Lamias, 2001). Internet surveys are less portable; they require greater burden of resuming a partially
completed questionnaire; and the survey must be completed at a computer connected to the Internet and equipped with a web browser. Even when participants have every condition to do the survey, they may choose not to start. Even when they start, they may drop out anytime for the increased burden (if there is any). Varied web survey designs may produce unlike levels of perceived burden (Jeavons, 1998). For a scrollable or static web design, participations make their decisions based on the content of the survey; while for an interactive design, information come to respondents gradually as they answer consecutive questions, which may cause high levels of dropout. This happens less in mail surveys, because people can inspect the entire questionnaire and make judgments based on the content, length, design, and other elements of the instrument.

**Mixed-mode approach**

Because of its lack of generalizability, many researchers (e.g., Dillman, 2000; Fricker & Schonlau, 2002; Schonlau, Fricker, & Elliott, 2002; Sills & Song, 2002) recommend a mixed-mode strategy, which is to provide participants alternative ways to do the survey. Dillman (2000) suggests putting a mail address in the cover letter of Internet surveys in case some people prefer to print the questionnaire out and mail it back. Others researchers would use Internet as a supplemental means to traditional survey research, such as putting a URL in the mailed questionnaire. In a sequential multi-modal web/mail survey by Schonlau et al. (2003), respondents were first encouraged to reply via Internet and in a follow-up were also offered mail questionnaires. Interestingly, although the Internet mode was emphasized and hard copies of the questionnaire were not mailed out until the second follow-up, and even though the target population (high school
students) was presumed to be relatively computer savvy, almost two thirds of the respondents preferred responding by mail. The authors postulate that this finding is likely to be true for other populations as well.

(2) Data Quality

More and more focus has been placed on the quality of the content in the responses and observations rather than on the generalizability of findings (Sills & Song, 2002).

“No opinion” responses

In noninteractive self-administered surveys (paper-based or scrollable web surveys), there are only two design choices with respect to “no opinion” answer: either explicit or absent. In the situation that the “no opinion” option is explicit, researchers should be aware that some respondents may be tempted to satisfice by choosing the “no opinion” answer as an easy way out (Krosnick & Fabrigar, 1997). However, in interviewer-administered surveys, there is one more choice: implicit, which is believed to reduce “no opinion” responses. Uncertain responses are possible even when uncertain response options are not initially provided to a respondent; that is, when such a response is volunteered, it is accepted often after an additional attempt to elicit a more definitive answer by the interviewer.

How about in interactive self-administered surveys (computer-assisted self-interviewing or interactive web surveys)? Derouvray and Couper (2002) state that all
three options are potentially available to the Internet survey designer. In other words, interactive web surveys permit a design similar to interviewer-administered surveys. Results showed that the implicit “no response” option with a prompt if the item is not answered yielded the most complete data in their study. The finding suggests that the interactive nature of some web surveys can be used to reduce item-missing data.

**Response formats**

Since Internet surveys are more flexible and creative in designing than traditional modes, new response formats have occurred. One of them is drop-down boxes, which has a height of only one text line and hides the list of response options. However, compared to radio buttons, a response format that allows questionnaire to be looked like conventional mail surveys with which most respondents are familiar, drop-down boxes are assumed more difficult to use. Drop-down boxes require two mouse clicks to select a response option, whereas radio buttons only require one, because drop-down boxes hides the response options from the respondent until it is clicked. When using radio buttons, all response options are visible by default.

One Belgian experimental study evaluates the effect of the two response formats on data quality (Heerwegh & Loosveldt, 2002b). Results indicate that a slight preference for radio buttons might be justified. Participants need more time to complete the questionnaire using drop-down boxes and resulted in significantly more dropout. In addition, as Krosnick (1991) noted, the likelihood of satisficing increases as task difficulty boosts. Since drop-down boxes are considered to be more difficult to use, it might lead to more satisficing and thus to reduce data quality. Dillman (2000) noted
another drawback of drop-down boxes much earlier. To him, drop-down boxes can
disorient respondents when they are offered in batteries of items that share the same
response format (e.g., liked scales).

**Login procedure**

It is common practice to control web survey access to individuals in the sample
through the use of access codes such as personal identification numbers, user names, and
passwords. Currently, there are three login procedures: automatic (no access code to be keyed in), semiautomatic (one access code to be keyed in), and manual (two access codes to be keyed in).

Several studies have shown that the login procedure can influence data quality in
terms of the number of pages of the questionnaire filled out, the amount of time spent on
completing the questionnaire, and the amount of information provided to sensitive
questions. Although automatic login procedure may increase the response rate,
respondents assigned to this condition may proceed more quickly through the web
questionnaire, indicating lower degree of willingness to devote much cognitive effort
when answering the questions (Crawford et al., 2001; Heerwegh & Loosveldt, 2002a).

In a more recent Belgian study, Heerwegh and Loosveldt (2003) found that the
semiautomatic login procedure produced the highest response rate, the highest amount of
completely filled-out-questionnaire, the longest survey duration which was believed to be
an indirect measure of higher data quality, and the lowest amount of nonsubstantial
answers to some sensitive questions. The manual login procedure elicited the highest
confidentiality rating. In addition, the manual login condition not only elicited the most responses to the open-ended questions, but also obtained more information to them.

These evidences, at the same time, show that automatic login procedure should not be preferred simply because it would reduce respondent burden (Heerwegh & Loosveldt, 2003). And the benefits of the manual or the semiautomatic login procedures compensate for the higher login burden imposed on the respondent by these login procedures (Crawford et al., 2001; Heerwegh & Loosveldt, 2002a; Heerwegh & Loosveldt, 2003).

These researchers gave out reasons. First, manual login procedure provides more sense of confidentiality. They assume that the need to key in access codes could also lead to an added sense of assurance regarding the confidentiality of the responses. Based on this presumption, respondents from the automatic login procedure should feel the least confident, as they may remain unaware of the fact that their response are protected by access codes as these codes are easily overlooked and do not attract any special attention. Second, researchers expect that less motivated participants will be filtered out by more burdensome login procedures at the start of the survey. Therefore, the manual login procedure will produce the highest complete response rate and more thorough consideration of the responses given participants’ higher levels of motivation.

**Future Research Recommendations**

Based on the above literature review on Internet surveys, I conclude that, representative sample is hard to achieve in the near future, mainly for the reason of
inadequate coverage of Internet among general populations. However, research can put emphasize on how to improve response rates and Data quality. Specifically, I propose the following future research in raising response rates and improving data respectively.

First, I want to know more on the role of incentives in Internet surveys. Why do incentives seem ineffective in Internet surveys? There are several possible reasons. First, incentives via Internet are not tangible. Second, participants may have less trust on online money-transferring services. Future research should utilize more formal financial intuitions, such as well-known banks. Another distrust issue grows from the current culture of Internet surveys. Individuals may be socialized to expect a prize draw in an Internet survey. Receiving a cash equivalent may be out of expectation and therefore heighten feelings of suspect.

I also think that the size of the incentives in Bosnjak and Tuten’s experiment might be too low ($2), compared with much more burden Internet surveys would produce and much more effort participants should pay if they start the survey. I suggest adding different sizes of incentives to the future design (at least one condition should be more attractive than $2). This would enable a deeper exploration of the levels of incentive required to positively affect response rate in Internet surveys.

I also recommend trying different delivery methods for the incentives, even the combination of several methods. The result that mailed incentives had no effect on web responses in Schonlau, Asch, and Du’s study was limited, because the incentives were introduced in the second stage. I want to see how things would (or would not) change if researchers send incentives by mail at the very beginning. For example, researchers can inform potential respondents via e-mail that they are being asked to participate in a
survey and a small monetary incentive will be mailed to them soon if they send their mailing address by e-mail. Sending participants reasonable amount of incentives with intensive follow-ups may be another useful strategy.

Second, I am interested in specific web design issues, some of which I think needs to be explored deeper. I am much interested in testing the “implicit prompts” effect proposed by Derouvray and Couper in a longer survey. Will a prompt for every unanswered time have the same desired effect (consistency issue)? I also expect the less cooperativeness of the respondents because people would get tired of those reminders in a longer time. Therefore, will less motivated people be more likely to satisfice, or even drop out? This question actually addresses a very important issue in the design of Internet surveys. In respect of reducing completion burden, static, rather than more interactive, web design is preferred (less dropout). However, interactive prompts are also recommended considering the data quality. My another concern is that, because reading a lot of words on the computer is somewhat more straining than reading a paper, the static questionnaire might be too monotonous and will cause respondent weariness especially in a longer survey. Actually Dillman (2000) has advocated using drop-down boxes sparingly.

Although many literatures tend to agree that a conventional questionnaire layout is most likely to produce the best results, and the basic philosophy for Internet survey design under current conditions is to be conservative, I still expect future research could take effort to find a balancing point between interactivity and completion burden.

Another balancing issue that I think is also important as well as interesting relates to the two big concerns of Internet surveys, that is, the sometimes contradiction between raising response rates and improving data quality. When using strategies to improve data
quality, such as applying semiautomatic login procedure to exclude people who are less motivated, researchers might run the risk of oversampling the motivated people. Although some research found that using a manual login procedure did not decrease response rates, whereas it did increase the overall degree of data quality (Heerwegh & Loosveldt, 2002), still we cannot ignore the potential problem of nonrepresentativeness of the sample, unless the population of the survey is also the motivated people.

References


Schonlau, M., Fricker, R., & Elliott, M. (2002). *Conducting research surveys via e-mail and the web*. Santa Monica, CA: RAND.
