THE USE OF WEB-BASED INTERNATIONAL SURVEYS IN INFORMATION SYSTEMS RESEARCH

Michael Lang
School of Business & Economics
National University of Ireland, Galway
University Road
Galway
Ireland
E-mail: Michael.Lang@nuigalway.ie

There has been much interest of late in the use of Web-based surveys. However, the methodological issues of Web-based surveys are poorly understood. While this new medium permits low-cost wide-scale access to international populations and is capable of generating a high number of responses within a short period of time with fully automated data capture and real-time validation, the reality is that without systematic guidelines to conduct such research, its results are entirely unreliable. This paper provides an outline of the methodological issues, and suggests how they may be addressed by the technology.

Keywords: Electronic Surveys; Web-based Surveys; IS Research; Survey Design Guidelines.

1. Introduction

The prominence of survey methods within IS research has come under justifiable scrutiny in recent years. Nevertheless, if used appropriately, surveys can provide “a reasonably accurate description of real world situations from a variety of viewpoints” (Galliers 1992). However, a major hindrance with survey-based IS research is dropping response rates, now typically of the order of 10% - 15%. Where researchers find that the population of interest within their own country is small, as is often the case with many European nations (e.g. Ireland, Finland, Slovenia), a response rate of 15% of the sample can mean that the data collated is statistically inadequate. The use of appropriate inducements and an optimised research design can boost response rates, but in any event it is unlikely to rise above 35% (Falconer & Hodgett 1999) which may still be insufficient. It may therefore be necessary to extend the scope of the research to include populations from contiguous nations.

Unlike North America, it is comparatively rare in Europe for surveys to be conducted across a federation of states. To a significant extent, cost has been a constraint in regard to collating data from international samples using traditional methods. There has been dramatic growth of late in the use of Web-based surveys as researchers are lured by the potential of low cost access to geographically distributed populations. Though coverage remains a barrier in many disciplines, this is not such a problem in IS research because of higher Internet penetration. There is empirical evidence to suggest that email/Web-based surveys are cheaper than postal surveys and yield responses that are faster, more complete and more accurate (McCoy & Marks Jr. 2001; Schaefer & Dillman 1998; Klassen & Jacobs 2001; Comley 1998). Web-based surveys offer the additional advantages of real-time response validation, automated data entry, and programmable context-sensitive skip patterns.
It would therefore appear that the Web has the potential to be the saviour of survey-based research. However, the reality is that so far the Web has been greatly misused as a medium for executing surveys, and there has been blatant disregard of methodological considerations. As with all technologies, there are trade-offs between efficiency, data quantity, and quality (Alreck & Settle 1996). In order to realise the benefits that Web-based surveys offer, it is important to develop a set of guidelines that ensure acceptable response quantity and quality. Thus far, little attention has been given in the mainstream IS literature to the use of Web-based surveys as a research tool, and Web-based researchers lack the same guidance as is available to traditional pen-and-paper survey designers where there is a rich history of methodological research. Of course, much that applies to paper-based surveys also applies to Web-based surveys, but there are a number of aspects that are particular to this new mode of delivery. This paper outlines the principal methodological and technological issues involved in implementing a Web-based survey within an international context for IS research, and provides guidance on how these issues may be addressed.

2. Issues in the Design of Web-based Surveys

The principal issues discussed in this section are, in order: international aspects; sampling methods; response effects; measurement error; authentication mechanisms; and codes of conduct.

2.1 International Aspects

Web-based research almost by definition has an international aspect. Conducting surveys across two or more countries involves additional complexities that have rarely been given much attention within IS research. It is often taken for granted that using the same sampling method and delivery modes in different countries means that results are directly comparable. However, merely using the same research design is insufficient justification, and comparability must be established of its own right. As Webster (1966) points out, “a bias or error inherent in any given method may interact with differential factors in each country so that the results will not be comparable”. It is therefore necessary to select the most reliable and efficient methods for each country on an individual basis.

Definitional problems in international surveys can be problematic, particularly where the same term takes on different meanings in different national contexts. This is further compounded by the matter of linguistic equivalence. Although most linguistic irregularities can be overcome by methods such as “back translation” or “parallel blind translation”, there remains the problem whereby some concepts cannot be directly translated into other languages or meaningfully described within certain cultures (Mayer 1980). While it is true that English is an official language of more countries than any other (Linguasphere 2000), the international distribution of a survey that is only available in an English-language version can introduce bias. In actuality, there may be little additional overhead involved in translating a survey to other languages, as was the experience of Coomber (1997) who used English, French and German versions. Of course, one should also translate the pre-notification and follow-up correspondence, and in so doing respect the different cultural norms that apply to the use of salutations and other elements of writing style. Where multilingual versions of a survey are used, it may be better to avoid the use of open-ended questions as the responses to these would have to be translated back.
Caution must be taken not to treat Internet populations as being representative of international populations. Samples drawn from international newsgroups and mailing list servers (“listservs”), even where those forums are moderated and members’ identities are verified before admission, remain prone to sampling error, coverage error, non-response error, and various other biases. For example, the breakdown by country of membership of the ISWORLD mailing list is presented in Table 1 (Ives 2001). This list is open to IS academics worldwide and uses the English language only. It is clear that members are predominantly from North America and other first-world English-speaking nations.

<table>
<thead>
<tr>
<th>USA</th>
<th>1,483</th>
<th>Israel</th>
<th>23</th>
<th>Slovenia</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>214</td>
<td>Norway</td>
<td>23</td>
<td>Argentina</td>
<td>4</td>
</tr>
<tr>
<td>Great Britain</td>
<td>161</td>
<td>Korea</td>
<td>19</td>
<td>Thailand</td>
<td>4</td>
</tr>
<tr>
<td>Canada</td>
<td>123</td>
<td>Italy</td>
<td>18</td>
<td>Hungary</td>
<td>3</td>
</tr>
<tr>
<td>Germany</td>
<td>110</td>
<td>Portugal</td>
<td>18</td>
<td>Mexico</td>
<td>3</td>
</tr>
<tr>
<td>New Zealand</td>
<td>51</td>
<td>Singapore</td>
<td>18</td>
<td>Peru</td>
<td>3</td>
</tr>
<tr>
<td>Netherlands</td>
<td>43</td>
<td>Taiwan</td>
<td>17</td>
<td>Russian Federation</td>
<td>3</td>
</tr>
<tr>
<td>Switzerland</td>
<td>37</td>
<td>Denmark</td>
<td>15</td>
<td>Colombia</td>
<td>2</td>
</tr>
<tr>
<td>South Africa</td>
<td>36</td>
<td>Japan</td>
<td>15</td>
<td>Kenya</td>
<td>2</td>
</tr>
<tr>
<td>Sweden</td>
<td>36</td>
<td>Brazil</td>
<td>14</td>
<td>Kuwait</td>
<td>2</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>34</td>
<td>Spain</td>
<td>14</td>
<td>Romania</td>
<td>2</td>
</tr>
<tr>
<td>Finland</td>
<td>32</td>
<td>India</td>
<td>11</td>
<td>Saudi Arabia</td>
<td>2</td>
</tr>
<tr>
<td>Ireland</td>
<td>30</td>
<td>Belgium</td>
<td>9</td>
<td>Venezuela</td>
<td>2</td>
</tr>
<tr>
<td>Austria</td>
<td>23</td>
<td>Greece</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>23</td>
<td>Malaysia</td>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Membership of the ISWorld Mailing List.

Temporal equivalence is another consideration in international surveys. If there are time lags between countries, - as may happen between those that are more advanced and those that are less advanced, or where certain cultures are more willing to embrace new technology than others, - the results could be badly skewed. The form of the survey instrument is also important. Some cultures are much more patient than others as regards questionnaire length or responses to open-ended questions, and response styles can also vary significantly (Mayer 1980). A researcher should be aware of these issues and should use appropriate statistical tests to detect and compensate for any biases therefore arising.

### 2.2 Sampling

In business surveys, accurate sampling is vital; if done improperly, the entire research study is crippled. A survey researcher who aspires to produce findings that are scientifically reliable must carefully define the study population, engage random sampling techniques, and then strive to attain as high a response rate as possible (Dillman & Bowker 2001). In reality, IS survey researchers often have to make undesirable but pragmatic compromises for a variety of reasons.
A frequent criticism of IS researchers is that they tend to use “convenience” samples such as mailing lists rather than probabilistic sampling techniques (Kraemer & Dutton 1991; Pinsonneault & Kraemer 1993). However, as Lucas (1991) points out, a major reason why many IS research studies omit to use probabilistic sampling techniques is not by preference but simply because of practical difficulties in obtaining random samples. For example, participants in a survey are often selected on the basis of an expressed willingness to cooperate, and are therefore not random.

The definition of accurate sampling frames for surveys in IS research is notoriously difficult. Very often, no suitably detailed register exists of the population of interest. It can be particularly hard to acquire contact details of named individuals that perform specific functions within organisations. Industry databases, such as those provided by government agencies or third-party companies, often provide only very general or ambiguous descriptions of an organisation’s products and services, usually list just the names of company directors or general contacts, rarely provide e-mail addresses, and are often riddled with inaccuracies.

2.2.1 Sample Size

The precision of a survey depends upon the sample size as a proportion of the overall population. It is not always possible to accurately determine the size of the population, especially with Web-based research. Hill (1998) therefore poses the question: how large should a sample be in Internet survey research? It is important to be methodically discriminate in gathering data and not fall foul of “vacuum-cleaner-like collection”. Roscoe (1975) claims that sample sizes of less than 30 or larger than 500 can seldom be justified, and Alreck & Settle (1996) maintain that about 10% of the overall population is adequate. Using the formula in Dillman (2000, p. 206), it is seen that if one were to accept a confidence level of 95% and sampling error of ±3%, a sample size of 1067 is adequate for an infinite population. However, given that response rates to business surveys are typically as low as 10%-15%, it may be argued that sample size should be increased so that sufficient data is captured for analysis.

2.2.2 Coverage

It is critical that the sampling frame that is used constitutes a representative subset of the overall population of interest. If this is not so, it matters not how large the sample is for it cannot be said to be truly representative.

For Web-based surveys of the general public, coverage error is likely to be high (Dillman & Bowker 2001). Respondents to Web-based surveys are typically substantially younger, better educated, more computer-oriented, more comfortable with and receptive to Web-based surveys, and more affluent than society as a whole (Zhang 1999; Taylor 2000; Batagelj & Vehovar 1998). When nearly all members of a population have Web access, coverage becomes less of a problem (Dillman & Bowker 2001). While it is still the case that many organisations do not facilitate company-wide access to the Internet, and that workers in jobs not commonly involving computer-based work may have limited Internet access, it seems fair to assume that for research within IS and other disciplines where Internet penetration is high, the risk of coverage error is significantly reduced.

2.2.3 Sampling Techniques for Web-based Surveys

With Web-based surveys, it is difficult if not impossible to use probabilistic sampling techniques. Lazar & Preece (1999) recommend that where it is impossible to get a strictly
random sample, the goal should be to get as diverse a response as possible. It is then up to the researcher to convincingly argue that his sample does not differ in any meaningful way from a random sample.

E-mail addresses of named individuals within organisations are not readily available, because many e-mail users are understandably reluctant to disclose their contact details on account of security and privacy issues. Vehovar et al (2000) report that they could find only 21% of e-mail addresses for the organisations within their sampling frame, and these were mostly of an impersonal form such as info@company.com. There is no comprehensive online directory of e-mail addresses, and because e-mail addresses change frequently, they are much more difficult to locate than mailing addresses or telephone numbers. Sampling techniques using e-mail addresses as a means of contact are therefore not random, because they are based on availability.

If it is not possible to acquire individual e-mail addresses, an alternative or complementary strategy that is often used is to send a broadcast message to dedicated, specialised newsgroups or listservs around which potential participants may be expected to congregate. However, numerous problems attach to the use of newsgroups and listservs as sampling frames. Firstly, studies reveal that response rates are relatively low because the communication is broadcast to a large anonymous group and is therefore impersonal (Zhang 1999; Schaefer & Dillman 1998). Secondly, it can be very difficult to authenticate who a respondent is or where they come from. Thirdly, this method of solicitation breaches the fundamental rule that a researcher should always define and go after his sample rather than let the sample define itself and come to him (Eaton 1997). The problems of self-selection are amplified where invitations to participate are broadcast to multiple online forums, and the survey is openly available to anonymous respondents. In such situations, inference to any larger population is scientifically unjustified (Dillman & Bowker 2001). This is not to say that listservs and newsgroups cannot be used as sampling frames. Forums that are moderated and apply strict admission criteria for members may be appropriate in certain circumstances. However, the researcher would need to collect personal data about the respondents and then compare this aggregate data with that of the overall population to assure that they represent a plausible microcosm of the universe. Mechanisms would also need to be put in place to prevent duplicate submissions and impersonation – a possible way of doing this would be to ask the respondent to register in advance, whereby he is randomly allocated an ID and password which are sent to a nominated e-mail address.

The membership listings of professional associations can also form appropriate sampling frames for Web-based surveys, particularly those that have large membership as a proportion of the overall population, and where all or most of the members have access to the Internet. However, most professional associations are organised on a national basis, and only a few have international chapters. It is important to heed the point by Mayer (1980) that where international research is being conducted, the sampling frames for each country should be selected independently of one another. For example, in one country the membership of a particular association may be 70% of the overall population, but in another country the membership of the same (or comparable) association may be just 10% of the population. It is therefore necessary to refer to national census data of the industry sectors being studied to form an impression of how representative the membership of an association is of the wider population. It is also important to consult the admission procedures of professional associations as these can cause potential bias or inconsistency; for example, some associations only permit company membership as opposed to individual membership.
Direct marketing databases compiled by reputable vendors can be useful for acquiring names, job titles, and postal and e-mail addresses for individuals within organisations. However, these data sources are costly, particularly for larger sample sizes, with individual records typically being priced at €0.50 - €1.20 each. In practice, it may be best to use a number of different sampling techniques in conjunction with each other. Even then, it may be very difficult to be sure that the sample is indeed representative of the audience of interest. In light of this, Farmer (1998) refers to the concept of “sifting”, whereby a universe of potential respondents can be over-sampled, and only those participants who accurately represent the population the researcher aims to study are included in the data analysis.

2.3 Response Effects in Web-based Surveys

2.3.1 Response Rates

Non-response error may arise because those members of the sample who did not respond would have responded substantively differently from those who did respond. Although non-response error is not the same as response rate, a higher response rate reduces the likelihood of non-response error. Few IS surveys have achieved response rates of 50% or above (Kraemer & Dutton 1991; Pinsonneault & Kraemer 1993), a level that according to Babbie (1973, p. 165) is barely adequate in social science research. Realistically, it seems that response rates in IS research are most likely to be in the range 10% - 35% (Falconer & Hodgett 1999).

The use of incentives, the length and topic of the questionnaire, pre-notification and follow-up procedures, endorsement by a university or professional body, advice of cut-off dates, visual design and questionnaire formatting, the delivery medium used, personalisation, confidentiality, the reputation of the researcher, the proximity of the researcher to the participants, and the quality of the sampling frame are all factors that can significantly affect response rates. This means that straightforward comparisons of one study against another are difficult. This difficulty is compounded by the fact that different studies have tended to compute response rates inconsistently. For Web-based surveys, response rates are sometimes inflated because undeliverable “bounced” e-mail may be excluded in computing them, unlike with traditional postal mail where the researcher may not be notified of undelivered mail. Even with e-mail notification, there is no exact way of determining how many of the intended participants received or read the message because of problems such as deactivated or dormant accounts, server errors, automated filtering or deletion, and the presence of gatekeepers. If the invitation to participate is broadcast on a newsgroup, bulletin board or listserv, then it is almost impossible to accurately estimate the sample size and thus the response rate, particularly if the posting appears on multiple related forums with overlaps between subscribers.

Most mixed-mode studies find that e-mail/Web-based surveys generate much faster responses than traditional media. This ought not be surprising because e-mail communication is almost instantaneous and in most cases the difference in response times between e-mail and postal mail is about 6-10 days (Cobanoglu et al. 2001), which could be attributed purely to the logistics of the postal service. As regards response rates, there is much less agreement. They can be as low as 8% (Smith 1997) or as high as 90% (Zhang 1999). Though Web-based surveys are quite distinct from e-mail surveys, disk-by-mail, and other computer-assisted interviewing (CAI) techniques, they share some common aspects. As there has been very little research thus far that compares Web-based surveys with other modes, one must look to comparisons of postal mail versus e-mail and other CAI techniques in order to gain insights.
that may guide the design of Web-based surveys. Table 2 provides a summary of research into response rates to mixed-mode surveys. However, comparisons between studies are difficult because (1) different studies manipulate different variables, (2) for some studies, respondents were forced to use a certain mode, but in others were given an opportunity to choose from two or more possible modes e.g. Web or post, (3) some of the studies are internally inconsistent e.g. using follow-up for Web-based distribution but not for postal distribution, (4) mixed-mode response rates are skewed because of coverage error, and (5) the examples are drawn from a variety of disciplines. Nevertheless, what is clear is that response rates to e-mail and Web-based surveys tend to be lower than for traditional modes such as postal mail, and that where respondents are given the choice of responding via e-mail/Web as opposed to traditional mail, they will generally opt for traditional mail. It would appear that most people are still more comfortable with traditional modes. Interestingly, this is true even where the sampling frame is wholly drawn from Internet-based populations (Mehta & Sivadas 1995; Mavis & Brocato 1998). A number of possible explanatory factors are discussed in the next section.

<table>
<thead>
<tr>
<th>Study</th>
<th>Description</th>
<th>Response Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Mehta &amp; Sivadas 1995)</td>
<td>Sample drawn from popular Internet newsgroups. E-mail addresses were harvested, institutional details and postal addresses found within signatures.</td>
<td>Postal mail (USA) 83%; E-mail (USA) 63%; E-mail (International) 64% (with pre-notification) Postal mail (USA) 45%; E-mail (USA) 40%; (no pre-notification)</td>
</tr>
<tr>
<td>(Schaefer &amp; Dillman 1998)</td>
<td>Survey of university faculty.</td>
<td>Postal mail 57.5%; E-mail 58%</td>
</tr>
<tr>
<td>(Mavis &amp; Brocato 1998)</td>
<td>Mixed-mode survey of subscribers to DR-ED medical education listserv. Subscribers mostly from North America.</td>
<td>Postal mail 77%; E-mail 66%</td>
</tr>
<tr>
<td>(Vehovar et al. 2000)</td>
<td>Mixed-mode survey of e-commerce usage amongst SME’s in Slovenia. Only interested in respondents who have Internet access.</td>
<td>Phone 52%; Postal mail 39%; Fax 32%; Web 26% (adjusted)</td>
</tr>
<tr>
<td>(Klassen &amp; Jacobs 2001)</td>
<td>Mixed-mode survey of manufacturing and service sectors in Canada.</td>
<td>Postal mail 23%; Fax 20%; Disk-by-mail 21%; E-mail 14%</td>
</tr>
<tr>
<td>(Cobanoglu et al. 2001)</td>
<td>Mixed-mode survey of 300 hospitality professors (USA)</td>
<td>Postal mail 26.3%; Fax 17% E-mail/Web-based 44.2% (adjusted)</td>
</tr>
</tbody>
</table>

Table 2: Response rates to mixed-mode surveys.

2.3.2 Critical Factors that Affect Response Rates to Web-based Surveys

Most of the benefits cited of Web-based surveys are on the side of the researcher. If the Web is to become an effective delivery mode for survey research, then it is important that it should be as least as convenient and as comfortable for participants as traditional modes such as pen-and-paper. Foremost amongst participants’ concerns are (1) confidentiality/anonymity, (2) ease-of-use and (3) time-efficiency. Failure to consider these aspects is likely to have a negative impact on response rates.

Privacy concerns can have substantial negative impacts upon response rates for online research. Because e-mail surveys automatically identify the respondent, anonymity is
compromised and response rates may be affected (Lazar & Preece 1999). With Web-based surveys, this is less of an issue. For very sensitive research topics, Coomber (1997) suggests that participants could use public- or open-access Web terminals, or just print the questionnaire and return it by post. This latter option negates one of the benefits of Web-based surveys, that of automated data entry. The major disadvantages of anonymity are that it is impossible to authenticate participants, to eliminate multiple responses from the same participant, or to follow-up accurately with non-respondents. A compromise is to allocate random identifiers/passwords to invited participants. This way, it is possible to authenticate legitimate respondents, but it is not possible to directly identify individuals from the data set. However, Stanton & Rogelberg (2001) warn that “unless participants perceive the processes involved in allocating passwords or identifiers as random, they may have substantial concerns about anonymity, which may, in turn, modify response patterns”.

Ease-of-use is another major issue. Poor survey layout may detract from response rate; hence usability testing should be done (Lazar & Preece 1999). A survey need not have an ornate interface; indeed a simple interface may be better. Dillman et al (1998) report a completion rate of 93.1% for the “plain” version of a Web survey, but 82.1% for the “fancy” version. Based on observations of subjects as they attempted to fill out Web surveys, Dillman & Bowker (2001) suggest that two sources of significant frustration that can lead to drop-out are a basic lack of computer knowledge and poor questionnaire design. Potential problems include the presence of excessive navigation controls, not knowing how close one is to the end of the questionnaire, or not knowing what to do with drop-down lists and other interface objects.

Lack of time is another major cause of non-response by managers to business surveys (Falconer & Hodgett 1999). Based on his experience in designing Web-based surveys, Farmer (1998) remarks that any questionnaire which takes more than 15 minutes to complete has a very high probability of non-response error. A key question that must be asked is: are Web-based surveys time-efficient for participants? In one of the earliest Internet surveys of the general public, Bertot & McClure (1996) reveal that although 77.3% of respondents found their on-line questionnaire easy to complete and 86.1% found it intuitive, a substantial number (40%) felt that it was not time-efficient as regards the effort required to complete it. In this regard, Batagelj & Vehovar (1998) report that graphical browsers are substantially faster than text-only browsers in the completion of Web-based surveys, and that single-screen questionnaires are faster than multiple-screen questionnaires. As a consequence, multiple-screen questionnaires had a slightly higher drop-out rate, but this should probably be seen as a tolerable trade-off given that multiple-screen questionnaires have been elsewhere seen to increase item completion rates (Smith 1997). Intuitively, one would suspect that Web-based surveys are not as good as pen-and-paper for long open-ended responses, as most people cannot type as fast as they can handwrite. This is therefore a significant limitation of Web-based surveys.

2.3.3 Incentives to Respond to Web-based Surveys

According to Dillman (2000), survey response rates are tied to the concept of social exchange theory, a critical element of which is the use of incentives. Incentives that have traditionally been used in survey research include small monetary amounts, lotteries, and final reports. With Web-based surveys, the use of material incentives such as cash prizes, vouchers, holiday competitions or physical goods can affect response quality by encouraging multiple submissions and rubbish data (Comley 2001). Indeed, it has been found that offering financial incentives in Web surveys does not necessarily reduce non-response error (O'Neil &
Penrod 2001). If material incentives are offered, then they should probably be related to the research topic – such as a subscription to a periodical or a chance to win a subsidised trip to a conference.

Perhaps the single best incentive that an academic researcher can offer is a free report of the findings in a format that is interesting and useful for respondents. Respondents should also be advised that their participation in the study is an opportunity for them to provide feedback and opinions that can help academics to produce better, more relevant research and teaching.

2.4 Measurement Error

Measurement error arises because of problems within the instrument itself, such as poorly worded questions, aspects of the respondent’s behaviour, or mode effects. In the design of Web-based surveys, particular care must be taken to trap and eliminate measurement errors. What the designer of a Web survey sees on his screen may be very different to what the respondent sees, because of various problems such as differences between browsers and operating systems, colour depths and palettes, screen resolutions, visual distances, text wrapping, device characteristics, font sizes, the use of non-standard fonts, character sets, plug-ins, image file formats, and support for languages such as Java, Javascript, HTML extensions, and Cascading Style Sheets (CSS).

Web-based technologies are continuously changing at a rapid pace, and people are adopting these changes at different rates (e.g. installing the newest version of a plug-in or upgrading a browser), so Web-based researchers must find ways to mitigate the affects of the uncertainty introduced by these variations (Dillman & Bowker 2001). It is therefore critically important that Web-based surveys be adequately pilot-tested because they may function differently on different platforms. Consideration must also be taken of users’ level of expertise (Stanton & Rogelberg 2001), and it may be necessary to include additional special instructions such as, for example, how to use drop-down lists or radio buttons. An advantage of Web-based delivery is that it is possible during pilot-testing to record timings and to detect potential problems with all or part of a questionnaire.

Ideally, the survey should be hosted on a server over which the researcher has control, as opposed to those freely available from service providers (e.g. geocities) where banner and pop-up advertisements are likely to be highly intrusive. The Web server should have adequate bandwidth and provide quick response times. It should also be tested to ensure that it can withstand the heavy traffic loads that typify the first few days of an on-line survey.

Where the survey is made available in multiple languages, caution must be taken with the translations in order to assure comparability and so as to avoid measurement error arising out of different interpretations.

2.5 Authentication and Treatment of Multiple Responses

With traditional paper-based survey, the passing-on of photocopies by solicited respondents to unsolicited respondents has not been much of a problem. However, for Web-based surveys where the invitation to participate arrives by e-mail, it is easy for respondents to forward it on to others. Therefore, Web-based survey researchers should carefully define their samples, and then engage some mechanism to effectively screen unintended participants. These mechanisms should be sufficiently secure as to lock out unwanted respondents and would-be hackers, yet not so complex as to dissuade legitimate respondents. Common techniques
include the allocation of respondent identifiers, the use of password-protected directories, and authentication based on network address.

Zhang (1999) used a unique randomly generated 11-character ID which was authenticated against a database. 92% of participants were successful entering their ID at the first attempt, and nobody needed more than three attempts. Schleyer & Forrest (2000) reported problems with their ID implementation because of confusion between the letter “O” and the digit 0, and between the letter “I” and the number 1. Therefore, it is best to avoid using these four characters. Furthermore, if letters are used in an ID or password, they should be case insensitive.

Another problem that can arise is where the same respondent either deliberately or inadvertently submits multiple responses. It must then be decided whether to discard all or some of the responses. A suitable policy is to record all responses with a timestamp and to retain only the most recent response. Once a participant has completed the survey, his ID should be disabled to forbid re-entry.

2.6 Standards of Conduct and Etiquette

As Web-based surveys are a relatively new mode, researchers are as yet unclear on how to solicit potential participants without intruding upon their privacy or breaching Internet etiquette. Sending e-mails to potential participants who have not agreed beforehand to participate may be unacceptable to many people. By giving pre-notification of an upcoming survey and giving participants the opportunity to decline in advance, response rates can be greatly improved and the risk of negative backlash is alleviated (Witmer et al. 1999; Mehta & Sivadas 1995).

Unsolicited e-mail received from a stranger is often treated with some suspicion, particularly those from free e-mail addresses (e.g. yahoo, hotmail). It is important that a researcher should permit a potential participant the opportunity to verify his identity (Smith 1997), - perhaps by just listing full contact details and Web page details in an e-mail signature.

The researcher should also state his policy on confidentiality and provide a warranty that data will not be sold to third parties. Compliance with codes of practice for Internet research (e.g. ESOMAR 1999) and with legislation (e.g. EU Directive on the Protection of Personal Data 95/46/EC) should be made explicit.

3. Conclusions

Used appropriately, Web-based surveys can produce valid results if conducting research on the use of computer products or the Internet itself, within academic institutions or networked organisations, with the early adopters of new technology, or for some kinds of youth research.

However, thus far, the Web has been greatly abused as a medium for survey research, and there seems to be little understanding of the complex methodological issues involved. Many people seem to be under the rather false impression that Web-based surveys are inexpensive, and can quickly and easily generate high numbers of good quality responses. As with all other modes of survey research, good quality responses cannot be expected except where careful attention has been given to sampling procedures and other critical aspects of research
design. Notably, the European Society for Opinion & Marketing Research (ESOMAR) warn that:

“Any Internet surveys which fall seriously below the high standards promoted by ESOMAR and other leading professional bodies will make it more difficult to use the medium for genuine research and could seriously damage the credibility of such research, as well as being an abuse of the goodwill of Internet users generally”

(ESOMAR 1999)

The design of Web-based surveys requires that a researcher not only have a sound knowledge of the methodological aspects, but also has the technical knowledge necessary to implement those aspects. IS researchers are in a privileged position because many of them have an understanding of research methods and of software design. For those with the requisite knowledge, the availability of free open source software (e.g. Linux, Apache Web server, MySQL database, PHP) can greatly reduce the setup costs. However, the fact remains that even within the domain of IS, coverage is often low, as exemplified in Table 1. This means that for the foreseeable future, Web-based surveys should properly be used as part of a mixed-mode research design, alongside other media such as postal mail, telephone, and e-mail.

4. References


