December 2003

Experiences in e-survey development for IS research: Lessons from the use of automated control tools

Eusebio Scornavacca
Victoria University of Wellington

Joao Becker
Federal University of Rio Grande do Sul

Karen Dowling
Victoria University of Wellington

Experiences in e-survey development for IS research: Lessons from the use of automated control tools

Eusebio Scornavacca[a], Joao Luiz Becker[b] and Stuart J. Barnes[c]

[a] School of Information Management, Victoria University of Wellington
P.O Box 600, Wellington, New Zealand
eusebio.scornavacca@vuw.ac.nz

[b] School of Management, Federal University of Rio Grande do Sul
Rua Washington Luis, 855, Porto Alegre - RS – Brazil, CEP 90010-460
jlbecker@ea.ufrgs.br

[c] School of Information Management, Victoria University of Wellington
P.O Box 600, Wellington, New Zealand
stuart.barnes@vuw.ac.nz

Abstract

The application of the Internet to traditional business and administrative activities has introduced considerable digitisation and process automation. One key area of development from a research perspective is the use of electronic (e-) surveys, based on Internet technology. E-surveys can bring many benefits from a research perspective, including extremely low marginal costs, automation of processes, and the ability to collect and manage very large samples. However, experience in the use of e-surveys has found considerable challenges in achieving a quality sample frame and response rates. This paper explores the development of an e-survey tool for assessing information needs of growth enterprises in Brazil. A key learning from the use of the survey is the use of control methods for both improving the response rate, and, as a consequence, the sample frame.

Keywords

E-survey; control methods; survey automation

Introduction

Despite misgivings and experiences with the much publicised dot-com crash, e-commerce maintains explosive growth (ITU 2001). Predictions suggest e-commerce will surpass US$1 trillion by 2005, up from US$350 million recorded in 2000 (Mosquera 2001). The benefits presented by the new paradigms of the Internet medium are many and varied; customers and companies alike may benefit from the new ways of exchanging information, communicating and conducting trade (Hoffman & Novak 1996).

The application of Internet technology to traditional business and administrative transactions has seen many types being transferred from paper (and other traditional modes) to electronic forms. From a practical research perspective, one major area of development to emerge from the Internet era is the administration of electronic (e-) surveys (Simsek 1999, Cook, Heath & Thompson 2000, Taylor 2000, Churchill 2001, Schubert & Hepworth 2000; DeVaney 2001,

Ranchhod & Zhou 2001 Boyer et al. 2002, Liaw 2002, Shannon & Bradshaw 2002). The Web is fast becoming an important data collection medium. However, the introduction of Web survey techniques in preference to other modes of data collection, such as telephone and mail, raises some important methodological issues. Each method has distinctive discrepancies. For Web surveys, the low cost and automated procedures are typically attractive, and properly designed and programmed surveys have proven to have high validity. However, key challenges exist in administering Web surveys, one of the main ones being the typically low response rates associated with this mode of data collection.

This article provides insight into specific issues associated with Web surveys, and control tool methods that can be used to enhance them. In particular, it explores the conception and implementation of a semi-automatic Internet survey system. One key development in the delivery of the electronic survey examined in this research is the use of control tools for follow-up communications with the survey sample. The particular combination of control tools in this study provides a highly successful mode of achieving enhanced response rates. This is considered one of the core contributions of the paper. The article is organized as follows. The next section provides a brief overview of the existing research on electronic surveys and survey methodology. Section three provides a summary of the development of the system, and section four examines the results of the e-survey. Finally, section five provides a summary and conclusions regarding the use of e-survey instruments.

The Literature on Survey and E-Survey Methods

There is a rich and long-standing literature in the IS and management fields on survey methods (Pinsonneault & Kraemer 1993, Mattar 1994, Hoppen, Lapointe & Moreau 1996, Babbie 1999, Churchill 2001). The methodology for traditional survey modes - such as face-to-face, telephone and mail - is well elaborated in the literature (Dillman 1978, 2000, Groves, Biemer, Lyberg, Massey, Nicholls & Waksberg 1988), as are the methods for statistical inference and probability sampling (Kish 1965). As a consequence, sample surveys based on these standards have been successfully conducted in countless empirical projects providing statistically valid and reliable data. There are, of course, possible problems in such surveys, ranging from frame bias and non-response errors to measurement issues, but a carefully designed and comprehensively tested instrument, based on a quality sampling frame, probability sample, and sufficient response, provides the basis for a valid and reliable measurement of social phenomena (Vehovar, Lozar Manfreda & Batagelj 2001). Notwithstanding, other issues, such as departure from methodological standards, lack of resources, or breach of professional ethics, can lead to deceptive or inaccurate research findings.

The use of e-survey methods, which have grown extraordinarily in popularity, brings a different set of problems (Boyer, Olsen, Calantone & Jackson 2002, Cook, Heath & Thomson 2000, Epstein, Klinkenberg, Wiley & McKinley 2001). With standard survey modes, the quality of the survey estimates seems to depend only on whether sufficient resources are dedicated to the measurement process. However, with Web-based e-surveys this is not yet the case, and thus the e-survey is often held to lack validity (Vehovar et al. 2001, Shannon & Bradshaw 2002). The e-survey, as we define it here, involves a self-administered questionnaire (without the presence of an interviewer) delivered via a standard Web browser. Responses are transferred electronically to a server through a network, typically the Internet. The questionnaire is typically based on HTML, and may use graphics, audio and hypertext

links to provide a richer survey experience. Typically, respondents are provided with a survey invitation and Web address via e-mail. Other types of e-survey solely conducted via e-mail are also possible, but are not the focus of this paper.

E-surveys have numerous benefits over traditional methods (Shannon & Bradshaw 2002). One of the chief advantages is the low cost of administration, due to the "peopleless, paperless" mode of data collection (Clayton & Werking 1998). There are, for example, no costs associated with paper, printing, envelopes, stamps, and related administrative work, or for data entry and editing. As compared with telephone or face-to-face interviews, the role of the interviewer is also redundant. The key start-up costs are associated with equipment, Web page design, and usability testing (Vehovar et al. 2001). Notwithstanding, the cost per interview falls rapidly with the number of responses (Clayton & Werking 1998, Boyer et al. 2002), and also enables extremely large samples, which may help to reduce sampling variance. Critics of statistical significance testing often point to the fact that statistical significance is a function of sample size and that most results will become statistically significant if a large enough sample is chosen (Kirk 1996, Nix & Barnette 1998). On the other hand, it is not necessary that representativeness increase monotonically with increasing response rate (Cook, Heath & Thompson 2000). Other benefits include the much shorter times involved in administering e-surveys, reduction in potential mistakes due to interpretation of the respondent's handwriting, and e-mail follow-ups (Simsek 1999).

E-surveys provide some important challenges for researchers. A critical problem concerns the quality of the sampling frame. If e-mail lists are used to create an initial sample, there are inherent problems. Usually these lists lack of structure, and a large proportion of the potential sample may not have access to the Internet (Bradley 1999, Simsek 1999). According to Bradley (1999), this is one of the greatest challenges on the development of e-survey techniques and an important area for the development of sampling techniques in order to minimize any kind of bias that might exist. As Shannon & Bradshaw (2002) point out, electronic surveys are not for everyone. Ranchhod & Zhou (2001) in a comparative study between mail and electronic surveys found that people who prefer to answer e-surveys are usually the ones who have a better understanding of the technology and use extensively the Internet as a communication medium. However, in some cases, where the target population is 'net enabled'', e-surveys prove a compelling mode of data collection. Such situations might encompass Internet related products or software, or target population such as students or middle-classes.

Developing an E-Survey for Assessing Information Needs of Brazilian Enterprises

The survey described here was part of a programme of research into information intermediation through Internet, focusing on the informational needs of Southern-Brazilian enterprises when expanding their businesses. This section describes the development and implementation of the instrument.

Sample selection

The e-mail list used as a sampling frame was taken from the database of SEBRAE-RS[i]. This database is very representative of the population, with a high degree of structure for sample selection. The sampling method used in the research was non-probabilistic, based on

convenience. The sample was chosen with the following criteria: to be a company located in the State of Rio Grande do Sul; to be registered at SEBRAE-RS; and to posses an e-mail address. As a result, a file was received from SEBRAE-RS containing 9,730 companies' e-mail addresses. After sending the first e-mail inviting participation in the survey, it could be detected, through return, that 1,767 companies had outdated e-mail addresses, leaving a group of 7,963 companies with valid e-mails.

Figure 1. Initial page of the questionnaire.

The survey questionnaire

Potential respondents were informed about the survey by e-mail, and invited to take part. The URL provided to the potential respondents led them to the questionnaire home page (see Figure 1), which provided information on the research objectives, partners involved, confidentiality, and about the issue of spontaneity in response. The respondents selected a button to move on to the main questionnaire.

The data collection instrument used in the survey was developed by a group of IT specialists involved with the research programme. The questionnaire focused on the information needs of growth enterprises. Specifically, the questionnaire opened with the following statement:

> "Before expanding a business, it is wise to obtain some information to support this important decision. Imagine that your company thinking about expanding its businesses and that you will be the person entrusted making this strategic decision. What information do you consider important to have in hand to support your decision-making process? Please, describe it in the fields below using keywords."

Figure 2. Keyword dictionary example.

Subsequently, the respondent is provided with a variety of keywords, and could select a hypertext link to read an explanation of each keyword (Figure 2). An example of a completed questionnaire is also provided. The decision to request the respondents to use keywords instead of open text to express their needs was motivated by three considerations:

- the evaluation of the possible use of an automatic information search system guided by keywords;

- the verification of the capacity for respondents' synthesis; and,

- the facilitation of the analysis process, since the number of respondents was close to one thousand.

At the end of this stage, the respondent is asked to confirm whether they wish to move to the next section, where some demographic data is collected. For this part of the questionnaire, the system consulted the database and identified the respondent's company data, which helped to automate the survey process. The use of ASP (active server pages) also allowed the capture of information - without the respondent informing - related to the operating system, manufacturer and version of the browser.

At the end of the survey, if the respondent pressed "continue" and the data was incomplete, a message appeared informing the respondent of the problem and associated field, and asked for confirmation that the respondent wished to continue. Subsequently, the respondents were asked if they wished to receive a brief report with the research results. Concluding the process, a message of gratitude was shown.

The next section describes some considerations about the construction and testing of the data collection system.

Piloting the survey

The pilot phase had three objectives: to evaluate the authenticity of the questions, to evaluate the graphical interface, and to validate the instrument. The pre-test sample consisted of 50 Master's course students from PPGA/UFRGS, 33 of which completed the survey. To verify the clarity and the understanding of the instrument, ten respondents were observed during the

questionnaire completion process, and follow-up interviews were conducted. Once collected the data from the pilot survey was analyzed in order to assess the likelihood of errors or imprecision in the composition of questions, complexity of the questions, inclusion of unnecessary questions, and exhaustion when filling out the questionnaire (Gil 1994).
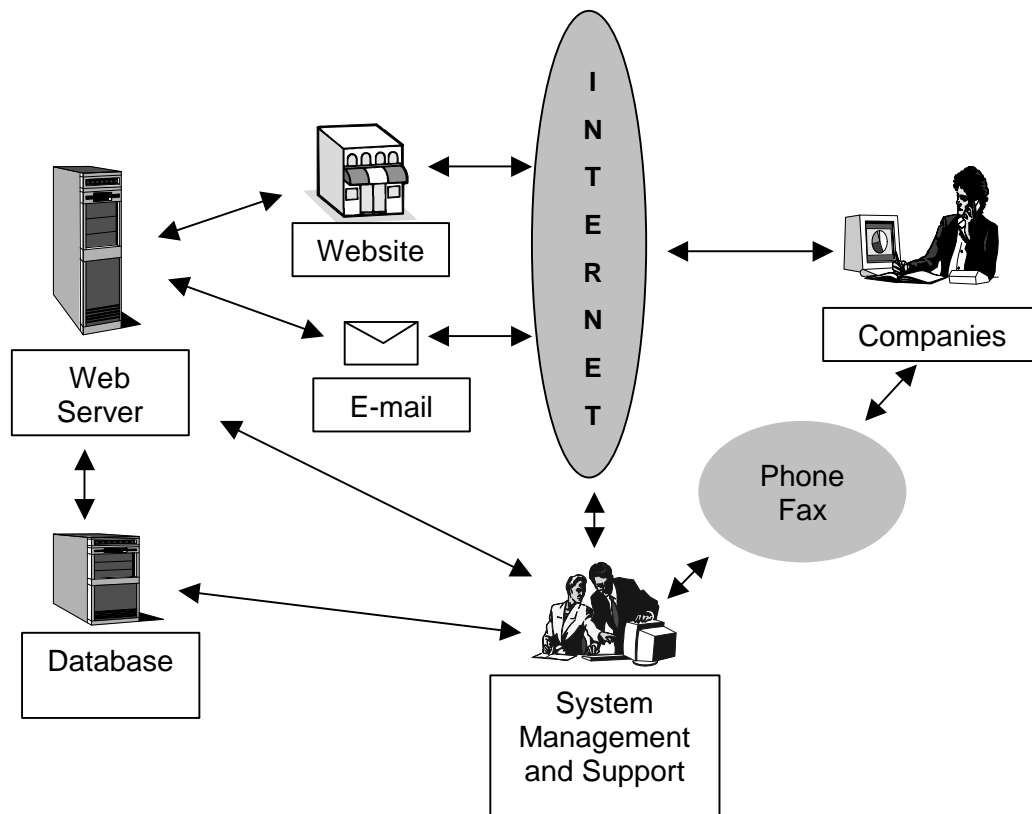

Figure 3. Data collection system.

Data collection system

One of the greatest challenges of this work was the development of the system to support data collection. Figure 3 illustrates the design of the final system. The core of the e-survey system consisted of the ASPs, Web pages and database hosted on the web server of EA/UFRGS. The database, in MS Access format, contained the information on the sample (identification code, electronic address, name, and so on) and also recorded the new data. The WebPages that the respondents accessed were developed in HTML and ASP. The scripts, developed in ASP, created an interface between the system administrator and the system (e.g., to send e-mails, control the process, and so on). Also, a support service was also put in place, by which the respondents could contact the researchers through e-mail, telephone or fax.

Before doing the pre-test (developed concomitantly with the instrument), some other tests (storing data, access and sending of e-mails) were used to check the system's performance. In the pre-test, the need to generate a program to lock the system and order the reading and recording of data could be identified. This device was generated so that two or more people could access and record data at the same time in the database without jeopardizing the level

of trust that the data could be read and registered. Observations of and interviews with pilot respondents also provided some useful feedback. For example, it was concluded that questions with blank fields for text input became easier to visualize on a darker, contrasting grey background (see Figure 2).

SCRIPT	DESCRIPTION
SENDMAIL	Sends registered in the database the invitation e-mails asking to participate on the survey to the companies. The e-mail is sent individually (the access link to the survey contains a code for each company).
RESET	Reestablishes access to the questionnaire to those whom accessed without answering it.
RETURNED	Registers, in the database, the companies whose e-mails returned due to inexistence of the e-mail address.
MARCAANR	Registers, in the database, the companies that received an e-mail for accessing and failed to respond.
MARCASKW	Registers, in the database, the companies that received an alert e-mail because they have not answered the key word.
MARCANR	Registers, in the database, the companies that received an e-mail reinforcing the invitation for participation in the survey.
RESETANR	Reestablishes access to the questionnaire to the respondents that accessed without answering.
RESETSKW	Reestablishes access to the questionnaire to the respondents that didn't complete the key word.
MAILANR	Sends e-mails to the respondents that accessed the questionnaire without answering it.
MAILSKW	Sends e-mails to the respondents that accessed the questionnaire without filling out the key word.
MAILNR	Sends the e-mails reinforcing the invitation for participation in the research (follow-up).
NEWMAIL	Modifies the e-mail addresses of the companies in the database - in case the respondent requests the change.

Table 1. ASP scripts developed.

Once done the modifications and necessary adjustments were completed, more than 200 tests simulating several use situations were conducted on the system. These tests included four respondents accessing and recording data at the same time, incidence of incomplete completion, failure to connect with the system, and so on. Tests were also completed on the ASP scripts, although for the e-mail sending ASP scripts, controlled simulations were accomplished before sending the messages to the sample.

The ASP scripts, described in Table 1, could be activated through a standard browser by any computer connected to the Internet. During the survey process, the ASP scripts facilitated the control process through the emission of reports indicating the executed routines and listing the participants involved in the process.
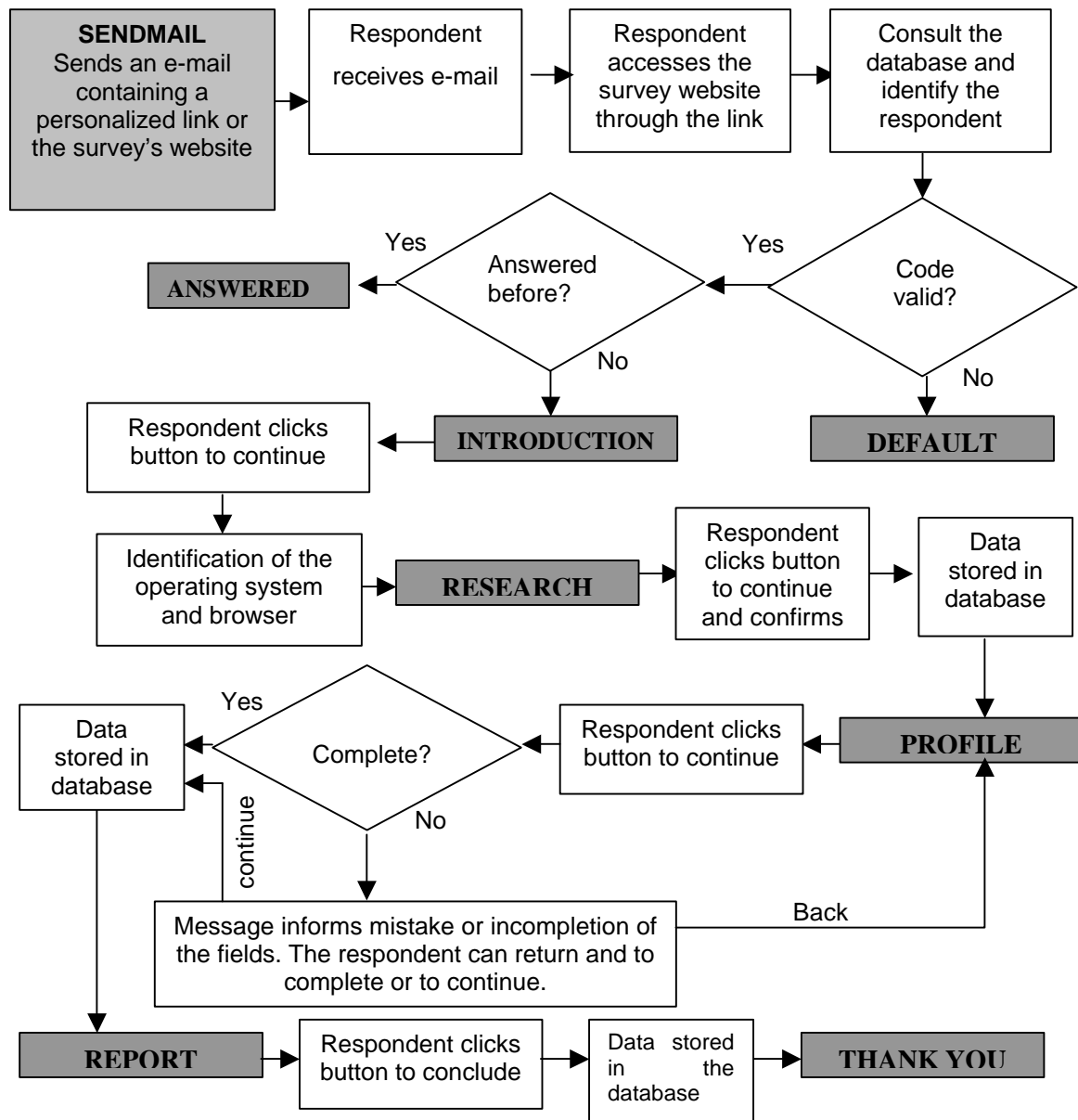
```
┌─────────────────┐    ┌─────────────┐    ┌─────────────┐    ┌─────────────┐
│   SENDMAIL      │    │ Respondent  │    │ Respondent  │    │ Consult the │
│ Sends an e-mail │    │ receives    │───▶│ accesses the│───▶│ database and│
│ containing a    │───▶│ e-mail      │    │ survey      │    │ identify the│
│ personalized    │    │             │    │ website     │    │ respondent  │
│ link or the     │    └─────────────┘    │ through the │    └─────────────┘
│ survey's website│                       │ link        │
└─────────────────┘                       └─────────────┘
```

Figure 4. Flowchart of site interaction.

The ASP scripts helped to facilitate the research, because they allowed the database to be manipulated without interrupting the data collection. Figure 4 presents the navigation flowchart of the site. The Web pages, mentioned in the above discussion, are shown in grey.

Contact e-mails for the survey

The strategy of data collection in this research was based on sending an e-mail to target respondents, explaining the objectives of the research and requesting that a person responsible for the company complete the questionnaire. Access to the questionnaire was accomplished through a URL presented in the body of the e-mail. These e-mail messages were generated automatically, and the code used was the company identification code of the SEBRAE-RS database. The use of individual links in the body of the e-mails had the intention of restricting and controlling access to the questionnaire (Bradley 1999, Simsek

1999). This system possesses the great advantage of ensuring that a company answers the survey once only. Also, it is possible to generate follow-up e-mails for those companies that failed to answer or partially answered the research instrument. Figure 5 shows how the URL was presented:
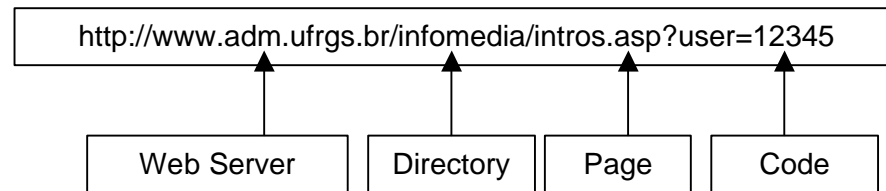

Figure 5. Web address of survey presented to potential respondent.

After properly validating the e-mail's text, the sending system was also checked. Firstly, a final system check-up was conducted, based on sending, access and recording tests. After that, in order to overcome some limitations of the server, the 9,730 e-mails in the database were grouped in clusters of 1,000. For each cluster, ten control users were generated. The control users contained data of the researcher and assistant. Upon completion of this process, the sending process was launched using one of the ASP scripts (SENDMAIL) and the reported sent list was verified. The reception of the e-mails by the control users verified the e-mail delivery process (numbered by the identification code). Note that the sending procedure was accomplished sequentially - one cluster at a time – thereby facilitating the control process.

Non-respondent control

After online data collection had been completed, an additional phase of data collection by telephone was carried out with companies that had not answered the survey (randomly selected). This procedure was used to check for the existence of a non-respondent bias. Such a bias would be verified if there are statistically significant differences among the results obtained in the previous collection (through the Internet) and the new sample (by telephone). If there is no evidence of differences, it gives stronger validity to the obtained data (Taylor, 2000; Churchill, 2001).

Firstly, the group of companies that had not accessed nor answered to the questionnaire was extracted from the database. Of this group, 100 companies were selected and ordered through a random draw. An interview protocol was developed, which served as a base for the two research assistants on how to carry out the interviews. After the training, four interviews were simulated and monitored in order to eliminate error and interviewer bias.

The companies were contacted based on the random draw. Thus, 80 companies were contacted. From this total, 30 were willing to answer the questionnaire by telephone. Among the other 50 companies, 31 alleged that the CEO of the company was not there, eight asked for the questionnaire by e-mail (but failed to answer it), six were on vacations, and five refused to answer. This is a common problem with telephone interviews. Simsek (1999), for example, observed that one of the disadvantages of data collection by telephone in comparison with e-mail is that it demands that the respondent be available to answer the call (even though it might have been previously scheduled).

For the statistical analysis of non-respondent bias, t-tests were used for the means and χ^2 tests for the frequencies. Alongside, an analysis was made of the keyword frequencies obtained in each respective sample. None of the tests found any significant differences between the respondent and non-respondent samples. Thus, after the series of comparisons and tests accomplished between the two samples, was concluded that there is no evidence of non-respondent bias (Taylor, 2000).

Questionnaire follow-ups

During the first day of data collection for the research, we discovered that many participants were accessing the website but not answering the questionnaire, or only completing demographic questions. As a consequence, for each one of these cases, a follow-up e-mail, which was sent in the same day of the access, was developed and validated. This procedure was carried out in the following way: cases were detected through consultation of the database, questionnaire access for the companies was reestablished, and e-mails were sent requesting an answer to the research.

In addition to the follow-up e-mails, automated e-mails were also sent to those who did not attempt to partake in the research. In essence, those who had not accessed the questionnaire were identified, and an e-mail was sent requesting them to answer the questionnaire. This system, validated in the pilot stage, was activated 7 and 17 days after the first sending.

Follow-ups are widely used in traditional survey methods. On this experience, it has proven to take great advantage from the benefits of automation.

Results

As shown previously, the survey started by sending an e-mail notification to 9,730 companies registered in the SEBRAE-RS database. The data collection was carried over during 21 days. The extended length of this period is due to the occurrence of a holiday, and also because the survey months of January and February are marked as being a typical period of vacation in the State of Rio Grande do Sul. Of the 9,730 e-mail addresses, 7,963 companies had valid e-mails. Of this group, 939 (11.80%) answered to the questionnaire. During the organization of the data, 143 questionnaires (15.20% of the 939) were excluded (104 did not present any key-words and 39 didn't present keywords pertinent to the theme). Thus, 766 (9.62% of the valid e-mails) answered the on-line questionnaire correctly, and 30 (0.38%) answered by telephone during the non-respondents' analysis, totaling 796 valid questionnaires (10.00% of the total of valid e-mails).

The response rate obtained is excellent when compared with other studies previously conducted in the same region using the Internet as the communication media. Recent studies carried out in Rio Grande do Sul obtained 5.48% (Vieira, Viana & Echeveste 1998) and 3.23% (Vieira 1999) of valid answers. This high level of answers can be attributed to the fact that this study used a very rigorous control system, which was lacking in the previous investigations. The control systems generated - besides allowing and facilitating the sending of follow-ups to the non-respondents - also ensured that no company answered the questionnaire more than once.

Figure 6 presents the number of questionnaires received daily during the data collection phase. The red line indicates the dates that the follow-up e-mails were sent.
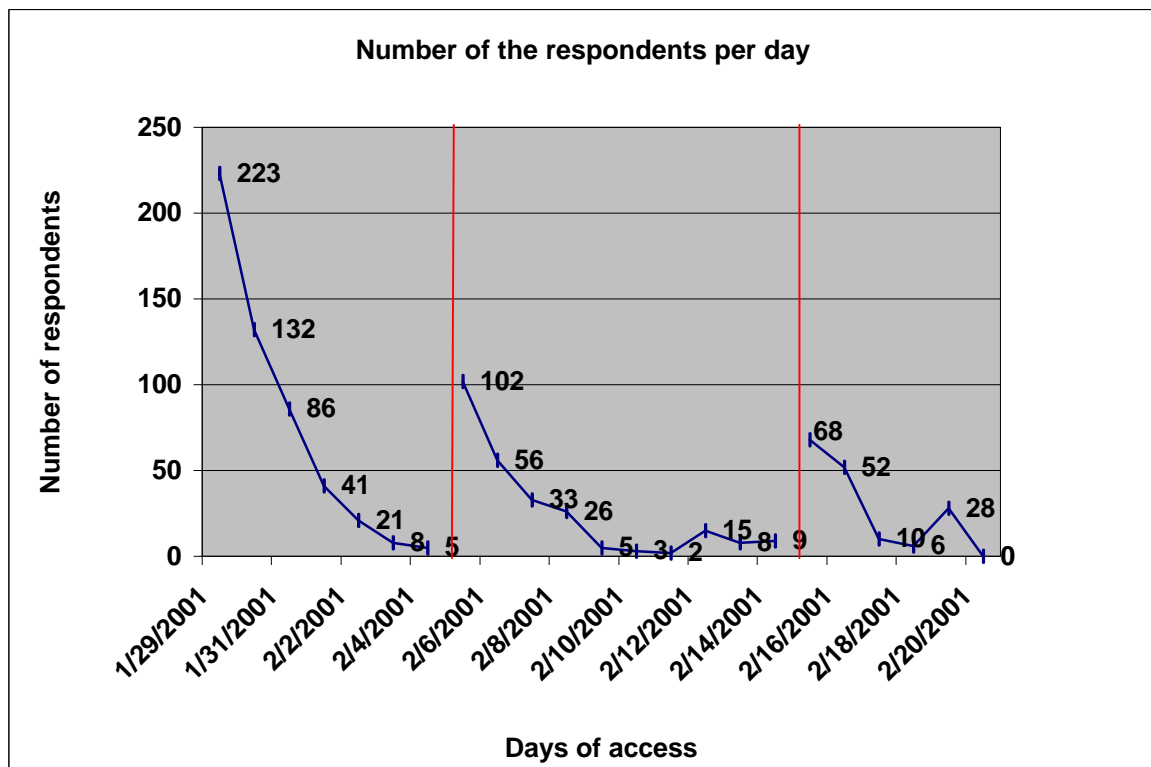
Figure 6. Number of daily responses.

Of the 939 answered questionnaires, 516 companies accessed the survey between January 29th and February 4th. In the period after the first follow-up e-mail (from February 5th to the 14th), the return was 259 companies. Finally, after the second follow-up e-mail on February 15th, 164 companies completed the questionnaire. Figure 7 provides more detail on the first two weeks of the data collection process. Specifically, a line of exponential tendency has been added to the response curves of the initial data collection phase and first follow-up phase.

Figure 7 demonstrates that the curves generated from the collected data (Week 1 and Week 2ii) are quite similar to their respective tendency lines (exp. week 1 and exp. week 2). This is also evidenced by the degree of explained variance found ($R2 = 0.989$ for week 1 and $R2 = 0.959$ for week 2). Successive follow-ups tend to lose impact among respondents, turning a decreasing answer rate. The fraction obtained by the integralsiii (from zero to infinity) of the curves Week 2 by Week 1 (4.5474/10.5320 = 0.4343) allows us to infer that the added value of the follow-up is equivalent to 43% of the effect of the first notification of the survey.
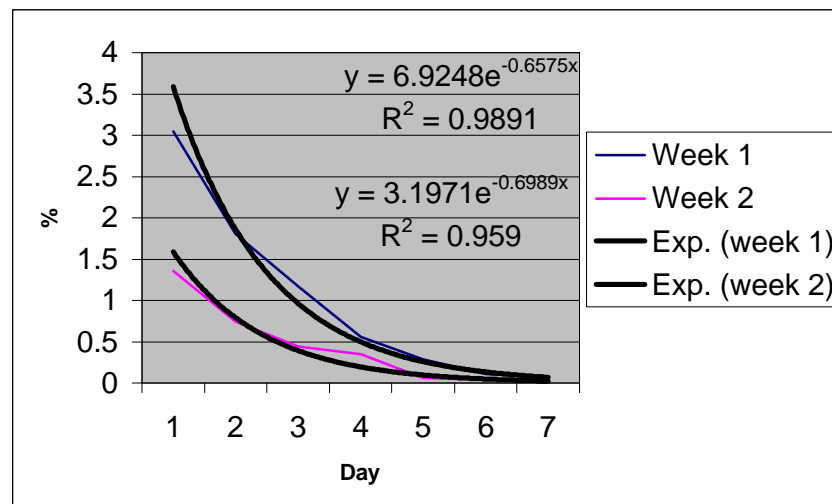
The chart shows percentage (%) on the y-axis (0 to 4) versus Day on the x-axis (1 to 7), with curves for Week 1, Week 2, Exp. (week 1), and Exp. (week 2). The fitted equations displayed are:

$$y = 6.9248e^{-0.6575x}$$
$$R^2 = 0.9891$$

$$y = 3.1971e^{-0.6989x}$$
$$R^2 = 0.959$$

Figure 7. Response rate analysis.

Of the 766 companies that answered the questionnaire online, 334 (43.60%) did not receive follow-up or alert e-mails. As mentioned above, follow-up and alert e-mails were used in a stratified way. Thus, the companies that accessed the questionnaire, but didn't answer the questions, were identified and received an alert e-mail on the same day of the access. These were referred to as "ANR - non-respondent access" cases. From the 450 ANR alert e-mails sent during the data collection period, only 69 (15%) accessed the website and answered the questionnaire correctly. Those respondents that accessed the questionnaire and provided demographic data, but didn't answer the question relating to their informational needs were another group that received an alert e-mail on the same day as the access. This group was referred to as "SKW - without keywords." Of the 169 SKW e-mail alerts, 55 (32%) returned and completed the questions. Interestingly, via the support activities, it could be noticed that the lack of response regarding informational needs was usually motivated by lack of interest, time or attention by the respondents. Follow-ups for the large body of companies that had not accessed the questionnaire excluded both the 1,767 e-mails that had invalid e-mail addresses (identified after the first sending) and those that had already received an SKW or ANR alert. Therefore, in the first follow-up, 7,325 e-mails were sent and, in the second follow-up, 6,930 messages were sent.

During the entire period of data collection, the support activities (by e-mail, phone and fax) were accessed by respondents 99 times.

Representativeness of the sample

As mentioned above, we were keen to assess the possibility of a non-respondent bias, and therefore undertook a control procedure with random sample of 30 companies, surveyed by telephone. Both statistical tests (t-test and $\chi2$) and a linguistic analysis of the keywords found no statistically significant differences. This contributed positively to confirmation of the representativeness of the sample. Comparison of the profiles of the samples did, however, find some statistically significant differences in gender and respondents' position within the organisation. In the telephone survey there was a statistically significant increment in the number of respondent women (at the 5 per cent level); in the e-survey, they represented 19.55% while by telephone they represented 42.86% of the participants. Regarding the

respondents' position, there was a fall in the number of business owners in the telephone survey; in the e-survey, 54.25% of participants were business owners, but this fell to just 14.81% in the telephone sample. As a consequence, the number of managers rose from 17.42% to 48.15%. Overall, with these exceptions, the research samples demonstrated only small differences in respondents' profiles. Since the telephone survey was conducted after three e-mail notifications, this might have been considered an insistence, and influenced the delegation of responsibility for answering the questionnaire to someone apart from the business owner. The gender differences of the samples could be a consequence of the differences in respondents' position (Oliveira, Oliviera & Dalfior 2000).

A key concern of the research was that the sample collected was representative of the geographical region under investigation. Different areas within the State had quite differing characteristics, and we wanted to be sure that the sample accurately represented the complex mix of organizations within the different zones of the State. Therefore, we used a χ^2 test in an attempt to verify whether the respondents' geographical distributions in the different areas of the State of Rio Grande do Sul is equivalent to those existent in the SEBRAE-RS database. Area codes were used to divide the State in four zones (Figure 8). Table 2 shows the representativeness of each area.
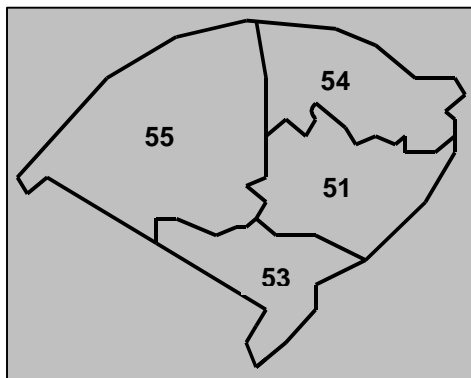
Área Code	SEBRAE	SAMPLE
51	61.28%	64.83%
53	4.19%	4.02%
54	23.30%	22.24%
55	11.23%	8.92%

Figure 8. Regions by area code.   *Table 2. Representativeness of the regions.*

On face value, the sample appears to be representative. The high concentration of respondents in zone 51 is observed in both the population and collected sample. This is easily explained because it includes the metropolitan area of Porto Alegre (the State's capital), Sinos Valley and Taquari. The Northeast Area (zone 54) is also a major area of population, and well represented in the sample. On the other hand, the areas of the South (zone 53) and Center-West (zone 55) are much smaller areas of population - these regions' economic base is primarily agricultural - but seem appropriately represented. Overall, χ^2 did not demonstrate significant differences in the sample and database population. Therefore, after these analyses, it was concluded that the sample of the population was representative (Taylor 2000, Churchill 2001).

Summary and Conclusions

This article has presented an Internet data collection system as a way support to e-surveys. The rigorous and the systematic approach adopted in the test and validation stages allowed the researchers to identify some key problems that, when corrected, made the system less susceptible to failure during the data collection procedure.

The system was both cost-efficient and allowed access to a very large and quite representative sample. Another important aspect of system development is to provide simplicity and functionality, both of which were adequately achieved in this study. The performance of the data collection system, in comparison to similar studies, can be considered excellent; the percentage of respondents was at least twice that obtained in previous studies accomplished with a similar sample (which, typically, did not use a rigorous control procedure for respondents). Another interesting finding from the research is that the survey completed via the Internet reached higher hierarchical levels in the organizations than the telephone survey (i.e., the non-respondents' analysis).

One of the key strengths of the developed system is the rigorous control of respondents. From the literature, this is an issue that presents one of the most challenging problems in the development of Internet surveys. Control tools helped increase response rates. On the other hand, it created several issues towards anonymity and confidentiality - which could be a factor that discourages some people from participating in surveys. This is a dilemma where the researcher is trapped between having control or protecting the anonymity to the respondents.

The e-survey system discussed in this paper was developed and implemented as part of a research programme examining the information needs of Brazilian growth enterprises. Further details on the survey itself can be found in another, related paper. The programme of research investigates the process of information intermediation through the Internet, and specifically, how this relates to Brazilian enterprises. Future research plans include further developments to the e-survey system, and more extensive work on the related methodology. Ultimately, the goal is to prepare and distribute an automated, generic e-survey support system (and associated methodology) that is freely available to the scientific community.

References

Babbie, E (1999), *Métodos de pesquisa de survey*, Editora UFMG, Belo Horizonte.

Betancourt, M (1999), *The best Internet businesses you can start*, Adams Media Corporation, Holbrook.

Boyer, KK, Olson, JR, Calantone, RJ & Jackson, C (2002), 'Print versus electronic surveys: A comparison of two data collection methodologies', *Journal of Operations Management*, vol. 20, no. 4, pp. 357-373.

Bradley, N (1999), 'Sampling for Internet surveys: An examination of respondent selection for Internet reseach', *Journal of the Market Research Society*, vol. 41, pp. 387-395.

Churchill, GA (2001). *Basic marketing research*, Dryden Press, Fort Worth.

Clayton, RL and Werking, GS (1988), 'Business surveys of the future: The World Wide Web as a data collection methodology' in *Computer-assisted survey information collection*,

eds. MP Couper, RP Baker, J Bethlehem, CZF Clark, J Martin, WL Nicholls & JM O'Reilly, John Wiley, New York, pp. 543-562.

Cook, C, Heath, F & Thompson, RL (2000), 'A meta-analysis of response rates in Web- or Internet-based surveys', *Educational and Psychological Measurement*, vol. 60, no. 6, pp. 821-836.

DeVaney, T (2001) 'Statistical significance, effect size, and replication: what do the journals say? *The journal of experimental education,* vol. 69, no. 3. pp. 310-320.

Dillman, DA (1978), *Mail and telephone surveys: The total design method*, John Wiley, New York.

Dillman, DA (2000), *Tailored design of mail and other self-administered surveys*, Wiley-Interscience, New York.

Epstein, J, Klinkenberg, WD, Wiley, D & McKinley, L (2001), 'Insuring sample equivalence across Internet and paper-and-pencil assessments', *Computers in Human Behavior*, no. 17, pp. 339-346.

Gil, AC (1994), *Métodos e técnicas de pesquisa social*, Atlas, São Paulo.

Groves, RM, Biemer, PP, Lyberg, LE, Massey, JT, Nicholls, WL & Waksberg, J, eds. (1988), *Telephone survey methodology*, John Wiley, New York.

Hoffman, DL & Novak, TP (1996), 'Marketing in hypermedia computer-mediated environments: conceptual foundations', *Journal of Marketing*, vol. 60, no. 3, pp. 50-73.

Hoppen, N, Lapointe, L & Moreau, E (1996), 'Um guia para a avaliação de artigos de pesquisa em sistemas de informação' in *READ – Revista Eletrônica de Administração*, PPGA-EA-UFRGS, Porto Alegre.

ITU (2001), 'European Internet commerce growing steadily', viewed 28 August 2002, <http://www.itu.int/itunews/issue/2001/05/inbrief.html>

Kirk, RE (1996) 'Practical significance: a concept whose time has come' *Educational and Psychological Measurament,* no. 56, pp. 746-759.

Kish, L (1965), *Survey sampling*, John Wiley, New York.

Liaw, SS (2002) 'An Internet survey for perceptions of computer and the world wide web: relationship, prediction and difference' *Computers in human behaviour,* no. 18, pp. 17-35.

Mattar, FN (1994), *Pesquisa de marketing: Metodologia, planejamento, execução, análise*, Atlas, São Paulo.

Mosquera, M (2002), 'Global e-commerce to hit $5 trillion in 2005', viewed 28 August 2002, <http://www.internetwk.com/story/INW20010523S0005>

Nix, TW & Barnette (1998) 'The data analysis dilemma: ban or abandon. A review of null hypothesis significance testing' *Research in School,* vol. 5, no. 2, 3-14.

Oliveira, N, Oliveira, RCM & Dalfior, SDR (2000), 'Gênero e novas perspectivas de trabalho: um estudo junto a mulheres gerentes de atendimento no banco do Brasil' in *Anais do XXIV Encontro da Associação Nacional de Programas de Pós-Graduação em Administração*, Florianópolis, September.

Pinsonneault, A & Kraemer, KL (1993), 'Survey research methodology in management information systems: an assessment', *Journal of Management Information Systems*, vol. 10, no. 2, pp. 75-105.

Ranchhod, A & Zhou, F (2001) 'Comparing respondents of e-mail and mail surveys: understanding the implications of technology' *Marketing Intelligence & Planning,* vol. 19, no. 4, pp. 254-262

Shannon, DM & Bradshaw, CC (2002), 'A comparison of response rate, response time, and costs of mail and electronic surveys', *Journal of Experimental Education*, vol. 70, no. 2, pp. 179-192.

Simsek, Z (1999), 'Sample surveys via electronic mail: a comprehensive perspective', *Revista de Administração de Empresas*, vol. 39., no. 1., pp.77-83.

Schubert, F & Hepworth M (2000) 'The implementation of an electronic survey tool to help determine the information needs of a knowledge-based organization' *Information Management & Computer Security*, vol. 8, no. 2, pp. 53-64.

Taylor, H (2000), 'Does Internet research work?' *International Journal of Market Research*, vol. 41, no. 1, pp. 51-63.

Vehovar, V, Lozar Manfreda, K & Batagelj, Z (2001), 'Sensitivity of electronic commerce measurement to the survey instrument', *International Journal of Electronic Commerce*, vol. 6, no. 1, pp. 31-51.

Vieira, BLA, Viana, DA & Echeveste, S (1998), 'Comércio eletrônico via Internet: uma abordagem exploratória' in *Anais do XXII Encontro da Associação Nacional de Programas de Pós-Graduação em Administração*, Foz do Iguaçú, September.

Vieira, BLA (1999), A Internet como canal de compras: um estudo junto aos usuários do provedor, Master's thesis, PPGA-EA-UFRGS, VIA-RS, Porto Alegre.

---

[i] SEBRAE-RS translated as the Brazilian Micro and Small Business Support Service. This is a non-profit independent institution, supporting the development of small-sized business activity. The institution is funded by both the public and private sectors, and the country's main fostering and research entities.

[ii] The data used in the curve "Week 2" is derived from the equation $y = y_x - (6,9248 * e^{(-0,6575x)})$, where y_x is the respondents' gross percentage obtained each day after the follow-up (Week 2), x is the order of the day of the week, and y is the respondents' net percentage obtained each day after the follow up. This operation allows discounting the number of answers theoretically expected in Week 2 without sending follow-up e-mails.

[iii] The integral of the curve $y = a \times e^{-bx}$, where the value is given by a/b, offers a theoretical estimation of the total percentage of return for each e-mail cluster sent.