

USING THE INTERNET AS A SURVEY MEDIUM: LESSONS LEARNT FROM A MOBILITY SURVEY OF YOUNG DRIVERS

Warren Harrison
Eastern Professional Services Pty Ltd

Ron Christie
RCSC Services Pty Ltd

ABSTRACT

The internet has the potential to be an important research tool, and is increasingly considered as a medium for conducting surveys in addition to (or in place of) more-traditional media such as mail and telephones. A survey of young drivers conducted in New South Wales made use of this technology to collect information concerning the role of driving in meeting their mobility needs. Participants were sampled through Government and Catholic secondary schools across the State, with schools agreeing to hand out invitation and reminder cards to senior students. The survey could be downloaded from a web site identified on the invitation and reminder cards, and survey data were automatically uploaded to the survey internet server. Participation in a prize draw was offered as an incentive. Ten thousand invitations and reminders were distributed through 120 schools, but the response rate was insufficient to meet the original aims of the study. This paper discusses the survey and sampling method, the use of the Internet as a research medium, and the implications of this project for future internet-based surveys.

INTRODUCTION

The Motor Accidents Authority of New South Wales (MAA) commissioned Eastern Professional Services Pty Ltd with assistance from RCSC Services Pty Ltd to undertake a survey of younger drivers to assess their mobility needs and their reliance on driving to meet those needs. The study arose out of an increasing interest in the potential value of exposure-reduction measures for younger novice drivers – such as night-time and passenger restrictions.

It is broadly assumed that young drivers will respond negatively to driving restrictions, and that any restrictions on mobility may be regionally or socio-economically inequitable. There is no published evidence, however, to confirm or disconfirm this. It was considered useful to collect data concerning the mobility needs of younger drivers and how they make use of private and public transport to meet those needs.

The current study used a survey methodology to collect information from a sample of young drivers that would contribute to further discussions of the potential benefits and costs of exposure-reduction measures. The survey was originally planned as a paper-and-pencil survey to be distributed through a sample of secondary schools and mailed back by participants, but schools were reluctant to agree to manual surveys due to the increased workload on staff. However, the high level of computer literacy amongst this group of drivers, and the general availability of computers and internet access throughout the community led to a decision to conduct the survey electronically as an alternative which did not burden schools and teachers.

The response rate was very poor, and there were insufficient responses to allow any meaningful analysis of the data. This paper details the research method and outcomes, and then goes on to discuss the response-rate issue. While the project did not meet its original aim of providing information that could be used in a policy setting, it does lead to consideration of the implications for the use of electronic technologies for survey-based research.

METHOD

Original Research Method

It was planned to use a survey instrument to collect data from a sample of young drivers. The survey instrument was designed specifically for the project to include demographic characteristics; lifestyle and mobility-needs issues, especially as they relate to transportation requirements; mobility preferences and behaviours; the availability and acceptability of alternative transportation methods; driving exposure patterns driving behaviours and safety through the Driver Behaviour Questionnaire, and self-reported crash and near-miss experiences; and perceptions of and contact with traffic law enforcement. The questionnaire was planned to take about fifteen minutes to complete, and a small incentive was to be offered (inclusion in a prize draw). A sample size of about 800-1,000 was considered sufficient.

Developing the Method

The survey was originally planned as a paper-and-pencil survey and logbook to be distributed to senior students through secondary schools. Schools would be asked to identify recently-licensed students and to provide them with a survey/logbook package that would include a reply-paid envelope. The logbook would require data entry for one week. Initial planning was based on an assumption that it would be necessary to invite 3,000 young drivers to complete the survey/logbook to achieve the desired sample size of 800 – an assumed response rate of just over 25%. Based on estimates of school sizes, it was calculated that obtaining agreement from 84 schools across NSW would provide a large-enough pool of potential participants.

Schools were selected geographically, based on ABS data concerning the distribution of 15-19 year olds across the State. The possibility that schools would refuse to participate was taken into account by selecting 102 state schools, and an additional 25 schools in the Catholic system (also selected to match the ABS data) – giving a total of 127 schools to be asked to participate in the project.

Concerns about the likely response of schools to a request to identify licensed drivers and manage the distribution of survey packs led to reconsideration of the survey method. It was proposed and agreed that the survey could be conducted using Internet technologies. Software was developed that would allow participants to complete the survey and the logbook over a seven-day period on their own computer, and that would return the data to the Eastern Professional Services Pty Ltd computer server automatically using FTP. This software was developed to allow participants to download the software installation package, install the software, complete the survey and logbook, return the data, and then uninstall the software.

It was intended that schools hand out invitation cards (and then reminders one week later) to all senior students inviting those with licences to take part in the project, using the same incentive. This would make the process easier for schools. A letter was sent to principals asking for their participation and outlining their involvement. Schools were then contacted by telephone to seek their agreement and to ask for an estimate of the number of students who could be given an invitation to participate in the project. Some principals had not read the letter and many had passed it on to another staff member. Fifty schools responded positively, and eighteen refused to participate. The remaining fifty-nine school principals (or delegates) could not be contacted – primarily because they did not return telephone calls or were not available when called on a number of occasions. The schools that agreed to participate provided a total estimate of 4,800 students for invitations. The average number of senior students in these schools was used as the basis for predicting the potential number of participants in the schools that could not be contacted, giving a total student population (for the invitations) of about 10,000.

A web site was developed and placed on the Eastern Professional Services Pty Ltd web server, along with the downloadable survey software installation package. A counter was set up to track site visits and downloads from the site. Ten thousand invitation cards and reminder cards were printed. They directed participants to the web site and gave a cut-off date of 16 April 2004, which was subsequently extended to 7 May for reasons outlined below. These were packaged and sent to the contact persons/principals in schools in the first week of March 2004.

The survey software collected demographic and safety-related information, and responses to the Driver Behaviour Questionnaire items. It included a daily log book that asked participants to provide some information about four trips (driving or using any other form of transport) for each day over a one-week period. Error checks were built into the software and it was designed to make it impossible for participants to edit responses to earlier items. Once they completed the final day's logbook, the software allowed them to submit the whole data set over the internet, with the data automatically uploaded to the server.

Problems During Data Collection

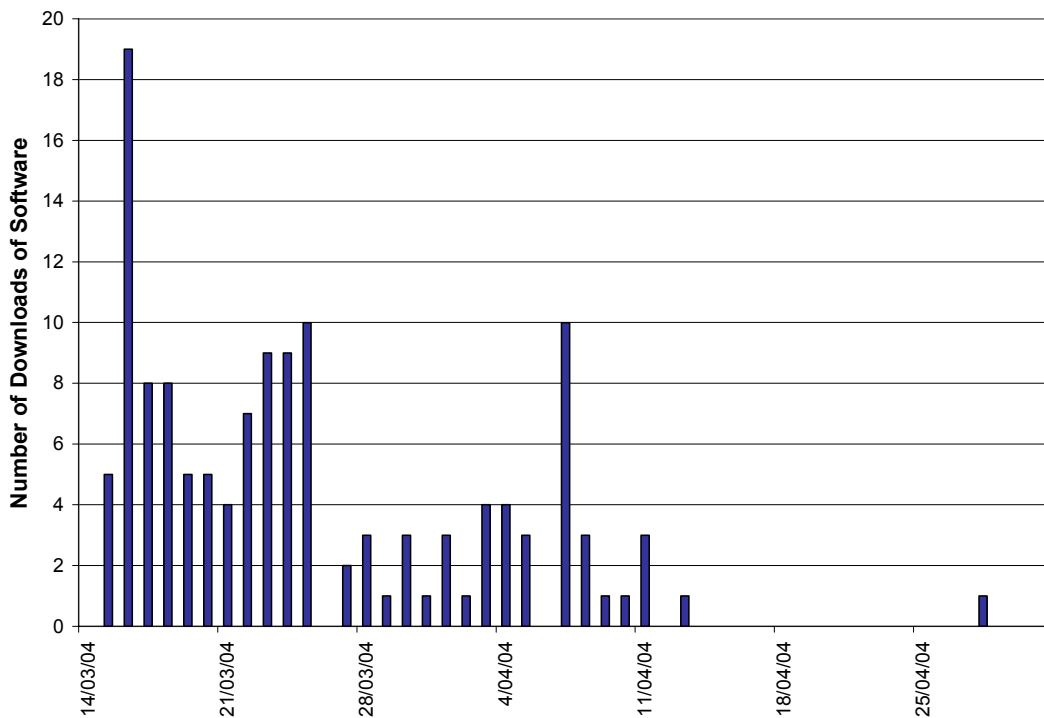
About two weeks after the web site was on-line, a regular update of the McAfee Antivirus software virus definition files mistakenly identified a Trojan in all software installation packages that were based on the Multimedia Fusion installation package – including the survey software installation package for this project. The problem was admitted by McAfee, but a fix was not made available for almost two weeks. No other antivirus packages had this problem. The concern here was that any participant who downloaded the installation package and who had McAfee software would have the package identified as containing a Trojan, and would then have it deleted from their computer. It was thought that any news of this would reduce the response rate, as the installation package would be perceived as a risky download.

The software installation package was then rewritten to work around the problem, and a new version placed on the web site. About four days elapsed between the McAfee problem starting and the new version being placed on the site. The effect of this problem is unknown but is unlikely to be substantial because the base download rate for the survey by this stage was already very low (see below).

Tracking of downloads suggested that the response rate was very low, so a small sample of schools was contacted in late March to assess the response of students to the package. Schools indicated that they had distributed the material to target students, and that the response from students had been generally positive. Schools were all positive about the use of an internet delivery method. There was no obvious reason for the low response rate. A letter was sent to all schools as a further reminder – the letter was couched as a “thank you” letter but was intended as a reminder if schools had not yet handed out the invitation cards, and as a prompt to encourage schools to remind students about the survey.

SURVEY RESULTS

A total of 267 hits were recorded on the survey web site, with a total of 134 downloads of the survey software. Twenty-nine data sets were uploaded by participants who completed the survey and logbook. Based on an assumption that all 10,000 invitations and reminders were handed out (this is discussed below), this equates to a web-site-visit rate of 2.7%, and download rate of 1.3%, and an overall response rate of 0.3%. About half of the visitors to the web site downloaded the survey software, and about 22% of people who downloaded the software completed their involvement in the study and returned their data. The overall response rate was surprisingly low. The graph (below) shows the pattern of downloads of the survey software over the time period during which it was available. The majority of survey downloads occurred in the first two weeks, and only a small number of downloads occurred after the thank-you/reminder letter was sent to schools.



DISCUSSION

Surveys on the Internet

Internet-based surveys provide a number of cost-related benefits, are thought to provide more-complete and valid responses to items where there might be reasons to be concerned about privacy, and provide opportunities to develop and implement complex survey designs (that skip questions and check responses) without having to rely on an interviewer or respondent to interpret skip instructions accurately (Tourangeau, 2004). These positive characteristics are countered by potential sampling problems in internet-based surveys.

In this context, however, it is important to note that all survey methods have associated sampling problems. Survey methods aim to produce data that can be generalised across the entire population of interest. To meet this aim, the sample needs to be an unbiased sample of the population in relation to the items of interest in the survey. Sampling bias results from a number of sources:

- **The differential inclusion of some population members in the pool of people available to be sampled.** In the case of internet-based surveys, generalisation of results is limited to those members of the population of interest who have access to computers and the internet. The difference between internet users and non-users (see Law and Justice Foundation of New South Wales, 2004) suggest the need for caution when generalising from internet-based surveys dealing with issues that might be related to education level and socioeconomic status in particular.
- **The method used to sample from the available population.** Accessing members of the internet-active population is difficult as there is no central collection of data that allows identification of email addresses, and recent developments in privacy and internet-related law make it difficult to contact internet users directly. Contact through the internet is also being actively blocked by anti-spam software. It is almost certainly necessary to use non-internet methods to access potential sample members for internet-based surveys.
- **Differences in the decision to take part or refuse to participate.** There are likely to be important differences between people who agree to take part in a survey and those who refuse. These differences may interact with survey content and make it difficult to generalise to all population members. This is complicated because refusers may have different underlying motivations for refusing – such as people who always refuse to take part in surveys, those who may have competing motivations at the time of the survey and refuse, and those who may refuse because they are not interested in the topic. This is compounded in internet surveys by the possibility that potential sample members who view the web site for the survey and take part in the survey have made two decisions rather than one. While a respondent to a telephone survey invitation is passively involved in the recruitment process and needs only to decide to continue a telephone interview started by the interviewer, the respondent in a web-based survey must make an active decision to view the site and then must make an additional decision to perform the activities required to complete the survey. This active role

and the decision making processes involved add another potential bias into the sampling process.

- **Some participants may choose not to complete the survey.** Once involved in a survey, it is still possible for participants to withdraw their agreement to take part, and in sensitive topic areas it is conceivable that there could be a relationship between potential survey responses and the decision to withdraw. It is easier to discontinue an internet-based survey as the participant does not need to deal with the psychological issues raised by doing so.

The potential for sample biases to influence the outcomes of surveys is so great that survey methods cannot generally be relied upon to produce data that can be generalised to the larger population without broad cautionary comments. It is possible to use stratified or quota-based sampling techniques and weighting techniques during data analysis if it is possible to compare the demographic characteristics of the sample with those of the population. These techniques are unlikely to solve the sampling bias problem, however, as it is still not possible to compare the characteristics of the sample on the issues of interest in the survey with the same characteristics in the population. If it was possible to do this, the survey would be unnecessary.

There is consistent evidence that response rates to surveys are declining. Tourangeau (2004) noted that it is now difficult to obtain response rates in telephone surveys higher than 60 percent – even when aggressive follow-up procedures are used. This appears to be related to three factors – many homes now have access to technological features that allow call screening (eg Link and Oldendick, 1999); the actual refusal rate for potential participants once contacted appears to be increasing (Groves and Couper, 1998); and demographic changes make it more difficult to obtain information from some members of the population. Declining response rates are a significant problem for all survey methods because they increase the potential for sampling bias – as additional factors cause people to avoid contact with interviewers or refuse to participate, the likelihood that reasons for refusal will be correlated to the topic of the survey increases, and the potential for biases to limit the generalisability of surveys increases. Response rates in internet-based surveys are generally poorer than response rates in telephone surveys (Couper, 2001)..

While there are methods that can reduce non-response rates, they add to the potential bias problem of survey methods in ways that are difficult to quantify. People who respond to aggressive recruitment and follow-up methods may differ from those who don't in ways that are not easily measured; the use of incentives may change the underlying motivational characteristics of the sample; and the use of panels of regular participants raises many concerns regardless of the apparent representativeness of these panels on a range of demographic variables.

The use of an internet-based survey was a reasonable decision given the general state of evidence concerning this method, the likely negative response to another paper-and-pencil survey in schools, and the potential cost savings for a large sample when compared to alternative methods. Internet methods have their own sampling and response-quality issues but in general terms seem to be as acceptable as other survey methods (e.g. Gosling, Vazire, Srivastava, & John, 2004) .

Truell, Bartlett, And Alexander (2002) investigated the potential use of internet-based surveys by comparing the response numbers and response times to an internet-based survey and a mail survey. They concluded that there were clear advantages to using internet-based surveys, especially in relation to response times, but that the method was limited to those potential participants with access to the internet – making generalisation to the population problematic.

The potential for the internet to be used in surveys of specific populations that are hard to access using traditional methods has been an important theme in survey research on the internet. In particular, there has been a number of studies concerning sensitive issues in the gay community (eg. Halkitis, Parsons, & Wilton, 2003), and some studies dealing specifically with the potential value of this method for some populations (eg. Mathy, Schillae, Coleman, & Berquist, 2002). As Truell et al. (2002) noted, some subgroups of the population have very high levels of internet access and internet-based survey techniques are likely to be especially useful.

In the context of the survey reported for this project, the use of an internet method seems appropriate given the relatively high level of internet access amongst late-teens in Australia. The Law and Justice Foundation of New South Wales (2004) noted that about 85% of people aged between 14 and 24 had access to the internet.

The choice of a survey method was made in the context of research literature that suggests that survey outcomes using dynamic survey presentation (using software to present items on multiple “pages”) were similar to those using static presentation (a survey on a single web page)(Norman et al., 2001). The survey used here included a logbook to be completed over a seven-day period, so a dynamic system based on a small, downloadable program was used. There is no reason in the literature to believe that this approach should have caused any specific problems, but Tourangeau (2004) does note that this is an area in need of additional research.

So What Went Wrong?

While the literature concerning sampling bias and response rates in internet and other surveys gives cause for some concern about the application of any survey method, there was no reason to believe that the response rate would be as low as it was. The worst case located in a detailed search of relevant literature was reported by Kent and Lee (1999), who emailed 8,300 invitations to participate in a survey and obtained a response rate of 2.9% (compared to the current project’s rate of less than one percent). It does need to be noted, however, that the peer-reviewed literature is unlikely to include examples of studies that failed to generate a sufficient sample size, so it is unclear whether there have been other examples of the recruitment problem in the current study. Possible reasons for the low response rate in the current study include the following:

- **It is possible that schools failed to distribute the invitation and reminder cards.** As noted earlier, contact with a small number of schools suggested that this was not the case. Even if all the schools that were not contacted initially by telephone (before sending the invitations and reminders) failed to distribute the cards, the response rate to the survey was still well under one percent. It does not seem likely that the low response rate can be accounted for as a result of a failure to distribute the invitation and reminder cards.

- **It is possible that student-access to the internet was lower than anticipated.** This seems unlikely given the evidence from other sources and the age group of potential participants. The large number of young novice drivers participating in the University of Sydney's current Drive Study – which also uses an internet based survey – suggests that internet access is not likely to have been a problem.
- **It is possible that accessing, downloading, and using the survey software presented problems for participants.** It is not likely that there were problems accessing the web server used to distribute the software. The web address was provided on the invitation and reminder cards, and the site was checked regularly. No potential participants contacted Eastern Professional Services Pty Ltd (contact details were provided on the invitation and reminder cards) for guidance accessing the site or using the software. The file was not large – it was less than one megabyte and took about 40 seconds to download with a dial-up connection.

It is possible that some potential participants may have been concerned about downloading and installing software given general warnings about the risk of doing so. The web site included information that the file was virus-free and that the software would leave nothing behind on the computer once the participant had completed the logbook, but this may not have reduced concerns for all potential participants. This may explain the difference between the number of visitors to the web site and the number of times the software was downloaded, but only 267 visits to the web site were recorded. It is unlikely that concerns about downloading the software contributed strongly to the low response rate.

It was noted earlier that McAfee antivirus signature files mistakenly identified a Trojan program in the installation package, and that it was a number of days before a workaround could be developed. It is possible that some participants may have downloaded the file and had a virus alert generated either at that time or when they tried to run the program. No participants contacted Eastern Professional Services Pty Ltd about this, and the small number of downloads (134) means that this was not a common experience. It is difficult to conceive of an informal communications network that could influence the decision making of a large number of students across New South Wales in such a short time period.

The software managed uploading data once the participant had completed the survey and logbook. It is possible that some participants may have had problems uploading the data, although this part of the software was thoroughly tested and no participants contacted Eastern Professional Services Pty Ltd. More importantly, only 134 people downloaded the software, suggesting that data upload problems were not the cause of the poor response rate.

The low number of site visitors and the poor download numbers suggest that the response-rate problems were not related to technical aspects of the survey or to any human-factors problems in making use of the software. The problem is more likely to relate to factors underlying potential participants' decisions to access the web site or not – i.e. their initial decision to take an interest in the survey.

- **It is likely, therefore, that motivational factors may underlie the poor response rate.** It is interesting that these factors appear to have influenced a large number of young people at the same time, with a similar outcome. It is

possible that young people may not have been interested in the survey topic, or may not have been motivated enough to access the survey web site. It is difficult to believe that this would be so across all potential participants, and inclusion in a prize draw for a reasonably large voucher was thought to be a reasonable incentive. The apparent success of the University of Sydney Drive Study confirms that a broad lack of interest was not a factor in the failure of the project, although the University of Sydney used a small incentive for all participants rather than the opportunity to win a smaller number of large incentives as was the case here.

Students may have had competing motivations and time pressures. Senior students tend to have busy school and social lives and sometimes hold part-time jobs too. It is possible that these interfered with the survey. This also seems unlikely as an explanation for the low response rate. The survey period included a school holiday period, and while some students might have misplaced the survey invitation and reminder cards before then, it is difficult to understand how this could have produced such a low response rate. The ongoing success of the University of Sydney's Drive Study also argues against a "busy-ness" explanation.

It is also possible that young drivers have become less motivated about surveys as they are often targeted, through schools, to take part in research projects. While this might not have influenced a large number of students, it is also possible that there was some level of interference from the ongoing University of Sydney's Drive Study. It is possible to imagine a scenario where young drivers who are aware of the Drive Study and may have already contributed to it (or decided not to) could misinterpret the MAA survey invitation as being related to the same survey. The survey being conducted by the University of Sydney is web-based, has recruited a large sample through direct and indirect publicity, and addresses similar issues. The possibility that there was some level of confusion between the two surveys cannot be discounted and may be the best potential explanation for the failure of the current survey despite the use of direct recruiting.

The possibility that confusion between an already-running web-based survey and a new web-based survey led to the poor response rate could be assessed, but this would require an additional survey of young people that would best be conducted through schools. If this is the best explanation for the poor response rate then two tentative conclusions can be drawn.

The first is that the failure of this study may not, on its own, be sufficient to argue against the use of an internet delivery method for surveys in road safety in future. Had the poor response rate been related to people visiting the web site and not downloading the software, or downloading the software and not completing the survey, it might have been possible to blame it on some aspect of the survey medium itself. As very few potential participants even went to the web site, it is clear that potential participants either did not have access to the internet, or were unmotivated about the survey. As the second of these is the most likely, the medium may not have been the problem in this case.

Of course, there are other reasons for questioning the use of the internet as a survey tool except in situations where the population of interest has a high level of internet access and where the potential sample bias towards younger, better-off participants is not perceived to be a problem. In the present context, though, the poor response

rate does not appear to be related to the delivery medium, and the internet delivery method does not appear to be less sound than any other method.

The second tentative conclusion is that it may be difficult for participants to differentiate between survey instruments dealing with similar topics when using the same delivery method. This seems reasonable and may apply to survey media other than the internet. It would be reasonable to expect that delivery of two telephone or mail surveys on closely-related topics in a short timeframe would result in some potential participants raising the first survey when asked to take part in the second. In a telephone survey it might be possible to persuade potential participants that the second survey is unrelated to the first, but the current study used a recruitment system where it was not possible to deal with this type of concern.

It is not possible to be certain about why the current study failed to recruit enough participants, but the most likely interpretation appears to be possible interference in the decision to participate by a survey on a similar topic, promoted by similar government organisations, using the same medium.

REFERENCES

Couper MP. 2001. Web surveys: a review of issues and approaches. *Public Opin. Q.*64:464–94

Gosling, S.D., Vazire, S., Srivastava, S., & John O.P. (2004) Should We Trust Web-Based Studies? A Comparative Analysis of Six Preconceptions About Internet Questionnaires. *American Psychologist*, 59, 93-104.

Groves RM, Couper MP. 1998. *Nonresponse in Household Surveys*. New York: Wiley

Halkitis, P.N., Parsons, J.T., & Wilton, L. (2003) Barebacking among gay and bisexual men in New York City: Explanations for the emergence of intentional unsafe behavior. *Archives of Sexual Behaviour*, 32, 351-357.

Kent, R., & Lee, M. (1999) Using the Internet for market research: A study of private trading on the Internet. *Journal of the Market Research Society*, 41, 377-385.

Law and Justice Foundation of New South Wales (2004) http://www.lawfoundation.net.au/information/technology/internet_access.html

Link MW, Oldendick RW. 1999. Call screening: Is it really a problem for survey research? *Public Opin. Q.* 63:577–89

Norman KL, Friedman Z, Norman K, Stevenson R. 2001. Navigational issues in the design of online self-administered questionnaires. *Behav. Inf. Technol.* 20:37–45

Tourangeau, R. (2004) Survey Research and Societal Change. *Annual Review of Psychology*, 55, 775–801

Truell, A.D., Bartlett, J.E., & Alexander, M.W. (2002) Response rate, speed, and completeness: A comparison of internet-based and mail surveys. *Behavior Research Methods, Instruments, & Computers*, 34, 46-49.