

Online and Postal Data Collection Methods: A Comparative Study

Stewart Adam and Heath McDonald, Deakin University

Abstract

Computer-mediated marketing research has been enthusiastically embraced by marketing organisations and those servicing them, for many reasons. While researchers using the Internet (Net) and World Wide Web (Web) in its early years reported benefits such as high response levels, there are now issues in this regard. This paper reports on the outcomes of a probabilistic study involving football club members (subscribers) involving both postal and online (e-mail invitation and HTML Web form) data collection methods. The paper reports differences in both the response pattern and demographic profile of respondents between the groups such as to warrant further examination of the methods used in online marketing research, and to suggest the need for further study.

Introduction

Use of the Internet (Net) as a medium, and the World Wide Web (Web) as an evolving technology has increased productivity (Hanson, 2000) and allowed marketers to get information – both of low quality and high quality – much more quickly. Couper (2000, p. 466) notes "there is a wide array of approaches representing varying levels of quality and cost". Commentaries concerning the quality and cost of online marketing research have been made by many researchers such as Couper (2000), Dommeyer and Moriarty (2000), Weible and Wallace (1998), and Comely (1996). These authors, and others, discuss a number of issues arising from use of marketing research in general, and in online marketing research in particular, including declining response levels and the resulting representativeness of those responding. However, the results are conflicting and the issues warrant further examination.

Antecedent Studies

Traditional marketing research is suffering from falling participation rates (e.g., telephone survey hang-ups, poor mail return rates), rising costs (particularly in the case of personal interviews), respondent fears concerning misuse of personal information, and the time taken to conduct postal surveys (Forrest, 1999). In contrast, a number of claimed advantages are put forward for using online data collection methods in survey research, in particular: lower costs; faster turnaround; higher response levels; lower respondent error; broader stimuli potential through the inclusion of colour, graphics and sound; flexibility in the form of adaptive questioning; and even greater enjoyment (Forrest, 1999; Kehoe and Pitkow, 1996). Most of these advantages may be said to fall into the category of efficiency (Weible and Wallace, 1998). The issue of the fear of misuse of respondent information does not disappear, but rather is exacerbated in the online environment (Cho and LaRose, 1999), and has also contributed to falling confidence in online marketing and research (Australian Privacy Commissioner, 2001).

The costs of online surveys may be lower, depending on how software development costs are treated (Weible and Wallace, 1998; Zadeh, Adam and Deans, 2000). The need for

printing, envelope filling, two-way postage and data entry is removed in the case of online surveys, making them less costly. Online surveys offer faster turnaround times (Adam and Deans, 2000), particularly in the case of email alone (Comely, 1996; Cho and LaRose, 1999; Dommeyer and Moriarty, 2000; Schaefer and Dillman, 1998).

Higher response levels are often claimed for online marketing research involving either e-mail or HTML forms or a combination of the two, over postal surveys (Bachmann, Elfrink and Vazzana, 1996; Bachmann, Elfrink and Vazzana, 2000; Jackson and DeCormier, 1999). A detailed examination of the literature shows that although difficult to state with statistical certainty, over time there has been a decline in response levels for both postal and online surveys. Another benefit of online surveys is a claimed lower respondent error as determined by such measures as the completeness of response and quality of the response to open text box questions (Weible and Wallace 1998). HTML form based questionnaires can easily become too complex for respondents simply because the technology permits intricate rank and rate matrices to be presented (Zadeh, Adam and Deans, 2000).

A Comparative Study

Where antecedent studies have examined the claimed advantages of online surveys, a number of issues have arisen that may have affected the validity of these studies (Couper, 2000). Such issues include the use of small samples, low response levels, non-probability sampling and other matters such as the use of reminders and in some cases the application of incentives that do not allow for ease of direct comparisons. The study we now describe uses a common methodological framework between the two data-collection methods as pertains to such elements as sampling, reminders and incentives, in order to measure similarities and differences in outcomes, such as response levels and response quality.

Hypotheses

A number of hypotheses were developed based on past research findings concerning the speed and efficiency of the data collection methods. Those hypotheses are shown in this section, grouped under the various areas of interest in this paper.

Response Level and Speed

H1_a: Online data collection manifests a higher response level than the postal data collection method.

H1_b: Online data collection manifests a higher response speed than the postal data collection method.

H1_c: Online data collection is less expensive than is the case with postal data collection

Data Quality

H2_a: Online data collection manifests less item-missing non-demographic data than the postal data collection method.

H2_b: Online data collection manifests a greater uniformity in response than the postal data collection method.

Comparability of Resulting Data and Respondents

H3_a: Online data collection manifests a different mean response to scale items as the postal data collection method.

H3_b: Online data collection attracts respondents with a different demographic profile than the postal data collection method.

Methodology

The study employed a 70 item, seven page (postal version) questionnaire entitled "Club Name (withheld) Member Satisfaction Survey". The questionnaire dealt with the satisfaction of members of a professional Australian Rules Football (AFL) club, which we hereafter refer to as AFC to maintain the club's anonymity. In the main, Likert scale items were used, with an 11 point (0 – 10) "poor" to "excellent" scale.

In the year of the study (2001), the AFC claimed over 21,000 paid up members. All members were invited to participate in the survey by way of communication from the club's website. The AFC also maintains an 'opt-in' e-mail list for those members who wish to interact with the club via this medium. The club sent an e-mail to the 3,900 members on this list inviting them to complete an online questionnaire (HTML form), and advising them of a username and password to gain access to the questionnaire. The online responses were automatically recorded in a flat file database.

Additionally, 1,026 members who were not on the email list were randomly selected, and sent a questionnaire through the post to their home address, together with a reply-paid return envelope. An attempt was made to have the online questionnaire appear as close as possible to the printed version, except for the arguably preferred use of radio buttons (Couper, Traugott and Lamais, 2001) and in an arguably preferred 'scrollable' HTML form (Dillman 2000) that disclosed all questions to respondents.

The combined responses were then assessed for their representativeness. Both online and postal survey responses were compared with the overall profile of the full list of AFC members. No statistically significant differences were found between those who returned the survey and the full membership list, suggesting the sampling process was not biased toward a particular demographic or membership type.

Findings and Discussion

Findings from testing the hypotheses set out in the preceding section and related discussion are presented in this section of the paper.

Response Level, Speed, and Efficiency

The response levels are reported in Table 1. The postal data collection method resulted in a response level more than twice that of the online data collection method indicating rejection of $H1_a$ is appropriate. This finding is in line with Dommeyer and Moriarty's (1999) argument that online data collection methods do not result in higher response levels. The postal survey response level is the same as that achieved by Bachmann *et al.* (2000). Moreover, the online response level is similar to the levels achieved by Adam and Deans (2000) (17 per cent) and by Bachmann *et al.* (2000) (19 per cent).

Table 1. Response levels for online and postal data collection methods

	Online	Postal
Sent	3,900	1,026
Undelivered (invalid address)	700 (18%)	6 (<0.5%)
Completed	826	471
Response level	21%	46%
Total cost per usable response	\$1.16	\$4.84
Response speed: Cum. day seven	87%	9%
Response speed: Cum. day 10	91%	53%

The findings in this study support a number of earlier studies in that 75 per cent of the online questionnaires were completed within four days following the e-mail invitation to participate, and 40 per cent were received within the first 24 hours. When discussing response speed it is usual (although perhaps unfair) to present average response times in the case of postal surveys, and to include weekends and other days when post is not delivered. This artificially exaggerates the time taken to respond. The data presented here does not include non-postal delivery days to give a more accurate representation of the relative response speed. On this basis we find that $H1_b$ is supported, since postal data collection took an average 10.8 days compared to 3.9 days in the case of online data collection ($t=58.9$, $p=0.00$).

As shown in Table 1, the total costs of the online survey are lower than those involved in administering the postal survey, thereby supporting $H1_c$.

Data Quality

A comparison was made of the combined item-missing and 'don't know' responses in the postal survey and the number of 'don't know' responses in the case of the online survey.

On the basis of an analysis that excludes responses to demographic questions, $H2_a$ is rejected. On average there was a 23 per cent item-missing data/don't know in the case of the online survey and 19 per cent in the case of the postal survey; which is not a statistically significant difference. Item-missing data levels are high in both cases, as respondents were not required to comment on aspects of the AFC they may not have directly experienced (e.g., the club website and/or family days). Prevention of question skipping in an online questionnaire changes this outcome (Couper, Traugott and Lamias, 2001).

Numerically, there was more item-missing data in the postal surveys in four out of the five demographic questions. However, the level of item-missing data is significantly different

in only two of the five cases, *viz.*, occupation, and life cycle stage. These results suggest acceptance of H2_b, and that there is a higher incidence of item-missing data on personal demographic questions in the case of the postal survey. This finding supports Basi (2000), who suggested that those who complete online questionnaires complete more questions than those completing postal questionnaires.

Because of the ease with which questions can be answered online, it was felt that there might be a higher level of acquiescence from on-line participants. This would present itself as a lower variation in responses, with individuals not using extreme points of the scales. An analysis of both the way individuals answered the questionnaires and of the way the sample as a whole answered specific questions, indicates that in general, those responding online are less likely to use the full scale and are using end and mid-points less than mail respondents. The findings suggest that H1_c should be accepted, and highlight a possible weakness of online data collection.

Comparability of Resulting Data and Respondents

Analysis of responses shows that for the non-demographic questions, the groups are statistically different on 26 of the 65 questions (40 per cent of cases). Even when the online survey was matched against a more demographically equivalent group of postal surveys ($n = 182$), the groups still differed on 22 of the 65 questions (34 per cent). We reject H3_a in that the two groups of respondents do not respond to the questions in the same way using the two data collection methods employed here.

The two groups of respondents also differ in their demographic profile. Income, Age, Years as a Member, Occupation and Lifestyle Category were all significantly different between the two groups of respondents ($p < 0.01$). There are also marked differences within the demographic measures. The online respondents were clearly younger with only the under-nineteen age group statistically the same for both collection methods. The online respondents are more likely to be professionals, and report higher incomes. In terms of income, all categories were significantly different at the $p = 0.05$ level. We, therefore, accept H3_b as the two data collection methods clearly engaged respondents with different demographic profiles.

Conclusion and Future Research

There is a need for further research into measurement error where online data collection methods are employed, and particularly where HTML forms are used. The clear message for managers and practitioners from this research is that online surveys should not be expected to produce equivalent results to the more established methods of data collection, such as the postal survey examined herein.

References

- Adam, S., and Deans, K. R., 2000. Online Business in Australia and New Zealand: Crossing a Chasm. In Treloar, A., and Ellis, A., (Eds.). Proceedings of AUSWEB2K, The Sixth Australian World Wide Web. Cairns: Southern Cross University, 19-34. Available from <http://ausweb.scu.edu.au/aw2k/papers/adam/index.html>, accessed 17 December 2001.
- Australian Privacy Commissioner's website, 2001. Available from <http://privacy.gov.au>, accessed 20 December 2001.
- Bachmann, D., Elfrink, J., and Vazzana, G., 1996. Tracking the Progress of E-mail Versus Snail-Mail. *Marketing Research*. 8(2), 31-35.
- Bachmann, D., Elfrink, J., and Vazzana, G., 2000. E-mail and Snail Mail Face Off in Rematch. *Marketing Research*. 11(4), 10-15.
- Basi, R. K., 1999. WWW Response Rates to Socio-demographic Items. *Journal of the Market Research Society*. 41(4), 397-400.
- Cho, H., and LaRose, R., 1999. Privacy Issues in Internet Surveys. *Social Science Computer Review*. 17(4), 421-434.
- Comely, P., 1996. The Use of the Internet as a Data Collection Method. ESOMAR/EMAC Symposium. Available from <http://Virtualsurveys.com/papers/email.htm>, accessed 20 December 2001.
- Couper, M., Traugott, M. W., and Lamais, M. J., 2001. Web Survey Design and Administration. *Public Opinion Quarterly*. 65(2), 230-253.
- Couper, M. P., 2000. Web Surveys: A Review of Issues and Approaches. *Public Opinion Quarterly*. 64(4), 464-494.
- Dillman, D. A., 2000. *Mail and Internet Surveys: The Tailored Design Method*, New York: John Wiley and Sons.
- Dillman, D. A., 1978. *Mail and Telephone Surveys. The Total Design Method*, New York: John Wiley and Sons.
- Dommeyer, C. J., and Moriarty, E., 2000, Comparing Two Forms of an E-mail Survey: Embedded vs. Attached. *International Journal of Market Research*. 42(1), 39-50.
- Forrest, E., 1999. *Internet Marketing Research*, Sydney: McGraw-Hill.
- Hanson, W., 2000. *Principles of Internet Marketing*, Cincinnati: South-Western College Publishing.
- Jackson, A., and DeCormier, R., 1999. E-mail Survey Response Rates: Targeting Increases Response. *Marketing Intelligence & Planning*. 17(3), 135-140.

Kehoe, C. M., and Pitkow, J. E., 1996. Surveying the Territory: Gvu's Five WWW User Surveys. *The World Wide Web Journal*. 1(3), 77-84. Available at http://cc.gatech.edu/gvu/user_surveys/papers/w3j.html, accessed 12 December 1999.

Schaefer, D. R., and Dillman, D. A., 1998. Development of a Standard E-Mail Methodology: Results of an Experiment. *Public Opinion Quarterly*. 62(3), 378-397.

Weible, R., and Wallace, J., 1998. Cyber Research: The Impact of the Internet on Data Collection. *Market Research*. 10(3), 19-31.

Zadeh, H. S., Adam, S., and Deans, K. R., 2000. A Technical Response to Online Marketing Research Issues. In Treloar, A., and Ellis, A., (Eds.). *Proceedings of AUSWEB2K, The Sixth Australian World Wide Web*. Cairns: Southern Cross University, 409-420. Available from <http://ausweb.scu.edu.au/aw2k/papers/zadeh/index.html>, accessed 17 December 2001.