An Empirical Evaluation of Three Web Survey Design Principles

Benjamin Healey, Terry Macpherson & Bart Kuijten

Web-based surveys present market researchers with a number of advantages in terms of speed and cost compared with other modes of research. However, online questionnaires often suffer from low response rates and technical issues. This prompted Dillman, Totora and Bowker (1998) to put forward a series of respondent-friendly design principles aimed at improving web questionnaire response rates and data quality. While these principles appear useful, empirical evidence validating them is sparse. This paper reports the results of a research project that tested a subset of Dillman et al's (1998) principles relating to: the structure of the first question; the use of graphical symbols conveying point of completion; and the use of double banking for multiple response questions. The results do not lend strong support to the principles tested, and further research is therefore required to clarify whether their adoption can significantly improve response to online questionnaires.

Keywords: Survey Research; Internet Questionnaire Design, Response Rates, Data Quality

Introduction

Use of the Internet as a survey delivery mechanism has increased markedly over recent years. This is not surprising given the relatively swift diffusion of Internet access throughout many countries and the commonly acknowledged advantages of the medium for information dissemination and collection, such as speed and cost reduction. However, much use of the Internet for research purposes occurs in the absence of empirically established methodological guidelines for ensuring the best possible information is obtained. Indeed, in addition to the well known problems of coverage associated with the medium, many online surveys suffer from low response rates or poor data quality for reasons including technical issues or confidentiality concerns (Dillman 2000; Couper 2000).

In an attempt to remedy these issues, Dillman, Totora and Bowker (1998) put forward 11 principles of 'respondent-friendly design', aimed at decreasing "the occurrence of measurement and nonresponse error in [online] surveys" (p.3). Respondent-friendly design is defined as "the construction of web questionnaires in a manner that increases the likelihood that sampled individuals will respond to the survey request, and that they will do so accurately" (p.3). The principles cover a number of aspects of web questionnaire development ranging from whether respondents should be required to answer a question before moving on to another, through to the general formatting for presenting questions on a computer screen. We refer the reader to Dillman, Totora and Bowker (1998) should they wish to examine all 11 principles in detail.

Dillman et al. (1998) clearly state how they developed their principles - through their extensive experience in the development of and research into paper questionnaires, and via review of various web survey implementations and pre-tests. They are just as clear in acknowledging that "most of the principles remain untested in experimental situations" (p.7).

Despite some recent studies (e.g., Couper, Traugott & Lamias 2001; Crawford, Couper & Lamias 2001; Couper, Tourangeau, Conrad & Crawford 2004; DeRouvray & Couper 2002; Heerwegh & Loosveldt 2002), the principles remain only partially explored.

It was in the interest of submitting some of these principles to empirical examination that this study was undertaken.

Specifically, we chose to test three of Dillman et al's (1998) principles:

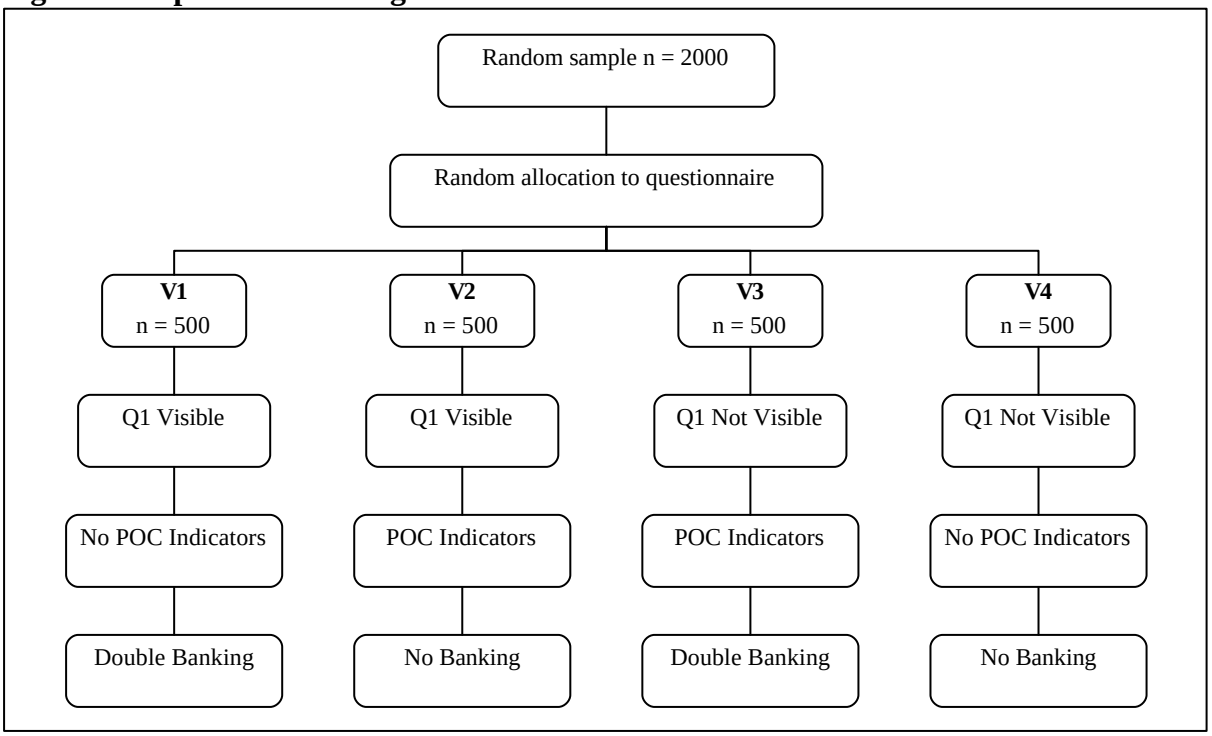
1. Begin the web questionnaire with a question that is fully visible on the first screen of the questionnaire, and will be easily comprehended and answered by all respondents
2. When the number of answer choices exceeds the number that can be displayed on one screen, consider double-banking with appropriate navigational instructions being added
3. Use graphical symbols or words that convey a sense of where the respondent is in the completion progress (i.e., a Point of Completion, or 'POC', Indicator), but avoid ones that require advanced programming

Method and Procedure

Experimental Design

Four versions of a web site questionnaire were constructed (labelled V1-V4 in Figure 1, below). Each version was distinguished by first question format and the presence or absence of a point of completion indicator, both of which were expected to influence unit nonresponse (i.e. whether a survey was completed). The design was not split further to test the double-banking effect as it was expected that this factor would influence item (i.e., whether a specific question was completed) rather than unit nonresponse.

Figure 1: Experimental Design



The Form of Test for each Principle

The following hypotheses were developed to enable the empirical assessment of Dillman et al's principles:

Question Visibility

H1. The use of a fully visible first question will lead to lower dropout rates for a web-site based survey.

Double Banking

H2. The use of double banking will increase respondents' consideration of all possible responses when the number of answers exceeds the amount that can be shown on a computer screen, and subsequently increase the overall number of responses.

H3. When presented response categories in a double banking format, respondents will be more likely to click responses in the left hand column.

H4. When presented response categories in the continuous column down the page, respondents will be more likely to click response categories near the start (top) of the column.

Point of Completion (POC) Indicator

H5. The use of a point of completion indicator will increase completion rates in a web-site based survey.

Each of the hypotheses had an associated metric to be evaluated in the study. Specifically, the test for H1 compared dropouts rates from survey Version 1 with Version 4, and Version 2 with Version 3. The test for H2, H3, and H4 involved recording the overall numbers of responses to the double-banked question along with the spread of responses within those treatments (banking vs. non-banking). Finally, the test for H5 compared completion rates from survey Version 1 with Version 2, and Version 3 with Version 4.

Sampling Frame

Residential email addresses were obtained from listings in the printed version of the New Zealand telephone directories. In the directories, email details are provided along with telephone numbers for a small proportion of residential telephone listings. People must request the listing and pay a small service fee. This frame was chosen as it allowed for the collection of email addresses for people who had specifically consented to their online details being made publicly available. Furthermore, the existence of a list of email addresses meant that clear measures of response could be made. Such calculations are not strictly possible in situations where, for example, responses are elicited via a website banner advertisement or through a newsgroup posting.

A random sample of 2000 addresses was drawn from the list obtained, and each was allocated at random to one of the four experimental treatments outlined earlier.

Survey Delivery

Invitations to complete the survey were sent to each respondent via a standard email merge. A reminder was sent to those who had not responded six days after the initial mailing. Each invitation included a customised URL which contained a unique identifier and led to a web page specific to the respondent's allocated treatment. Hence, by clicking on the URL, or typing it into their browser address bar, each respondent automatically logged themselves into the survey. Where a respondent entered the survey more than once, this was recorded against the database and their responses were over-written with any new ones supplied.

The unique identification numbers assigned to respondents ranged between 0 and 1,000,000 and were randomly generated. Hence, it was unlikely that anyone would have been able to enter the survey simply by guessing an identifier. In cases where an invalid identifier was used to access the survey, the respondent was presented with an error page requesting that they contact the survey coordinators.

The survey asked questions relating to general Internet usage and consisted of a series of web pages linked to an Access database via Active Server Page technology. Five pages were dedicated to the survey itself while another two served to indicate survey completion and communicate errors, respectively. Questions were typically closed in structure and so were presented using checkboxes (for multiple response) or radio buttons (for single response). A small number of open-ended questions were presented using text boxes. Hence, the survey employed the standard form controls supported by all major web browsers.

In addition to the questions, each survey page had a Massey University logo at the top left hand corner and a submit button at the bottom. In Version 2 and Version 3, a point of completion indicator was also present in the top right hand corner of each page. The indicator consisted of an HTML table with a cell shaded depending on the point in the survey. For example, the following was presented to respondents on the second page of the questionnaire:

Figure 2: Point of Completion (POC) Indicator

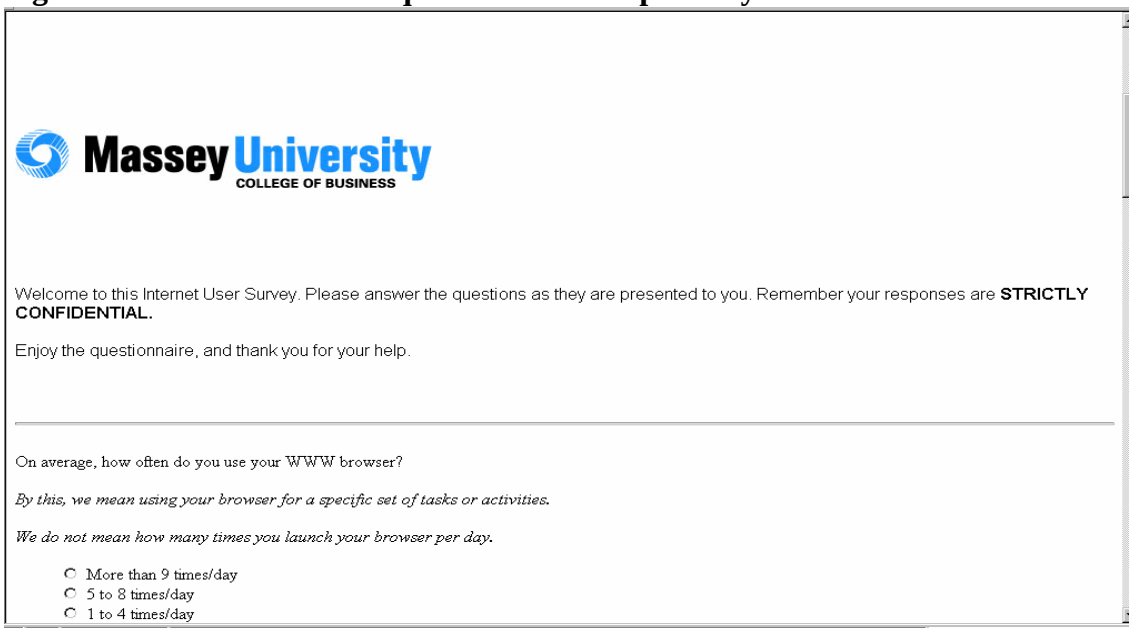
Percent of Survey Completed

0	20	40	60	80

The visibility of the first question in each of the questionnaires was manipulated and tested on common platforms (i.e., Microsoft Internet Explorer and Mozilla browsers) and screen resolution settings to achieve consistency of presentation. Figure 3 shows the first question as presented so that it is not fully visible (i.e., as in Version 3 and Version 4).

Additionally, response categories for double-banked questions were rotated across questionnaire versions to avoid item order bias. That is the left and right banks were swapped in the two double-banking questionnaire treatments (Version 1 and Version 3), while the top half and bottom half questions were swapped in the continuous column treatments (Version 2 and Version 4).

Figure 3: Screenshot of first question set to be partially visible on the first screen



Results and Discussion

Overall Response to the Survey

Forty nine percent of those sent an invitation began the survey. Excluding undeliverable addresses, this represents an initial response rate of 52% (45% fully complete plus 7% partially complete). The proportion of respondents who fully completed the questionnaire was lower, at 45%.

Table 1. Overall Response to the Survey

Response	Respondents	%*	%**
Completed survey	837	42	45
Partially completed survey	134	7	7
Refused	26	1	1
Undeliverable	153	8	-
No Response	850	43	46
Total	*2000*	*100*	*100*

* = Percentage calculated including undeliverables
* = Percentage calculated excluding undeliverables
The percentage columns do not sum to exactly 100 due to rounding error.

The Effect of Question Visibility

As presented in Table 2, the proportion of those that were presented with the first question but did not answer it was not significantly different across the treatments. Similarly, the proportions of those viewing the first page but deciding not to continue with the survey (by clicking through to the second page) were not statistically different. Thus, it would appear

that having the full question visible has little or no effect on whether or not respondents decide to complete the question or continue with the survey.

Table 2. First Question Visibility Results

	First Question Visible	First Question Not Visible
Number starting (presented with first question)	478	493
Proportion not answering the first question	0.09	0.11
Proportion not clicking through to second page	0.08	0.08

Note: All differences insignificant at the 90% confidence level.

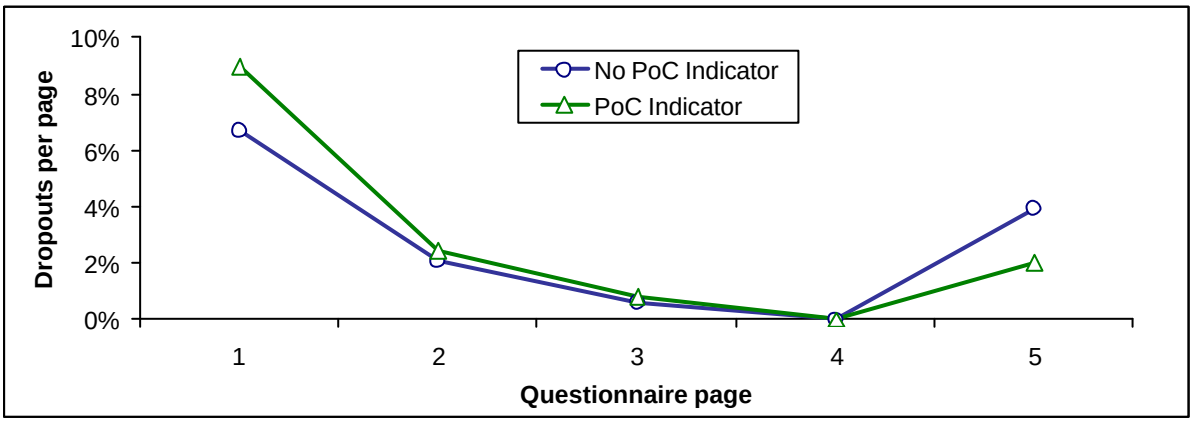
The Effect of a Point of Completion (POC) Indicator

Two measures of response were examined to assess the effect of the POC indicator: overall proportion of dropouts, and dropout rates per page. Table 3 presents the overall dropout rates, while figure 4 presents the breakdown by page.

Table 3. Overall dropout rates by Point of Completion Indicator treatment

	Without POC Indicator	With POC Indicator
Began Survey	481	490
Number not completing	64	70
Percent not completing	13%	14%

Figure 4. Page-by-page dropout rates by Point of Completion Indicator treatment



None of the differences at either the overall or page-by-page level were statistically significant. Hence, no strong support for the efficacy of the POC indicator is provided. Of note, however, is that the per-page dropout rates between the two experimental exposures steadily converge until the last (personal demographic) page of the survey, at which point a larger proportion of the no-indicator group dropped out. The page began with the following statement:

> "ABOUT YOURSELF: So that we can be sure we have a good cross section of people in our survey, would you please answer the following questions about yourself . Remember that all the responses remain STRICTLY CONFIDENTIAL."

Hence, no indication was given in the text that this was the last page of the questionnaire. As such, the results are consistent with the expectation that a POC indicator will reduce dropouts later in the survey, as respondents are more aware that they are nearly at the end. However, further research is required to determine whether this effect can be replicated and the degree to which it occurs in surveys of varying lengths. Furthermore, given that personal demographic questions can be off-putting to some respondents, future investigations should assess the extent to which varying placement of these questions alters dropout patterns.

The Effect of Double Banking

Table 4 shows that the double-banking treatments resulted in only 6% more checked items per respondent than for non-banked treatments. This insignificant effect suggests that the use of double banking in a web survey does not cause respondents to consider the list more thoroughly than if it was presented in a single column.

Table 4. Total Checked Items between Banking Treatments

	Double Banking	No Banking
Total Number of Checked Items	2507	2442
Average	5.7	5.4

Note: Differences insignificant at the 90% confidence level

Previous research relating to the layout of paper questionnaires (Dillman, 2000; Jenkins and Dillman, 1997) suggests that respondents tend to start reading in the upper-left quadrant of the page and tend to ignore instructions or response options on the right hand side of the page. In this instance, there were only 5% more responses in the right rather than left column of the double-banked questions (see Table 5). This small and statistically insignificant difference suggests that respondents in this medium are attending to both sides of the web page equally.

Table 5. Total Checked Items by Column in the Double Banking Treatment

	Left Column	Right Column
Total Number of Checked Items	1224	1283
Average	2.8	2.9

Note: Differences insignificant at the 90% confidence level

Finally, as presented in Table 6, there was a slightly higher number of responses in the bottom quarter of the continuous column treatment than in the top quarter. Given that the treatment was implemented so that the response set was larger than could be viewed on one screen, this effect (although insignificant) runs contrary to the claim by Dillman et al. (1998)

that spreading answers over more than one screen will result in a bias towards the first 8-10 categories.

Table 6. Total Checked Items by Vertical Area in the Non-Banking Treatment

	Top	Bottom
Total Number of Checked Items	1204	1238
Average	2.66	2.73

Note: Differences insignificant at the 90% confidence level

Conclusions and Directions for Further Research

No strong support was found in the present study for the ability of any of the three web survey design principles tested to improve response rates or data quality. While there were instances where the general direction of effect was as hypothesised, none of the differences between experimental exposures were statistically significant. Thus, at this point it appears that those conducting web-based surveys must look to other methods such as incentives and prizes, repeated contacts, visual question elements and reduced questionnaire length to improve response rates and quality. These approaches have been found to positively affect response across a variety of survey modes, including the web (Dillman 2000; Cobanoglu & Cobanoglu 2003; Cook, Heath & Thompson 2000; Deutskens, De Ruyter, Wetzels & Oosterveld 2004; Bosnjak & Tuten 2003; Tuten, Galesic & Bosnjak 2004).

It is possible that the lack of significant effects found here is due to our specific implementation of the principles or the composition of the sample surveyed. Thus, further research exploring alternate executions of the principles in different situations is warranted before any of the principles examined should be discarded. Specifically, alternate forms of the POC indicator, and use of double-banking over a range of response options, should be examined in web questionnaires of varying length. Furthermore, responses from samples sourced over a variety of frames should be examined.

References

Bosnjak M & Tuten TL (2003). Prepaid and promised incentives in Web surveys: An experiment. *Social Science Computer Review*, 21(2), 208-217.

Cobanoglu C & Cobanoglu N (2003). The effect of incentives in Web surveys: Application and ethical considerations. *International Journal of Market Research*, 45(4), 475-488.

Cook C, Heath F & Thompson R (2000). A meta-analysis of response rates in Web or Internet-based surveys. *Educational & Psychological Measurement*, 60(6), 821.

Couper MP (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, 64(4), 464-494.

Couper MP, Tourangeau R, Conrad FG & Crawford SD (2004). What they see is what we get - Response options for Web surveys. *Social Science Computer Review*, 22(1), 111-127.

Couper MP, Traugott MW & Lamias MJ (2001). Web survey design and administration. *Public Opinion Quarterly*, 65(2), 230-253.

Crawford SD, Couper MP & Lamias MJ (2001). Web surveys - perceptions of burden. *Social Science Computer Review*, 19(2), 146-162.

DeRouvray C & Couper MP (2002). Designing a strategy for reducing "No Opinion" responses in Web-based surveys. *Social Science Computer Review*, 20(1), 3-9.

Deutskens E, De Ruyter K, Wetzels M & Oosterveld P (2004). Response rate and response quality of Internet-based surveys: An experimental study. *Marketing Letters*, 15(1), 21-36.

Dillman D (2000). *Mail and Internet Surveys: The Tailored Design Method.*

Dillman D, Tortora R & Bowker D (1998). *Principles for Constructing Web Surveys. SESRC Technical Report 98-50*

Heerwegh D & Loosveldt G (2002). Web surveys: The effect of controlling survey access using PIN numbers. *Social Science Computer Review*, 20(1), 10-21.

Jenkins CR & Dillman DA (1997). Towards a theory of self-administered questionnaire design. *Survey Measurement and Process Quality*, 165-196.

Tuten TL, Galesic M & Bosnjak M (2004). Effects of immediate versus delayed notification of prize draw results on response behavior in Web surveys: An experiment. *Social Science Computer Review*, 22(3), 377-384.

Benjamin Healey and Terry Macpherson are both lecturers and Bart Kuijten is a former postgraduate student in the Department of Marketing, Massey University