

# Do Mail and Web Surveys Provide Same Results?

Katja Lozar Manfreda and Vasja Vehovar<sup>1</sup>

## Abstract

Web surveys are a relatively new data collection mode, used only since the mid 90's. However, they already represent an important part of the contemporary survey industry (<http://websm.org>). Despite their broad implementation, the validity of Web survey mode still needs to be confirmed, especially the convergent validity, e.g., the comparison of Web survey results with results of other survey modes. In our case, such validity shows whether different survey modes can be replaced by each other.

An establishment survey of educational institutions within the RIS (Research on Internet in Slovenia, <http://www.ris.org>) project at the Faculty of Social Sciences, University of Ljubljana, enables the comparison of a Web and a mail survey. Participants were randomly assigned to two experimental groups. The first group received a mail questionnaire (two follow-ups), while the other group was invited to answer a questionnaire on the Web (a mail advance letter with two follow-ups). The advance letters and questionnaires were made as similar as possible. Automated skips were implemented in the Web questionnaire while arrows and text were used for skips in the paper questionnaire. The substantive results, data quality and respondents' satisfaction from both questionnaires were compared. Results show that there are no major differences in substantive responses, however there do exist differences in item non-response with the Web mode having much higher item non-response rates.

---

<sup>1</sup> Faculty of Social Sciences, University of Ljubljana, Slovenia, email: [info@ris.org](mailto:info@ris.org).

## **1 Introduction**

Since they first emerged in the mid 90's Web surveys have become a profitable part of the survey industry (<http://websm.org>). Especially in marketing research they are often regarded as a mode that could replace mail and telephone surveys (e.g., Black, 1998; Cleland, 1996; Hollis, 1999; Jones Thompson, 1999). However, a proof of their validity is needed. By this we have in mind especially the convergent validity (Ferligoj et al., 1995: 88), that is the comparison of Web survey results with results of other already proven to be valid survey methods. Such validity is used when new research methods are developed (Splichal, 1990: 195) and it shows whether different survey modes can be replaced by each other. In this paper we are going to contribute to the research on the validity of the Web survey mode by comparing the results of the Web survey with another survey mode, e.g., with the mail survey mode. At this we would especially like to refer to the possible mode effect error (Groves, 1989: 11) that needs to be established with this new survey mode.

In Web surveys the medium of communication is different from that used in traditional survey modes. While for invitations to a survey mail, telephone or even face-to-face communication is used the questionnaire itself is completed using the Internet. Thus, there are important questions to ask about the impact of the mode of data collection on the respondents' substantive answers and on the quality of data in Web surveys. In this paper we therefore discuss the possible reasons that may lead to different responses in Web surveys in comparison to traditional survey modes, therefore to mode effect. Some of them may have a negative, others a positive influence on data quality. Following this, we are going to present research designs used to study mode effects in general. In the empirical part of this paper we present a study of mode effect where one of the described research designs was used. The study was designed in order to answer a practical question whether a Web survey would give the same results as a mail survey, given all the differences in the way surveys are conducted.

## **2 Specifics of the Web survey mode**

The specifics of the Web survey mode as regards its impact on the quality of data gathered with this data collection method can be discussed in a broader context of the mode effect as one of the measurement survey errors (Groves, 1989). According to the standard definition of the mode effect (Groves, 1989: 12), the mode effect occurs because the Web survey mode is used for collecting survey data from respondents and not some other survey mode.

The fact that the Internet or more specifically, the World Wide Web as one of the Internet services is used for completing the survey questionnaire introduces some specific effects to this type of survey data collection. Some of these effects can be contributed to the fact that Web surveys are *self-administered* surveys and are therefore similar to other self-administered modes. Other effects can be contributed to the *computerization* aspect of the Web survey mode and are therefore similar to other computer-assisted survey modes. Finally, there are some effects that may be *specific only for the Web survey* mode and not for other modes.

In comparison to mail surveys, the Web survey mode may introduce specifics which occur due to the *computerized questionnaire* used in Web and the paper-and-pencil questionnaire used in mail surveys. In general, computerized questionnaires may influence the respondents' answers due to the following advantages in comparison to the paper-and-pencil questionnaires (de Leeuw and Nicholls, 1996: 3.2)<sup>2</sup>:

- Absence of routing errors;
- Immediate data checks;
- New possibilities of formulating questions: randomized order of questions or items, eliminating order and recency effects; assisting in interactive coding of open-ended questions; employing question formats such as drawing line lengths as in psychophysical scaling.

These advantages represent also the characteristics of Web surveys and can decrease the mode effect error if the Web survey software is correctly programmed.

In comparison to telephone surveys (which are today mostly computer assisted), the Web can introduce specifics especially because Web surveys are *self-administered* while telephone surveys are interviewer-administered. In general, the advantages of interviewer-administered surveys for data quality lie in the fact that well-trained interviewers can explain unclear terms to the respondents, keep them motivated, assure the confidentiality of their answers, and probe incomplete and inadequate responses. They serve as intermediaries between the researcher and the respondents while in self-administered surveys there is no such intermediary (Couper, 2001). This absence of the interviewer can increase the mode effect error in Web surveys. On the other hand, self-administration can also decrease the mode effect error. It allows respondents to answer the questionnaire when they want, where they want and at the pace they want, a freedom which might increase their

---

<sup>2</sup> Additional advantages, such as elimination of the separate data entry phase and recording information on the questionnaire completion process have also been identified by de Leeuw and Nicholls (1996). However, they have a particular impact on issues such as time and money, therefore they are not of our concern when discussing the mode effect.

motivation to accurately answer survey questions. In addition, it may reduce the problem of social desirability bias<sup>3</sup> due to the lack of interviewer presence.

Besides the above specifics which Web surveys share with all computerized self-administered questionnaires, there are other specifics that are characteristic only for this mode. We believe that there are two main reasons why the Web survey may result in different respondent's answers than other survey modes:

1. different 'channel' capacity of the Web (and Internet in general) and
2. context effect of the Web (and Internet) usage.

## **2.1 Channel capacity of the Web mode**

By channel capacity of the Web mode we refer to ways of presenting information to respondents that are not present in other survey modes, either because they are not possible or are too expensive and time-consuming to develop and duplicate. Web survey instruments actually no longer consist only (or primarily) verbal features (words and numbers) but can make use also of rich visual features (Couper, 2001). These features include the use of multiple colors, special navigational features (e.g., indexes, tables of contents, progress indicators), still and moving images, animations, line drawings, sound, etc. These can be added to traditionally presented survey questions in order to illustrate them or simply to motivate respondents. In addition, survey questions can be presented differently, e.g., looking and functioning differently. For example, there can be questions with drop-down menus which do not appear in any other survey mode.

In general, these features are used to motivate respondents to answer survey questions more 'accurately' and therefore to increase the data quality. However, these features can also result in larger coverage error: some respondents cannot access a Web questionnaire using advanced programming features because the equipment they use does not support this. In addition, non-response can increase since respondents may lose patience while waiting for such a questionnaire to download. The stress associated with animation by visual features, sound, etc., on the one hand, and equipment that does not support them properly on the other hand, may also decrease the motivation to participate. Empirical examples of situations where advanced graphic design had a negative impact on participation in Web surveys are presented by Dillman et al. (1998) and Lozar Manfreda et al. (2002).

Beside this, new ways of presenting survey material can result in different mode effect, and this is our particular concern in this paper. In general, Web surveys present questions using different sensory modalities and respondents use

---

<sup>3</sup> Social desirability bias is defined as a tendency of individuals to give survey answers that present them in a better light (Robinson et al., 1991: 17).

different means to provide responses. As a result differences in their responses may occur (Lozar Manfreda et al., 2002; Smith, 1999).

## 2.2 Context effect of the Web mode

The situational variables surrounding the activities of answering a Web questionnaire may be different from those of reading and answering paper questionnaires or listening and answering to interviewers on the phone or in a face-to-face situation. The global computer network, i.e. the Internet, is used for questionnaire completion, a situation different from the local context of questionnaire completion in other survey modes. Several aspects of this context can introduce mode effects:

- The visual presence of the computer;
- The specific task of completing a Web questionnaire;
- Survey participation as a specific social interaction on the Internet.

### 2.2.1 Presence of the computer

At this we do not refer to the technical aspects of computerized questionnaires as discussed in the introduction to this section, but rather to the effects that the presence of the computer has on the actions of the respondents. For less experienced Internet users, the computer itself can have a negative effect because of their *limited capability to answer survey questions on a computer screen* with a keyboard and a mouse. It has been shown (comparing the performances of individuals when working with paper-and-pencil or computer) that people give more accurate responses and complete the task faster when working with paper-and-pencil (Hansen and Haas, 1988). In addition, the computer screen may be smaller than paper questionnaire resulting in the eventual loss of context or sense of the question content (Sperry et al., 1998: 352).

Besides the limited capability of answering survey questions, the use of a computer can also have an effect on answers due to *different perceptions of privacy*. Whether this effect is positive or negative is yet to be established. As hypothesized for computerized surveys in general, unfamiliarity with computers can lead to a 'big brother' effect, leading to more refusals and more socially desirable answers (de Leeuw and Nicholls, 1996: 3.9). This can be even more true for Web surveys, owing to the novelty of the survey mode, the larger possibility of combining data, and additional information collected without the explicit agreement of the respondent (Lozar Manfreda, 2001: 135). On the other hand, using a computer could also lead to the expectation of greater privacy by respondents, since responses are typed directly into the computer and cannot be

read by anyone who happens to find the questionnaire (de Leeuw and Nicholls, 1996: 3.9). This issue is further discussed below (section on specific social interaction on the Internet).

### 2.2.2 Specific task of completing a Web questionnaire

Due to the convenience of self-administration, it is very possible that Web survey respondents perform other tasks using their personal computer simultaneously alongside the task of answering the questionnaire. While they have a questionnaire open in one window, they may also download, read and sent emails, be engaged in simultaneous online conversations (for example using messengers, such as Yahoo messenger), look for information in another browser window or even perform other non-Internet-related activities. Because of this lack of concentration and care, respondents may be less careful when completing the Web questionnaire, resulting in larger item non-response, less elaborate answers to open-ended questions and less accurate responses to demanding questions.

In addition, individuals treat the text on the Internet differently than they treat printed text. On the Web people tend to read the text more quickly, to scan it with their fingers on the mouse ready to click on through to the next item (Bauman et al., 2000; Gräf, 2002: 79). They are also more impatient, more fastidious than off-line readers (Internet Rogator, 1998). Applying that to Web questionnaires suggests that question texts may not be as carefully read as they would have been on paper or when heard during an interview.

### 2.2.3 Specific social interaction on the Internet

Social interaction on the Internet differs from the offline social interaction. In general, “... the Net is an intricate and chaotic environment and its rules and parameters seem different to real life, so our behavior as users are likely to be different too” (Aspinall et al., 2000: 230). At this stage we are not going to discuss all differences in detail. We are only going to highlight those specifics that are related to survey data collection on the Internet. In particular, we will focus on those aspects that might affect social desirability bias and the reliability of answers in Web surveys.

One of the often-stated advantages of Web surveys (e.g., Aspinall et al., 2000: 237; Coomber, 1997) is their possibility to *reduce the social desirability bias*, owing to self-administration and the special context of the Internet. The latter can occur for several reasons:

- Online, the conventional norms of behavior and concerns about social presentation and judgment are greatly reduced or entirely absent (Joinson,

1998). According to the reduced social cues theory (Sassenberg and Kreutz, 2002: 230-231), in this case less information is given on social context. Communication situations are considered less as social interactions. Therefore social norms are considered less important, social cues less salient and as having less effect on behavior. For these reasons, Web surveys are more likely to endorse sensitive or non-socially-desirable behavior.

- Persons surfing the Internet often experience a loss of social control. This might lead to less inhibition in answering personal sensitive questions or in admitting socially undesirable behavior (Hertel et al., 2002: 139-140).
- On the Internet, people may have the impression that their answers cannot be associated with their persona, owing to greater anonymity and lower possibility of being identified (Sassenberg and Kreutz, 2002: 231-233). In Web surveys, this may allow them to express less socially desirable behavior.

While the above discussions actually praise the (perceived) anonymity of Web questionnaires, this can also have a negative effect. According to Hertel et al. (2002: 140) the (perceived) anonymity in Web questionnaires can lead to *lower data reliability* particularly when participants are not highly motivated, or enjoy playing with different identities. In addition, anonymity is actually not assured and those concerned about security on the Web may produce *higher non-response or less honest responses* on sensitive topics (Couper, 2000: 474).

### 3 Measuring mode effect

The effect of the Web mode of data collection can be approached by two questions<sup>4</sup>:

1. Would a Web survey give the same results as another survey (mail or telephone), given all the differences in the way these surveys are conducted?
2. What is the marginal effect of the Web mode of data collection on survey statistics, assuming all other attributes of the design are the same as in the survey (mail or telephone) to which the Web survey is compared?

The first research question is more practical and easily measurable, therefore it is actually often addressed by Web survey methodologists, as shown below. The second research question tries to identify inherent properties of the Web mode which might produce differences between the modes that can be attributed merely to the communication channel and not to survey design characteristics. Due to the

---

<sup>4</sup> These questions were defined by Groves (1989: 502) when studying the mode effect of telephone in comparison to face-to-face surveys.

more complicated research design (as described below) there has been little research produced in order to answer this question so far.

When answering the first question, split sample experimental design can be used. In these cases each randomly chosen group of participants is administered one survey mode. For each survey mode, the procedures that are known to be the best for that situation tend to be chosen. Therefore, different procedures are usually chosen for one mode than for another. In this case, the differences in results can be caused by the medium of communication used for the questionnaire completion (i.e. by mode effect) or by any other possible causes:

1. If a Web and another survey mode are used to compare results on a sample from the general population, the coverage error may affect the results.
2. If the Web mode allows a larger sample than another, more expensive mode, the sampling error may introduce additional differences.
3. Even if a sample is taken from the same population with high Internet coverage and the same sample sizes are used, different non-response errors may occur in both modes additionally contributing to the differences.

When answering the second question, i.e. the question on marginal effect, panel design can be used. First, the respondents are administered one survey mode. Those who answer (and if necessary provide contact information for the next panel wave) are later administered another survey mode. Only those who participate in both waves are compared. In this case, the differences cannot occur because of coverage, sampling or non-response error since the same individuals are compared. Most likely, they can be attributed to the communication mode itself. However, also in this case, additional differences can occur for other possible reasons:

1. The experience from the survey in the first wave may alter the response behavior in the second wave.
2. The first survey may lead respondents to change their actual behavior related to the survey measures.
3. Actual differences can occur due to some events between the two waves.

In literature on Web surveys, usually the first research question is addressed: mode effect is measured by comparing results from Web and another survey mode while two different methodologies are used on comparable samples (samples extracted from the same population in an identical way). Ideally, empirical studies try to ensure a similarity of groups before collecting data using split sample design. By taking experimental groups of the same size from the same population, the coverage, sampling and error due to respondents are controlled. By using a questionnaire with the same questions and possible answers in the same order and of a similar graphical design, the error due to the survey instrument is controlled. By implementing procedures for achieving a response as high as possible, non-



response error is controlled. However, especially the latter is, for now, difficult to obtain, since Web surveys usually result in lower response rates than other survey modes (Lozar Manfreda, 2001: 120). Therefore, when differences in results are discussed, they can never be attributed to the mode effect alone, but often also to differences in non-response error. In literature, such studies using split sample design are reported by Bates (2001), Dillman et al. (2001), Jones and Pitt (1999), Kwak and Radler (1999), McNeish (2001), and Wygant and Lindorf (1999).

Frequently quasi-experimental designs are used to address the first research question. In this case the similarity of groups before collecting data is not assured and the design is called confounded design (Grisaffe 2000). They ensure the similarity of the two compared groups after data collection through thoughtful use of sophisticated weighting algorithms. However, if differences occur between the two modes, they can only conditionally be attributed to the mode effect itself. In addition to the possible sources of differences occurring in split sample designs, they can also be attributed to the selection of variables used in weighting algorithms. It is possible that not all key variables that differentiate between the compared groups have been used. Studies using weighting procedures in order to compare results from different modes are reported by Dietrich (1999), Flemming and Sonner (1999), Terhanian and Black (1999), and Vehovar et al. (1999).

Regarding the second research questions, i.e. the marginal effect of the Web mode, not much research has been performed so far. Although panel Web surveys are often used, the advantage of panel design in measuring the marginal effect of the Web mode has rarely been taken. One such study comparing the responses of the same individuals to a Web and a telephone questionnaire and to a Web and a mail questionnaire (therefore individuals were measured twice) has been performed by Vehovar and Lozar Manfreda (2002).

In this paper we concentrate on the practical question whether a Web survey can provide the same results as another survey mode, specifically the mail survey mode. We address the first research question regarding the mode effect which is a practical survey problem. It refers to the problem faced by a researcher choosing a mode of data collection for a particular topic within a given population. The sample is split into two (or more) randomly chosen sub-samples to which two (or more) different survey modes (with the most suited rules and procedures for administering a particular survey mode) are administered. Although with this type of research design the eventual differences in results cannot be contributed merely to the mode effect (as discussed above), this is the typical decision situation that survey researchers are faced with in practice.

## 4 Web versus mail surveys: A case study of mode differences

Within the Research on Internet in Slovenia (RIS) project<sup>5</sup> a study using split sample design was intended to test whether a Web survey would give the same results as a mail survey. The study was implemented within the 2000/2001 annual RIS survey of primary and secondary school institutions<sup>6</sup>. The 1999 study revealed that the Internet penetration rate among Slovenian school institutions is high (over 95% of primary and secondary schools had access to the Internet); therefore this population seemed appropriate to test the mode effect of a Web survey.

### 4.1 Methodology

#### 4.1.1 Data description

Two thirds of all Slovenian secondary and primary schools (the sample frame was obtained from the Ministry of Education, Science and Sport) were randomly assigned to two experimental groups, each of the size  $n=200$ <sup>7</sup>.

1. Mail group: Traditional mail survey with two-follow-ups: a reminder and a letter with a replacement questionnaire (TDM procedure; Dillman, 1978);
2. Web group: Mail advance letter with an invitation to a Web questionnaire and two similar follow-up mail letters.

The design of this experiment was prepared in order to eliminate as much as possible the differences in the design of the two surveys, with the mode remaining the only difference. Specific design measures to eliminate or control other survey errors were:

- The sampling error was controlled by using systematic random sampling choosing two samples of the same size.
- Sampling from the same population allowed controlling the problem of measurement error due to respondents: if it occurred, it was similar in both surveys since the respondents were from the same target population. The

---

<sup>5</sup> Conducted at the Faculty of Social Sciences, University of Ljubljana.

See <http://www.ris.org>.

<sup>6</sup> Actually also kindergartens, music schools and student dormitories were surveyed; however these were not included in the described experiment.

<sup>7</sup> One third of the institutions were assigned to the third experimental group (also of the size  $n=200$ ) which was used to test the effectiveness of the Web survey mode within a mixed-mode design. This group was primarily administered a Web questionnaire; however in the follow-ups respondents could also choose a replacement paper questionnaire. Results of this experiment are discussed in Lozar Manfreda et al. (2001).

schools' principal or person responsible for the computer technology at the school was asked to answer the questionnaire.

- The measurement error due to the survey instrument was controlled by using similar paper and Web questionnaires. The same questions and answers, and the same question order were used. In addition, in order to mimic the paper questionnaire, the graphical design of the Web questionnaire was very simple. There were several questions on one HTML page and a new page appeared only after automated skips. An ID number was printed in the lower right corner of the paper questionnaire. For the Web questionnaire, respondents needed to enter their ID number at the survey introductory Web site in order to access it.
- By taking units from the same target population with high Internet penetration, the problem of coverage error was eliminated. From the previous survey (1999) we knew that over 95% of schools had access to the Internet; therefore no special instructions were given in case a school would not have access to the Internet. We assumed that in the event that the school did not have access to the Internet the respondents in the Web group would ignore the quest for participation or try to answer the questions from another location, not from the school. It turned out that this was not a problem at all. Every school that participated in the mail survey had access to the Internet. In the Web group three respondents answered that they still did not have access to the Internet, but they planned it in the near future. Nevertheless, they obviously had the possibility to answer the questionnaire from another location where access to the Internet was available.

What we could not control with the design was the non-response error. Despite the same actions for non-response conversion (i.e. same use of follow-up mailings), the response was higher in the mail group: 89% versus 77%. In this case the differences in results might therefore occur due to mode effect, but also due to different non-response error, i.e. different people (institutions) responding to the mail and the Web questionnaire. However, the composition of the samples of respondents in terms of the type of institution and function of the persons within the institution who answered the questionnaire was the same. Therefore we assume that the differential non-response error is of minor concern at this study.

#### **4.1.2 Analysis**

A particular survey item (and not respondent) is of interest here. Characteristics of these items, such as scale type (quantitative, qualitative), closed versus open questions and type of information measured (attitude, behavior, fact) are taken as predictors of possible mode effect using bivariate analysis. Mode effect is measured with the substantial difference between the mail and the Web responses, i.e. with difference in responses as regards their content. For quantitative variables

the mean across respondents from the mail survey and the mean across respondents from the Web survey were compared. For qualitative variables, the percentage of interest was compared. The independent samples t test was used to test the statistical significance of the difference if estimates were to be given from both surveys.

In addition, item non-response in both surveys was compared. For each survey the item percentage of cases with an invalid response (no response, response 'don't know' or not usable response) was calculated and compared across modes. Also in this case, the characteristics of the items were taken as predictors of possible differences in item non-response using bivariate analysis.

A meta-analysis across survey variables was performed in order to integrate the findings from individual survey variables. As already explained, the variable t test testing the null hypothesis of no difference in substantive responses was performed for each individual item. The Stouffer combined test<sup>8</sup> was then used to statistically summarize the results of these independent t tests of the same hypothesis. The purpose of this test is to obtain a summary overall test of the hypothesis (Wolf, 1986: 18).

In the studies on the quality of survey data, multiple classification analysis (Andrews et al., 1973) is also usually performed across survey variables in order to test the interaction effects of different types of compared survey variables (for example, Hlebec, 1999). Unfortunately, in this study this type of meta-analysis cannot be performed due to the lack of possible combinations across different characteristics of compared survey variables (for example, all nominal variables demand factual data, therefore interaction among the type of information asked and type of measurement scale cannot be measured).

## 4.2 Differences in results

### 4.2.1 Substantial differences

For most of the questions from the studied survey among Slovenian secondary and primary schools there is no difference in the respondents' answers. Nevertheless, a

---

<sup>8</sup> The Stouffer combined test is used in meta-analysis as a method for combining results of independent studies (Wolf, 1986: 20). It is calculated using the following expression:

$$Z_c = \frac{\sum z}{\sqrt{N}},$$

where z are statistics associated with each independent test of the effect measured and N number of tests.

statistically significant difference (at  $p < 0.05$ ) occurs in 29% of items (31 out of 108).

There are some differences across the types of compared variables:

1. Qualitative vs. quantitative variables: A statistically significant difference occurs for all qualitative (nominal or ordinal) variables, while only for 25% of quantitative (interval or ratio) variables.
2. Type of information asked: Among quantitative variables, the statistically significant difference occurs more often for behavioral and attitudinal questions while less often for factual questions. It occurs the least often for quantitative variables which ask about behavior in the future (e.g., the frequency of usage in 12 months, or ‘Will Internet replace other communication media when communicating with ...?’).
3. Open vs. closed questions. There are 8 open questions, which ask for an exact number, such as the number of employees, pupils, computers, etc. Statistically significant difference occurs in only 1 of these questions. On the other hand, statistically significant differences occur in 30% of the closed questions.

**Table 1:** Statistically significant differences across survey items in the mail – Web comparison.

	<b>Number of items</b>	<b>% (number) of items with statistically significant difference (<math>p &lt; 0.05</math>)</b>	<b>% (number) of items with difference not statistically significant</b>
<b>Total</b>	108	29% (31)	71% (77)
<b>Quantitative variables (mean compared)</b>	103	25% (26)	75% (77)
<b>Qualitative variables (% of interest compared)</b>	5	100% (5)	0% (0)
<b>Attitudinal questions</b>	24 quantitative	25% (6)	75% (18)
<b>Behavioral questions</b>	64 quantitative	28% (18)	72% (46)
<b>Factual questions</b>	10 quantitative	20% (2)	80% (8)
<b>Factual questions</b>	5 qualitative	100% (5)	0% (0)
<b>Predicting behavior</b>	5 quantitative	0% (0)	5% (5)
<b>Open questions</b>	8	12% (1)	88% (7)
<b>Closed questions</b>	100	30% (30)	70% (30)

The direction of differences is variable and has no pattern. In some cases mail respondents report more frequent Internet usage or greater importance of individual usage, while in other less frequent or lower importance.

The Stouffer combined test ( $z=0.76$ ) shows that on average the difference in estimates from the mail and the Web survey is not statistically significant (at  $p < 0.05$ ). Nevertheless, some variation across the types of survey variables can be observed, as showed above.

### 4.2.2 Item non-response

For every survey item, the percentage of respondents that did not answer that item, answered with ‘don’t know’, or gave an unusable answer (for example, ‘almost all’ where an exact number was required) was calculated. The average item non-response for items in the mail survey is 8%, while in the Web survey it amounts to 17%. The difference is statistically significant at  $p < 0.0005$ .

In both surveys the item non-response occurs more often for quantitative than for qualitative variables. It also occurs more often for attitudinal and behavioral questions than for factual questions. For the Web survey, it occurs often also for questions asking as regards the predictions of behavior.

In the mail survey, the item non-response occurs more often for open than for closed questions. However, there is no difference among open and closed questions in the Web survey, although in both cases the item non-response is rather high.

**Table 2:** Item non-response across survey items in the mail – Web comparison.

	Number of items	Item non-response (% of cases with no answer)	
		Mail	Web
<b>Total</b>	108	8%	17%
<b>Quantitative variables (mean compared)</b>	103	8%	17%
<b>Qualitative variables (% of interest compared)</b>	5	4%	5%
<b>Attitudinal questions</b>	24	8%	18%
<b>Behavioral questions</b>	64	8%	17%
<b>Factual questions</b>	15	1%	14%
<b>Predicting behavior</b>	5	3%	17%
<b>Open questions</b>	8	14%	18%
<b>Closed questions</b>	100	7%	17%

In general the item non-response is therefore larger for the Web questionnaire. This is consistent with the findings of two other studies (Bates, 2001; Vehovar et al., 2001). In all these cases, no controls for the item non-response were used. Without this programming feature, it looks like the item non-response is larger in Web than in paper questionnaires. In our case reminders were not used since only higher versions of browsers would support them (those that support Java Scripts) and we were not sure whether our respondents use them.

### 4.2.3 Respondent satisfaction

At the end of the questionnaire we asked the respondents to evaluate the survey according to several criteria (subject, difficulty, length, graphics) on a scale from 1

(not adequate) to 5 (excellent). The average score on these items was the same in both groups, except for the graphic layout of the questionnaire. The Web respondents liked the layout of the questionnaire more than the mail respondents. In addition, respondents from the Web also gave an overall higher evaluation of the survey.

### **4.3 Overview of results**

Regarding the differences in Web and mail surveys measured with the described study, we can summarize the following:

- In most cases the mode did not influence substantial answers of the respondents. This suggests that despite the slightly lower response rate in the Web survey, the Web survey presents an attractive alternative to the mail survey.
- On the other hand, there is a difference in item non-response, especially for closed-ended questions between the two modes. Item non-response was statistically significantly larger for the Web questionnaire. It seems as if the respondents are less committed to completely answering the Web than the paper questionnaire, unless controls for the item non-response are used which force them to answer.

## **5 Conclusions**

This paper deals with the mode effect of Web surveys. First, some specifics of the Web survey mode were presented in relation to their possible influences on mode effect. Then, ways of measuring the mode effect of Web surveys were discussed. Finally, an empirical study aimed at answering practical research questions whether Web and mail surveys provide the same results was presented.

The presented study therefore tested whether Web and mail surveys would give the same results if two samples of identical size from the same population are surveyed using the same procedures (except for the mode of questionnaire completion). The design permitted control for all types of survey errors except for the non-response error: non-respondents to the Web mode may be different from non-respondents to the mail mode. Since response rates in both cases were relatively high, with a somewhat lower response for the Web survey, (however with an identical structure of respondents), we assumed that the eventual differences in results could be mostly attributed to the mode effect. Another type of research design, i.e. a panel research design measuring the same individuals with two or more survey modes, would allow to extract mode effect errors with more certainty, as showed in Vehovar and Lozar Manfreda (2002) where the

marginal effect of the Web mode in comparison to mail and telephone mode was measured.

Nevertheless, with the described study we showed that not many differences in substantive responses occurred. This might be the effect of the self-administration context of both modes: in both cases the respondents had to answer the questionnaire by themselves without any help from an interviewer. Research from other studies shows that mode effects occur more often when Web surveys are compared to some interviewer-administered surveys (for example, Vehovar and Lozar Manfreda, 2002; Willke et al., 1999; Wydra 1999).

In contrast to the minimal differences in substantive responses, important differences occurred as regards the item non-response. The item non-response was higher for the Web survey. This result is consistent with the findings of two other studies (Bates, 2001, Vehovar et al., 2001). In all these cases, no controls for item non-response were used. Without this programming feature, it seems as if the respondents are less committed to completely answering the Web than the paper questionnaire. This may be the effect of the context in which Web questionnaires are completed. During the Web questionnaire completion, it may be possible that Web survey respondents perform other tasks using their personal computer simultaneously with the task of answering the questionnaire. Because of this lack of concentration and care, respondents may be less careful when completing the Web questionnaire, resulting in larger item non-response.

A technical solution to the problem of item non-response in computerized questionnaires are controls for item non-response which were however not used in our study. In principle, forcing respondents to answer questions properly can prevent any item non-responses (or also inconsistent responses). However, respondents' frustration associated with these requirements is likely to lead to premature terminations (Dillman, 2000; Dillman et al., 1998; Zukerberg et al., 1999) and frustrations (Abraham et al., 1998: 836). In particular, forced reminders, i.e. forcing the completion of questions before allowing respondents to move on can lead to drop-outs (Comley, 2000: 331), or result in respondents answering the questions without due consideration. Soft reminders - when the program allows one to proceed even if the error was not corrected - is probably a reasonable alternative to hard (forced) reminders (Zukerberg et al., 1999). Alternatively, 'prefer not to answer' and/or 'don't know' categories can be provided for every item when hard controls are used (Dillman et al., 1999). Unfortunately, no experimental research comparing the effect of reminders for item non-response in Web surveys has been performed so far. In future, the experiments should be directed to the effect of different types of reminders. An optimum balance between their positive effect of decreasing the item non-response and the negative effect of decreasing the respondents' satisfaction, increasing partial non-response (i.e. premature abandonment of the Web questionnaire without finishing it), and decreasing other aspects of data quality is needed.



In this paper we concentrated on the survey error stimulated by the mode of survey data collection. However, each practical decision on the mode of data collection always includes also their costs. For this particular empirical study we showed elsewhere that the costs for the Web mode were about one half of the costs of the mail mode (Lozar Manfreda et al., 2001) thus strongly favoring the Web mode. Similarly, for another establishment survey, e.g., survey of Slovenian business companies in 1999, we showed (Vehovar et al., 2001) that with respect to costs-errors analysis, the Web survey mode was superior in comparison to mail mode. The Web questionnaire combined with traditional mail invitation produced no extreme errors in the estimates when compared to a conventional mail survey, although some larger error was observed when compared to a telephone survey. However, even in the latter case, when costs and errors are taken into account, the higher costs of telephone surveys with long questionnaires resulting in limitations in samples size may favor Web surveys. There, the non-sampling errors are somehow larger, however, this can be compensated with potentially larger sample sizes (Vehovar et al., 2001).

## References

- [1] Abraham, S.Y., Miller Steiger, D., and Sullivan, C. (1998): Electronic and mail self-administered questionnaires: A comparative assessment of use among elite Populations. *Proceedings of the Section on Survey Research Methods*. Alexandria: American Statistical Association, 833-841.
- [2] Aspinall, E., Batey, R., and Branthwaite, A. (2000): What gets through the Net? Net use in brand communication and research. *Proceedings of the ESOMAR Worldwide Internet Conference Net Effects*, April 10-12, 2000, Dublin. Amsterdam: ESOMAR Publication Series - Volume 237.
- [3] Bates, N. (2001): Internet versus mail as a data collection methodology from a high-coverage population. Paper presented at 2001 AAPOR Annual Conference, Montreal, Quebec, May 17-20, 2001.
- [4] Bauman, S., Jobity, N., Airey, J., and Atak, H. (2000): Invites, intros and incentives: Lessons from a Web survey. Paper presented at 2000 AAPOR Annual Conference, Portland, Oregon, May 18-21, 2000. Available from Wintellitech. <http://www.wintellitech.com/> (November 21, 2001).
- [5] Black, G.S. (1998): Internet surveys – A replacement technology. Paper presented at 1998 AAPOR Annual Conference, St. Louis, May 1998.
- [6] Cleland, K. (1996): On-line research costs about half that of traditional methods. *Advertising Age's Business Marketing*, **81**, B8-9.

- [7] Comley, P. (2000): Pop-up surveys. What works, what doesn't work and what will work in the future. *Proceedings of the ESOMAR Worldwide Internet Conference Net Effects 3*. April 10-12, 2000, Dublin. Amsterdam: ESOMAR Publication Series, **237**. (November 26, 2001).  
[http://www.virtualsurveys.com/papers/popup\\_paper.htm](http://www.virtualsurveys.com/papers/popup_paper.htm)
- [8] Coomber, R. (1997): Using the Internet for survey research. *Sociological Research Online*, **2**, <http://www.socresonline.org.uk/2/2/2.html> (Feb 15, 2002).
- [9] Couper, M.P. (2000): Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, **64**, 464-494.
- [10] Couper, M.P. (2001): The promises and perils of Web surveys. In A. Westlake et al. (Eds.): *The Challenge of the Internet*. London: Association for Survey Computing.
- [11] de Leeuw, E.D. and Nicholls, W. II (1996): Technological innovations in data collection: Acceptance, data quality and costs. *Sociological Research Online*, **1**. (November 21, 2001).  
<http://www.socresonline.org.uk/1/4/leeuw.html>
- [12] Dietrich, D. (1999): Consumer segmentation results using RDD vs. mail panel vs. Internet sample... There is a difference! Paper presented at the 20th Annual Marketing Research Conference, organized by AMA, San Diego, California, September 26-29, 1999.
- [13] Dillman, D.A. (1978): *Mail and Telephone Surveys. The Total Design Method*. New York: John Wiley & Sons.
- [14] Dillman, D.A. (2000): *Mail and Internet Surveys. The Tailored Design Method*. New York: John Wiley & Sons.
- [15] Dillman, D.A., Carley-Baxter, L., and Jackson, A. (1999): Skip pattern compliance in three test forms: A theoretical and empirical evaluation. *SESRC Technical Report 99-01*, Pullman, Washington.
- [16] Dillman, D.A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., and Berck, J. (2001): Response rate and measurement differences in mixed mode surveys using mail, telephone, interactive voice response and the Internet. Paper presented at 2001 AAPOR Annual Conference, Montreal, Quebec, May 17-20, 2001.
- [17] Dillman, D.A., Tortora, R.D., and Bowker, D. (1998): Influence of plain vs. fancy design on response rates for Web surveys. *1998 Proceedings of the Section on Survey Research Methods*. Alexandria: American Statistical Association. (November 26, 2001)  
<http://survey.sesrc.wsu.edu/dillman/papers/asa98ppr.pdf>
- [18] Ferligoj, A., Leskošek, K., and Kogovšek, T. (1995): Zanesljivost in veljavnost merjenja (Validity and reliability of measurement). *Metodološki zvezki*, **11**. Ljubljana: Faculty of Social Sciences.

- [19] Flemming, G., and Sonner, M. (1999): Can Internet polling work? Strategies for conducting public opinion surveys online. Paper presented at 1999 AAPOR Annual Conference, St. Petersburg, Florida, May 13-16, 1999.
- [20] Gräf, L. (2002): Assessing Internet questionnaires: The online pretest lab. In B. Batinic et al. (Eds.): *Online Social Sciences*. Seattle: Hogrefe & Huber, 73-93. (Forthcoming)
- [21] Grisaffe, D. (2000): Weighing your options on Web surveying? Look, then leap. Walker Information White paper, 2000. (November 26, 2001).  
<http://www.walkerinfo.com/resources/docs/whitepapers/ACF1BDB.pdf>
- [22] Groves, R.M. (1989): *Survey Errors and Survey Costs*. New York: John Wiley & Sons.
- [23] Hansen, W.J. and Haas, C. (1988): Reading and writing with computers: A framework for explaining differences in performances. *Comm. ACM*, **31**, 1080-1089. <http://www-2.cs.cmu.edu/~wjh/papers/framework.html>  
(November 21, 2001).
- [24] Hertel, G., Naumann, S., Konradt, U., and Batinic, B. (2002): Personality assessment via Internet: Comparing online and paper-and-pencil questionnaires. In: B. Batinic et al. (Eds.): *Online Social Sciences*. Seattle: Hogrefe & Huber, 137-153. (Forthcoming).
- [25] Hlebec, V. (1999): *Evaluation of Survey Measurement Instruments for Measuring Social Networks*. Doctoral dissertation. Ljubljana: Faculty of Social Sciences.
- [26] Hollis, N. (1999): Can a picture save 1,000 words? Augmenting telephone tracking with online ad recognition. *ARF's Online Research Day - Towards Validation*. New York: Advertising Research Foundation, 41-49.
- [27] Internet Rogator (1998): *Internet Rogator Help with Surveying*. <http://www.internet-rogorator.com/hm/help.htm> (November 21, 2001).
- [28] Joinson, A. (1998): Causes and implications of disinhibited behavior on the Internet. In J. Gackenbach (Ed.): *Psychology and the Internet: Intrapersonal, Interpersonal, and Transpersonal Implications*. San Diego: Academic Press.
- [29] Jones Thompson, M. (1999): When market research turns into marketing. *The Industry Standard*, August 23<sup>rd</sup> (February 15, 2002).  
<http://www.thestandard.com/article/0,1902,5995,00.html>
- [30] Jones, R. and Pitt, N. (1999): Health surveys in the workplace: comparison of postal, email and World Wide Web methods. *Occupational Medicine*, **49**, 556-558.
- [31] Kwak, N. and Radler, B.T. (1999): A comparison between mail and Web-based surveys: Response pattern, data quality, and characteristics of respondents. Paper presented at 1999 Annual Research Conference, organized by Midwest Association for Public Opinion Research, Chicago, Illinois, November 19-20, 1999.

- [32] Lozar Manfreda, K. (2001). *Web Survey Errors*. Doctoral dissertation. Ljubljana: Faculty of Social Sciences.
- [33] Lozar Manfreda, K., Batagelj, Z., and Vehovar, V. (2002): Design of Web survey questionnaires: Three basic experiments. *Journal of Computer Mediated Communication*, **7** (July 17, 2002).  
<http://www.ascusc.org/jcmc/vol7/issue3/vehovar.html>
- [34] Lozar Manfreda, K., Vehovar, V., and Batagelj, Z. (2001): Web versus mail questionnaire for an institutional survey. In A. Westlake et al. (Eds.): *The Challenge of the Internet*. London: Association for Survey Computing.
- [35] McNeish, J. (2001): Using the Internet for data collection – Just because we can, should we? Paper presented at 2001 AAPOR Annual Conference, Montreal, Quebec, May 17-20, 2001.
- [36] Robinson, J.P., Shaver, P.R., and Wrightsman, L.S. (1991): *Measures of Personality and Social Psychological Attitudes*. San Diego: Academic Press.
- [37] Sassenberg, K. and Kreutz, S. (2002): Online research and anonymity. In: B. Batinic et al. (Eds.): *Online Social Sciences*. Seattle: Hogrefe & Huber, 229-242. (Forthcoming)
- [38] Smith, B.A. (1999): Validating online research: Comparing data from telephone and Web-based surveys. Paper presented at the ARF's week of workshops, Online research workshop, New York, October 4-7, 1999.
- [39] Sperry, S., Edwards, B., Dulaney, R., and Potter, D.E.B. (1998): Evaluating interviewer use of CAPI navigation features. In: M.P. Couper et al. (Eds.): *Computer Assisted Survey Information Collection*. New York: John Wiley & Sons, 351-365.
- [40] Splichal, S. (1990): Analiza besedil. Statistična obravnava jezikovnih podatkov v družboslovnih raziskavah (Textual analysis: Statistic processing of language data in social science research). *Metodološki zvezki*, **6**. Ljubljana: Faculty of Social Sciences.
- [41] Terhanian, G. and Black, G.S. (1999): Understanding the online population: Lessons from the Harris Poll and the Harris Poll Online. *ARF's Online Research Day - Towards Validation*. New York: Advertising Research Foundation, 28-33.
- [42] Vehovar, V. and Lozar Manfreda, K. (2002): Mode effect in Web surveys. Paper presented at the 2002 AAPOR Annual Conference, St. Petersburg, FL, May 16-19, 2002.
- [43] Vehovar, V., Batagelj, Z., and Lozar, K. (1999): Web surveys: Can the weighting solve the problem? *1999 Proceedings of the Section on Survey Research Methods*. Alexandria: American Statistical Association, Alexandria.
- [44] Vehovar, V., Lozar Manfreda, K., and Batagelj, Z. (2001): Sensitivity of E-commerce measurement to survey instrument. *International Journal of Electronic Commerce*, **6**, 31-52.

- [45] Willke, J., Adams, C.O., and Girnius, Z. (1999): Internet testing. A landmark study of the differences between mall intercept and on-line interviewing in the United States. *Proceedings of the ESOMAR Worldwide Internet Conference Net Effects*. 21 -23 February 1999, London, 145-157.
- [46] Wolf, F.M. (1986): *Meta-analysis. Quantitative Methods for Research Synthesis*. Newbury Park: Sage Publications.
- [47] Wydra, D. (1999): Online tracking: A new frontier. *ARF's Online Research Day - Towards Validation*. New York: Advertising Research Foundation, 34-36.
- [48] Wygant, S. and Lindorf, R. (1999): Surveying collegiate Net surfers - Web methodology or mythology. *Quirk's Marketing Research Review*, July, 1999 (November 21, 2001).  
[http://www.quirks.com/articles/article\\_print.asp?arg\\_articleid=515](http://www.quirks.com/articles/article_print.asp?arg_articleid=515)
- [49] Zukerberg, A., Nichols, E., and Tedesco, H. (1999): Designing surveys for the next millennium: Internet questionnaire design issues. Paper presented at the 1999 AAPOR Annual Conference, St. Petersburg, Florida, May 13-16, 1999.