Comparison of the quality of qualitative data obtained through telephone, postal and email surveys

François Coderre and Anne Mathieu
*University of Sherbrooke (Canada)*

Natalie St-Laurent
*Ad Hoc Research*

Many claims have been made about the advantages of conducting surveys on the web. However, some concerns have been raised about the quality of the information gathered through this medium. The purpose of this research was to compare the quality of qualitative information obtained using three data-collection methods, in the context of the development of a scale for the measurement of corporate image. First, a study was carried out to generate a list of items that could be used to describe all elements of the corporate image of three firms as perceived by consumers. Different lists of items were obtained from telephone, postal and web-based surveys. Next, a qualitative study was conducted to assess the predictive validity of the lists of items obtained from each data-collection method. The results showed that the quality of qualitative data obtained through a web-based survey was comparable to that of information obtained through telephone and postal surveys, for two of the three target firms.

**Introduction**

The success of web-based research depends on the credibility of the data that it generates. Although it has been stressed that web-based data-collection methods have several advantages over other collection methods, such as low cost, short response delay and high selectivity (Weible & Wallace 1998; Forrest 1999; McNeisch 2000), there is still considerable discussion among researchers on an important issue: what is the predictive validity of web-based surveys? In order to address this issue, this paper first reviews the literature on the use of web-based qualitative research and then presents the results of a study conducted to compare the predictive validity of raw qualitative information gathered through email surveys with that of data obtained via telephone or postal mail.
Literature review

According to Cooper (2000), there are four major sources of error in surveys: coverage, sampling, non-response and measurement errors. Coverage error refers to a mismatch between the target population and the frame population (e.g. all residential telephone numbers). Sampling error occurs during the process of selecting a sample from the frame population. Non-response bias arises when there are significant differences between respondents and non-respondents on variables of interest. Finally, measurement error comes about when the value of a respondent’s answer on a measure differs from its true value.

In research on web-based surveys, coverage, sampling and non-response errors have received a great deal of attention (for a review, see Cooper 2000; Llieva et al. 2002). In contrast, measurement error has received little attention from researchers (Miller 2001).

In qualitative research, it is usually assumed that coverage, sampling and non-response errors are not the important issues they are in quantitative research. Because the objective of qualitative research is more to understand a phenomenon than to make inferences to a population, researchers are more concerned with the quality of the information that is gathered – that is, with measurement error. Web-based surveys have unique characteristics that could affect the quality of the information gathered. These are discussed below.

Web-based qualitative research

Although the advantages of conducting web-based qualitative research are numerous (see Tse 1999), its use also raises concerns among researchers. It has been reported that web-based participants may find it difficult to set down their feelings or ideas in writing (Tse 1999); that because web-based participants are able to read and reread their responses, their answers are likely to be better thought out and may be less spontaneous (Comley 1997); that the interviewer’s inability to capture non-verbal cues may limit the information elicited from participants; and that, because a higher level of commitment is needed to complete a web-based interview and participants may also feel that their anonymity cannot be preserved, the drop-out rate may be higher (Folkman Curasi 2001).

Of particular interest for the purposes of this research are the findings regarding answers to email open-ended questions and web-based in-depth interviews. Studies show that email recipients are more likely than mail recipients to respond to open-ended questions and to add comments to
questions (Bachmann et al. 2000); they write more comments and their responses are more insightful (Mehta & Sivadas 1995; Schaefer & Dillman 1998; Taylor 2000).

When comparing 24 semi-structured interviews conducted face to face with the same number of semi-structured interviews conducted on the web, Folkman Curasi (2001) found that the web-based interviews contained both the strongest and the weakest transcripts; they were less detailed (but could attain greater detail when follow-up questions were used); and they were far more grammatically correct and seemed less spontaneous: that is, the sentences had a clear beginning and end, with none of the ‘umming’ and ‘ahhing’ associated with face-to-face interviews.

Although the findings reported in the literature seem to favour the use of the internet for qualitative research, the criteria (e.g. number of recipients who answer open-ended questions, number of recipients who are more likely to add comments in a survey, or number of words contained in recipients’ answers) retained to assess this medium’s information quality are of limited usefulness since they provide only indirect evidence of the information’s validity. A stronger test is needed to assess the quality of the information gathered via this method. The research presented below was performed with this aim.

**Method**

The procedure used to assess the predictive validity of email open-ended questions was based upon the one followed to assess the predictive validity of corporate image. First, the construct of corporate image is introduced, and then two studies that were conducted are presented.

**Measurement of corporate image**

Corporate image is a multi-dimensional concept that represents a portrait comprising the beliefs and attitudes that people hold about a firm (Dowling 1986; Brown & Dacin 1997). Although authors do not agree on the number of dimensions connected with this concept or their labels, they generally use a two-step procedure to identify them. First, they conduct an exploratory study on a limited number of consumers in order to identify a list of items that can describe consumers’ beliefs and attitudes towards the firm. They then carry out a quantitative study on a much larger number of consumers and use multivariate analysis to reduce the items to a few dimensions. To assess the predictive validity of the dimensions identified,
A regression analysis is performed, using as its dependent variable a global measure of attitude towards the firm, and as independent variables the firm’s scores on the various dimensions.

A similar procedure is used in the present research. However, it will test the predictive validity of the list of items itself rather than that of the dimensions that could be generated from the list. The reason for this is that the identification of dimensions involves subjectivity and the present research is interested in assessing the validity of the raw information obtained via the internet.

**The qualitative study**

The first step in the development of a scale is to specify the concept of interest (here, three firms) and then to generate a number of items (associations or attributes) that consumers connect with the pre-designated concept.

**Subjects and design**

Three groups of 40 subjects were selected from a panel of households. The first group was composed of 40 subjects with home internet access. The other two groups were selected to match the demographic characteristics of the internet group as closely as possible – the three groups were very similar in terms of gender, education and age.

**Selection of the firms**

Three firms were selected for the purposes of this research: Bombardier, Ultramar and Videotron. There are several reasons for this choice. First, all three are well-known firms operating in Canada; this was important given that the target population of the study was Canadian consumers. Second, they operate in different sectors: Bombardier is a diversified international firm operating in the transportation sector; Ultramar operates gas stations; and, at the time of the study, Videotron was a cable firm providing television channels and internet access. Third, consumers’ familiarity with Videotron as an internet access provider was expected to be higher among the internet group than among the other two groups because those in the former group were more likely to use such services – thus allowing the study to compare the quality of the information gathered by the three
Data-collection methods when subjects’ involvement was likely to differ on the variable of interest.

**Data collection**

Data were collected via the internet (i.e. an email containing the questions was sent to subjects), mail and telephone. Two open-ended questions probed subjects’ beliefs regarding each firm: they were asked to indicate whether they thought the evaluated firm was a good firm and why, and also whether they liked the firm and why. The items generated through this procedure were then analysed by three groups of two judges, whose task was to eliminate all redundant items. In order to reduce bias that might result from fatigue, each pair of judges evaluated three firm/modality lists of items, according to a fractional design. In this way, for each firm, a different list of items was obtained for each data-collection procedure.

**Results**

Table 1 gives some information on the lists of items generated by the three data-collection methods for each firm. As can be seen, there are no important differences between the methods in terms of the average number of items per respondent, the total number of different items on the list or the number of unique items generated. However, the average number of items per subject differs across firms: subjects were more likely to generate

<table>
<thead>
<tr>
<th></th>
<th>Email (n = 40)</th>
<th>Mail (n = 40)</th>
<th>Telephone (n = 40)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bombardier</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average number of items per subject</td>
<td>3.1</td>
<td>3.2</td>
<td>3.2</td>
</tr>
<tr>
<td>Total number of different items</td>
<td>18.0</td>
<td>25.0</td>
<td>21.0</td>
</tr>
<tr>
<td>Number of unique items</td>
<td>2.0</td>
<td>4.0</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>Videotron</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average number of items per subject</td>
<td>2.1</td>
<td>2.1</td>
<td>2.1</td>
</tr>
<tr>
<td>Total number of different items</td>
<td>23.0</td>
<td>23.0</td>
<td>17.0</td>
</tr>
<tr>
<td>Number of unique items</td>
<td>2.0</td>
<td>4.0</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>Ultramar</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average number of items per subject</td>
<td>1.9</td>
<td>1.4</td>
<td>1.4</td>
</tr>
<tr>
<td>Total number of different items</td>
<td>19.0</td>
<td>20.0</td>
<td>15.0</td>
</tr>
<tr>
<td>Number of unique items</td>
<td>1.0</td>
<td>3.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>
items for Bombardier ($X = 3.2$) than for Videotron ($X = 2.1$) or Ultramar ($X = 1.6$). This finding was expected, since Bombardier is a well-known and respected firm for the target population of the present study.

The above analysis tends to indicate that the number of items collected through email open-ended questions is similar to numbers obtained via mail or telephone. To provide a more stringent test of the comparability of the three data-collection methods by assessing the predictive validity of the information gathered using these methods, a quantitative study was run.

The quantitative study

The quantitative study was conducted in order to compare the predictive validity of the raw information gathered in the qualitative study. The overall approach consists in assessing whether the beliefs that were identified using the three data-collection methods in the qualitative study are good predictors of consumers’ attitudes towards the three firms.

Subjects

The sample used in this study was drawn from a panel of households representative of the target population. Of the 1512 members of the panel, 576 respondents completed a self-administered questionnaire, which constitutes a 38.6% response rate. The sample was mostly feminine (58.9%), married (71.5%) and relatively educated (30.5% had attended university). The average age of respondents was approximately 45.

Data collection

A postal survey was conducted to assess consumers’ attitudes and beliefs regarding the three firms. The postal survey was preferred to telephone and email surveys, given the length of the questionnaire. The questionnaire included three sections. Attitude towards the three firms was measured in the first section and beliefs about the three firms in the second. The questions in the third section were unrelated to the present study. All respondents reported their attitudes and beliefs about the three firms.

Measures

Attitude towards the firms was measured using Chattopadhyay and Nedungadi’s (1992) three-item, seven-point Likert scale: I am favourable
to firm x; I like firm x; firm x is a good firm. A principal-components and reliability analysis (available upon request) was conducted and shows appropriate psychometric properties. The results confirm the three scales’ unidimensionality – all three items load on the same factor, for each firm. In addition, the resulting reliabilities exceed 0.85. The average of the three items was therefore used as a composite scale of attitude towards the firms.

Beliefs about the firms were measured with a seven-point Likert scale, using the list of items identified in the first study. For the purpose of data collection, the lists of items obtained from the three data-collection methods were pooled. For instance, in the case of Bombardier, 25, 23 and 18 items were identified via the mail, telephone and email collection methods, respectively. However, since some items appear on more than one list, a series of 30 different items were evaluated by the subjects.

Results

In order to compare the predictive validity of the three collection methods, separate regression analyses were performed for each method/firm combination, using as dependent variable the attitude towards the firm, and as independent variables, the list of items generated in the first study by the method concerned. In total, nine regression analyses were conducted. To facilitate comparison, $R^2_{adj}$ were computed, given that the number of independent variables differs across regressions. The results are presented in the first three columns of Table 2.

As Table 2 shows, small differences of $R^2_{adj}$ are observed across methods and firms. The values range from 0.451 for Ultramar/telephone to 0.509 for Videotron/mail. Given that different independent variables were used in each regression analysis, it is not possible to assess whether the differences observed are statistically significant. However, a visual exami-

<table>
<thead>
<tr>
<th></th>
<th>Email</th>
<th>Mail</th>
<th>Telephone</th>
<th>The three methods together</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bombardier</td>
<td>0.489</td>
<td>0.493</td>
<td>0.498</td>
<td>0.497</td>
</tr>
<tr>
<td></td>
<td>(18)</td>
<td>(25)</td>
<td>(23)</td>
<td>(30)</td>
</tr>
<tr>
<td>Videotron</td>
<td>0.473</td>
<td>0.509</td>
<td>0.471</td>
<td>0.511</td>
</tr>
<tr>
<td></td>
<td>(23)</td>
<td>(23)</td>
<td>(17)</td>
<td>(30)</td>
</tr>
<tr>
<td>Ultramar</td>
<td>0.469</td>
<td>0.465</td>
<td>0.451</td>
<td>0.467</td>
</tr>
<tr>
<td></td>
<td>(19)</td>
<td>(20)</td>
<td>(15)</td>
<td>(24)</td>
</tr>
</tbody>
</table>
nation of the table seems to indicate that the level of predictive validity of the three data-collection methods is similar, except perhaps for the case of Videotron where the $R^2_{adj}$ is higher when the mail method is used, and the case of Ultramar where the $R^2_{adj}$ is lower when the telephone method is used.

Additional analyses were conducted to determine whether the predictive validity of the information gathered using one data-collection method may be improved when it is combined with information collected using the other two methods. Thus, for each firm, the $R^2_{adj}$ was estimated when all the items obtained from the three data-collection methods are used as independent variables. The results are reported in the fourth column of Table 2. Next, following the procedure suggested by the Fisher test, the values from the fourth column of Table 2 (i.e. the full models) were compared with those of the first three columns (i.e. the reduced models). Nine comparisons were performed. The results are reported in Table 3.

No significant improvement was found in the $R^2_{adj}$ from the Bombardier data when the full model was compared with the reduced models based on the mail, telephone and email methods alone (all $p$ values $> 0.05$). Analysis of the data on Videotron revealed significant differences when the full model was compared with the reduced models based on the telephone ($F_{(13,537)} = 4.47, p < 0.001$) and the email ($F_{(7,537)} = 7.10, p < 0.001$) method alone; but not when it was compared with the reduced model based on the mail method alone ($F_{(8,537)} = 1.51, p > 0.05$). Finally, for Ultramar, significant improvement was observed in the $R^2_{adj}$ when the full model was compared with the reduced model based on the telephone

<table>
<thead>
<tr>
<th>Firm</th>
<th>Method</th>
<th>SSE ($F$)</th>
<th>SSE ($R$)</th>
<th>Df ($F$)</th>
<th>Df ($R$)</th>
<th>$F^*$</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bombardier</td>
<td>Email</td>
<td>261.1</td>
<td>270.78</td>
<td>539</td>
<td>551</td>
<td>1.66</td>
<td>$p &lt; 0.05$</td>
</tr>
<tr>
<td></td>
<td>Mail</td>
<td>261.1</td>
<td>265.36</td>
<td>539</td>
<td>544</td>
<td>1.76</td>
<td>$p &lt; 0.05$</td>
</tr>
<tr>
<td></td>
<td>Telephone</td>
<td>261.1</td>
<td>264.94</td>
<td>539</td>
<td>549</td>
<td>0.79</td>
<td>$p &lt; 0.05$</td>
</tr>
<tr>
<td>Videotron</td>
<td>Email</td>
<td>535.0</td>
<td>584.45</td>
<td>537</td>
<td>544</td>
<td>7.10</td>
<td>$p &lt; 0.01$</td>
</tr>
<tr>
<td></td>
<td>Mail</td>
<td>535.0</td>
<td>547.03</td>
<td>537</td>
<td>545</td>
<td>1.51</td>
<td>$p &gt; 0.05$</td>
</tr>
<tr>
<td></td>
<td>Telephone</td>
<td>535.0</td>
<td>592.80</td>
<td>537</td>
<td>550</td>
<td>4.47</td>
<td>$p &lt; 0.01$</td>
</tr>
<tr>
<td>Ultramar</td>
<td>Email</td>
<td>380.6</td>
<td>384.83</td>
<td>544</td>
<td>550</td>
<td>1.01</td>
<td>$p &gt; 0.05$</td>
</tr>
<tr>
<td></td>
<td>Mail</td>
<td>380.6</td>
<td>386.86</td>
<td>544</td>
<td>549</td>
<td>1.79</td>
<td>$p &gt; 0.05$</td>
</tr>
<tr>
<td></td>
<td>Telephone</td>
<td>380.6</td>
<td>407.30</td>
<td>544</td>
<td>554</td>
<td>3.82</td>
<td>$p &lt; 0.01$</td>
</tr>
</tbody>
</table>
method alone ($F_{(10,544)} = 3.82, p > 0.001$); but not when it was compared with the reduced models based on the mail ($F_{(5,544)} = 1.79, p > 0.05$) and email $F_{(6,544)} = 1.01, p > 0.05$) methods alone.

**Discussion and conclusions**

The results of this research are interesting in several respects. The analysis of the answers to open-ended questions in the qualitative study seems to indicate that the quality of the information gathered through the email survey is similar to that of information gathered through mail or telephone surveys, according to the criteria usually reported in the literature. For instance, it was found that the average number of items per respondent elicited by open-ended questions in the email survey did not differ from the numbers obtained in mail and telephone surveys. It was also found that both the total number of different items and the number of unique items generated by the email survey were very similar to those generated by mail and telephone surveys. These findings are in line with those reported in the literature supporting the use of web-based surveys (e.g. Mehta & Sivadas 1995; Bachmann *et al.* 1996; Schaefer & Dillman 1998).

However, when a more stringent test is used to compare the quality of the information gathered using the three data-collection methods, a somewhat different picture emerges. First, the predictive validity of the information gathered through the email survey was equal or superior to that observed with the other two data-collection methods, except in the Videotron case, where a lower level of predictive validity was observed in comparison to the mail method. A possible explanation for the lower level of predictive validity observed for the Videotron information may lie in the fact that subjects in the email group were likely to be more involved with this firm, as a major provider of internet services. Because of this higher level of involvement, these subjects may have developed very specific beliefs about Videotron that were not representative (or not as exhaustive a representation) of the whole population’s beliefs.

If this is the case, then these results suggest that when the issue under investigation is of equal interest for internet users and the overall target population (as was the case for the Bombardier and Ultramar data), one can be confident that the quality (i.e. the level of measurement error) of the information gathered by email survey is similar to that for data gathered through mail or telephone surveys. However, when the issue under investigation is one for which internet users and the overall target population are likely to differ in terms of interest, one should not rely exclusively...
on the information gathered through a web survey. In this case, more than one data-collection method should be used, as suggested by the triangulation approach. These findings reinforce the traditional warning issued by several researchers, cautioning that one should not make inferences from the results obtained though web-based surveys when there are significant differences between respondents and non-respondents (i.e. non-response bias) on variables of interest.

A second important finding from the quantitative study concerns the lower level of predictive validity observed when the telephone method was used to collect beliefs about Videotron and Ultramar. This finding may be explained by the time pressure associated with the telephone method. Unlike those in mail and email surveys, subjects participating in telephone surveys may lack the time to think deeply and may therefore be more likely to report only those beliefs that are most salient in their memory. Therefore, the list of items generated by this method may not be as exhaustive as those generated using data-collection methods where respondents do not answer under pressure.

Those using the results of this study should keep in mind that it looked only at the quality of the information gathered using open-ended questions in a study whose objective was to elicit a list of beliefs about an object, and that the number of subjects surveyed (40) was minimal. More complex forms of qualitative study can be conducted via the internet, such as in-depth interviews and focus-group discussions. More research is needed to assess the level of measurement error of these more involved types of qualitative study.

Acknowledgement

The authors would like to thank the Bombardier Chair in Brand Management for the financial assistance it provided for the conduct of this research.

References


