USING MICROSOFT EXCELT AS AN ALTERNATIVE SURVEY INSTRUMENT

John M. Rose. Phone: 02 9351 0168  JohnR@its.usyd.edu.au  Fax N/a
Institute for Transport Studies, University of Sydney, Sydney, NSW Australia, 2006

Iain R. Black. Phone: 02 9036 9143  I.Black@econ.usyd.edu.au  Fax 0293516872
Discipline of Marketing, University of Sydney, Sydney, NSW Australia, 2006

Christina Ioannou.  C.Ioannou@econ.usyd.edu.au  Fax N/a
Discipline of Marketing, University of Sydney, Sydney, NSW Australia, 2006

Alejandra Efron.  alejandra@its.usyd.edu.au  Fax N/a
Institute for Transport Studies, University of Sydney, Sydney, NSW Australia, 2006
Abstract
Over the past ten years, the use of the Internet and e-mails as communication tools has become ubiquitous. In the survey arena, the rising costs of gathering survey data have been partly compensated by the introduction of a range of new, relatively cost effective survey instrument. The use of the internet has led to an ever increasing number of online surveys, which not only reduce the cost per respondent in terms of information gathered but also time spent in the entering and cleaning of data. In contrast to internet surveys which require considerable upfront investment in terms of programming, the use of Microsoft Excel as a means of gathering survey data requires minimal programming skills on behalf of the researcher whilst offering the same benefits afforded by internet surveys. In this paper we present a case study investigating the use of Microsoft Excel as a survey instrument and assess its usefulness against a range of criteria. The case study examines a multi-attribute survey conducted to create a new psychological scale using a local (Australian) population of students.
Introduction:
At the heart of research into travel surveys methodologies specifically, and applied research more generally, is the search to develop improved methods for design and implementation whilst increasing the validity and reliability of results. An area that has shown great promise with this quest is the use of software programs to design and implement surveys and a wide range of internet technologies to collect electronic data, (for example the WWW (Aslnih et al., 2003), email (Best and Kreuger 2002) and Bulletin boards (Sweet, 2001). The benefits of the internet and associated software to research can be seen throughout the research process, however this paper will concentrate on three stages: (1) the use of software to design the survey instrument, (2) implementation, including sample frame development and recruitment, delivery of data collection tool, and communication with the respondents, and (3) preparing the data for analysis. This paper will present a case study describing the use of Microsoft Excel as a survey instrument and its implementation via email. This will then form the basis of an assessment of its advantages and limits for Internet survey design.

Aim:
The aim of the paper is to assess the use of Microsoft Excel as a survey design, implementation and data preparation instrument.

Objectives:
- To explain how a medium sized survey was designed and implemented using Microsoft Excel.
- To describe a number of benefits of using Microsoft Excel.
- To describe a number of issues that were encountered and how these were overcome.
- To describe the features and limitations of excel according to key design, implementation and data preparation criteria.
- To compare email delivery to online hosting of surveys.

Literature review:
In order to establish criteria against which to assess Microsoft Excel, an initial review of literature addressing the benefits and concerns (regarding respondent, sample and data issues) with online and electronic data collection techniques will be presented. Following this, literature will be reviewed describing the use of online hosting techniques and their advantages and disadvantages.

Benefits and Issues of Online surveys and electronic data collection
There are a number of benefits for both researchers and respondents from using Internet technologies in research. According to Thomson et al. (2003 the key advantages when compared to traditional paper and pencil methods are cost and speed. Cost savings fall into a number of categories; stationary savings come from the lack of paper, envelopes, ink and postage (Sheehan and McMillan, 1999). Administration costs including photocopying, letter fulfilment, typing and scanning are reduced as are the costs of labour intensive parts of data preparation including data entry, cleaning, and coding (McFarland, Ryan, and Paul 1998). However Thomson et al. (2003) see these cost benefits being countered to an extent by high start up cost caused by software and
hardware purchases and staff training. For example a commercial license for the widely used Websurveyor product can be $1000 per month.

The benefits of internet technologies are also seen in the speed with which surveys are delivered and returned and with the automation of data entry and preparation including data entry, cleaning, and coding (Dommeyer & Moriarty, 2000; Sheeban, 2001). Based on these benefits some authors suggest that internet based surveys are more cost effective than paper alternatives (Dommeyer & Moriarty, 2000; Kraut & Saari, 1999; Schmidt, 1997; Sheehan, 2001). In addition these technologies allow the researcher to use a full range of colour, sound, art and video as stimulus material or to improve the presentation of the survey and allows data to be collected on a wide range of research questions.

Despite these benefits there a number of respondent, sample and data issues that online surveys must address. Thomson et al. (2003) classified these into six common concerns. These concerns will be used as a basis for the assessment of Microsoft Excel conducted by this paper. It is clear that not all of the population have convenient access to the Internet. There are issues regarding whether one response per respondent can be ensured, particularly when incentives are used (Tierney, 2000). Start-up costs in terms of training and hardware and software costs are thought to be considerable (Donovan, 2000). Connection problems and network outages are a well-known feature of the Internet and this leads to a concern over whether respondents will be able to access and successfully complete Internet surveys. The presentation of the surveys can be affected by local computer setting (browser configuration, user preferences, monitor settings) and this leads to the concern that the researchers lose control of this important area. The sixth concern involves uncertainty whether respondents will accept this method. Work by Tierney (2000) suggests that this method is acceptable to many potential respondents. Further research is required in this area to highlight if particular groups in society are resistant to online research and how (if at all) they differ from respondents.

For an online research instrument to be judged as useful, it should be able to address a number of typical respondent concerns that are exacerbated by lack of trust in technology. These include; questions over reassurances of anonymity, data security and delivery. The much published computer hacking and virus activity may be seen to compromise the anonymity required for collecting valid data (Kraut & Saari, 1999; Magnan et al., 2000) as well as reducing response rates because of concern over becoming infected with a virus. This may be a factor in McDonald and Adam (2003) who found a lower response rate to online surveys compared to traditional methods and that respondents were different from non-respondents.

Forms of Online Surveys

The most common form of online survey used, are those centrally hosted on a remote server (Rhodes, Bowie and Hergenrather, 2003, and McDonald and Adam, 2003). Potential respondents are recruited using a variety of means including email and pop up advertisement inviting them to complete the survey. Respondents either follow the link embedded in the email or pop up and are directed to the webpage containing the survey. The survey is then completed whilst the respondent is online. This method has the advantage of high degrees of control over the presentation and content of the survey and can allow for real time data entry and rapid automated coding. In addition, preset statistics can be programmed to provide researchers with interim results at important intervals and final results immediately after data collection has finished.
This method is undermined by the expense of such services, the need to find reliable hosts and number of concerns over recruitment. The recent controversies over unsolicited email (Spam), and subsequent measures to reduce it (both legal and software based) make it very difficult to recruit respondents without a pre-existing relationship. The resistance to pop up advertisements also makes it harder to recruit respondents as click through rates have fallen dramatically (Dreze and Hussheer 2003).

Examples of this technique include Tierney (2000) who used online (and email surveys) to investigate effectiveness of a tourism promotional website. Methodological findings include that despite being effective in collecting data that showed websites are important in peoples travel plans, there was a low response rate and that respondents were different from non-respondents. Thomson et al. (2003) assessed reactions to web-based surveys when an organizational climate survey was moved online. As part of wider methodology, a web-based climate survey was administered (n = 403). Initial concerns were raised over anonymity, however the main finding was that most respondents were open to online surveying, and the Web-based medium did not appear to discourage participation from any subgroup.

Survey design features.

There are a number of questionnaire and research design features that Excel must be capable of constructing when used to construct internet surveys. Schaefer and Dillman, (1998) discuss the importance of well crafted introductory letters and invitations as these can increase the response rate. In addition to the numerous requirements regarding wording, questionnaires should be “respondent friendly” (Dillman 1978, Dillman, Sinclair and Clark, 1993). The use of instructions, clear layout of questions and answer formats, appropriate font size and type and appropriate number of questions per page address this.

A design platform must allow a full range of scale properties described in nominal, ordinal, interval and ratio scales to be used. It should also be capable of providing open, closed, graphic and written response formats and allowing randomisation of response options so to combat order effects (Green and Tull, 1978). It is also important to minimise respondent burden by effectively including skip patterns and question sequencing and should be capable of including a wide range of stimulus material including multimedia effects (Dillman, 2000). There should be no restrictions on the number of questions that can be asked and the length of the questionnaire. It is also important for internet data collection methods to be able to code answers automatically into a data file (Ioannou and Black, 2004).

Summary

This review of literature has detailed a number of cost and data quality criteria against which to assess Microsoft Excel. Researchers should be able to design presentable surveys with a full range of question types and response formats, and be able to implement these surveys quickly, cheaply and with a high degree of automated data preparation. The software should also be easy to use and
require minimal training for both researcher and respondent. A case study will now be presented detailing the use of Microsoft Excel to design and implement a scale item development survey.

Case study: Scale Development

Background
This case study is based on a study into satisfaction typologies which involved the development of a psychological scale (Oliver, 1997). As part of the data collected for this research a survey comprising of 10 sections was emailed to 450 students studying at an Australian university. A response rate of 74% was achieved; the high rate can be explained by the sample and the excellent sample frame.

The use of Excel as the instrument to design the survey was driven by a lack of funds to purchase commercially available software, the need for fast data collection and preparation. One of the researchers had experimented previously with the “forms” (see Figure 1) toolbar and suggested these functions, when combined with many others available on Excel, (such as cut and paste, paste link, formulas, macros, worksheets and password protection) could be used to create a survey.

Figure 1. Forms Toolbar

1.1. Design of the questionnaire
The questionnaire design process was guided by the nine-step framework of Churchill and Iacobucci (2002). This process can be summarised as follows; specifying the information sought, considering the method of administrations and the type of questionnaire to be administered, determining the content/response/wording/sequence of each question, determining the physical appearance of the questionnaire and pre-testing the questionnaire.

1.2. Types of Questions and Response formats
The questionnaire predominantly comprised of closed response questions, multiple item Likert scales with a five-point response format and semantic-differential scales with a seven point response format (see figure two, four and five). The key feature from the “Forms” toolbar used to design each of the response formats was the “Option” buttons inserted within a “Group box”. Each
option box was inserted into the group box in a specific order and each group box was cell-linked to a cell in the answer sheet. For example, if there were five option boxes within the group box and the respondent clicked on the first option box (see Figure 2 row 1) the number 1 would appear in the cell-linked cell (B1) on the answer sheet (see Figure 3, column 1). If they clicked on the fifth box (see Figure 2, row 2) then the number 2 would appear in the cell-linked cell (C2) on the answer sheet (see Figure 3, column 2). This demonstrates the program’s ability to design a range of questions easily and automate data entry and capture.

**Figure 2:**

<table>
<thead>
<tr>
<th></th>
<th>Scale Item</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>This is one of the best cars I could have bought</td>
<td>⬜</td>
<td>⬜</td>
<td>⬜</td>
<td>⬜</td>
<td>⬜</td>
</tr>
<tr>
<td>2</td>
<td>This car is exactly what I need</td>
<td>⬜</td>
<td>⬜</td>
<td>⬜</td>
<td>⬜</td>
<td>⬜</td>
</tr>
<tr>
<td>3</td>
<td>This car hasn’t worked out as well as I thought it would</td>
<td>⬜</td>
<td>⬜</td>
<td>⬜</td>
<td>⬜</td>
<td>⬜</td>
</tr>
<tr>
<td>4</td>
<td>I am satisfied with my decision to buy this car</td>
<td>⬜</td>
<td>⬜</td>
<td>⬜</td>
<td>⬜</td>
<td>⬜</td>
</tr>
</tbody>
</table>

**Figure 3:**

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Questions</td>
<td>5.01</td>
<td>5.02</td>
<td>5.03</td>
<td>5.04</td>
</tr>
<tr>
<td>2</td>
<td>Answers</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

**Figure 4:**

I would have to say that my car is............

<table>
<thead>
<tr>
<th>Important</th>
<th>Of no concern to me</th>
<th>Irrelevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>⬜</td>
<td>⬜</td>
<td>⬜</td>
</tr>
<tr>
<td>⬜</td>
<td>⬜</td>
<td>⬜</td>
</tr>
</tbody>
</table>

**Figure 5:**

5.2 Overall, how do you feel about your car?

Please click on the circle above the word that best matches your answer

<table>
<thead>
<tr>
<th>Terrible</th>
<th>Unhappy</th>
<th>Mostly Dissatisfied</th>
<th>Mixed</th>
<th>Mostly Satisfied</th>
<th>Pleased</th>
<th>Delighted</th>
</tr>
</thead>
<tbody>
<tr>
<td>⬜</td>
<td>⬜</td>
<td>⬜</td>
<td>⬜</td>
<td>⬜</td>
<td>⬜</td>
<td>⬜</td>
</tr>
</tbody>
</table>

**Navigation**

Skip patterns were not used however if formulas such as (IF, VLOOKUP) were inserted into the design, then sophisticated patterns can be created aiding navigation. Navigation is controlled by locking cells so that only certain cells can have information inserted into them and by inserting macros to control navigation between sections and worksheets (see Figure 6). The expertise
required by the researcher to know about and use these functions can be thought of as an intermediate level of knowledge with Excel. It does however depend on a strong knowledge of questionnaire design.

Presentation
A number of simple features were used to ensure that the survey was easy to read and presentable. The “Format-Cell-Font” function allows manipulation of size, font, effect and colour. This was used throughout the design process to create an aesthetically pleasing style and font scheme for the questionnaire. Respondent instructions and response formats were colour coded using “Format-Cells-Patterns”. The “Forms-Toggle grid” feature was used to create a mono-coloured background. In order to create a spacious layout and avoid crowding, the questions/scales were spread over multiple spreadsheets within a single workbook. A Macro was assigned to a Button (Figure 6) at the bottom of each spreadsheet in order to join the spreadsheets together.

Figure 6

CLICK HERE TO PROCEED TO SECTION 4 PART A

The ability to spread the questions over a number of spreadsheets also disguised the length of the questionnaire, which is especially useful for longer questionnaires. The canvas-style quality of the application (as opposed to a template design) provides the researcher with the ability to insert their own creativity into the design of the questionnaire and to apply the variety of features at their own discretion.

Order Effects
The ordering of questions is an important consideration in questionnaire design as a bias may occur if responses are in a consistent order. In order to overcome potential order bias, multiple versions of the original questionnaire can be created by rearranging the order of questions/scale items in each of the questionnaires. In this case, Excel provided a simple and efficient way to do this. Once the master questionnaire was designed, scale items in Section 2, 3 and 4 of the questionnaire were randomised using the cut and paste feature to create five versions of the questionnaire (yielding a total of 6 questionnaires). The key benefit of this application was the ability of each option box to be cell linked to the same cell within the answer sheet. In this way, if a scale item was positioned as item number 5 within Questionnaire no. 1 and it was positioned as scale item number 14 within Questionnaire no. 2 it would always appear in the same cell (e.g. B21) within the answer sheet. This meant that the responses were all coded in the same order, even though there were 6 versions of the questionnaire with questions/scale items arranged in a different order in each version.

Administration of the Questionnaire
The electronically configured questionnaires were password protected and uploaded onto the University Blackboard system. The survey was completed as part of a tutorial with each tutorial class randomly allocated one of the six versions of the questionnaire. A non-probability sampling method was used, and a screening process introduced prior to the completion of the questionnaire in order to ensure that only those respondents eligible to participate were included in the study.
Participants were briefed during the tutorial that the questionnaire was part of a study into consumer satisfaction with automobiles and that they would be required to electronically provide answers to a series of questions/scales. Respondents were instructed to download the survey from the online blackboard system onto their personal student drive. This method of delivery was used because the questionnaire was going to be completed during tutorial time and each student had access to the blackboard system. Alternatively, the questionnaire could have been emailed to the respondents.

Following the completion of the questionnaire, the respondents were instructed to save the questionnaire and anonymously email it back to the researcher. In order to preserve the confidentiality of the respondent in adherence with ethics requirements, the respondents were asked to send the questionnaire from their university email account that does not disclose the full name of the email account user.

Each questionnaire deposited in the researcher’s inbox was then sequentially saved to the hard drive (001, 002, 003 etc.). Subsequently, each questionnaire was opened and the row of answers from the answer sheet was then pasted into the data view of the Statistical Package for Social Sciences (SPSS) for subsequent data analysis. At this stage data cleaning was performed and additional coding steps were taken. Whilst in this case the data was transferred to SPSS for advanced multivariate analysis, it should be mentioned that Excel has its own Statistical package add-in (e.g., PACE XL, Analysis ToolPak) that potentially could be used to conduct data analysis directly within Excel.

Learning Process

A important issue to be addressed is how easy Excel is to use for survey design and how long it takes to train researchers in the specific features. The researchers in this case evaluated themselves as experienced (>5 years) and with an intermediate level knowledge of Excel functions. There was a range in research experience and knowledge between the authors ranging from novice to experienced researcher. The specific training comprised a short demonstration (approximately 20 minutes) of the features when the Excel “Forms” toolbar were explained. Additionally, examples of how to create macros and the method of cell-linking worksheets were provided. Following this, additional time was invested in self-education in order to become familiar with the capabilities of the Form toolbar features and to experiment with the dynamics of the application. Overall, the process of learning how to use the Form features was relatively straightforward and required a basic to intermediate knowledge of Excel spreadsheet software.

Design Limitations

There were a number of limitations with Excels design capabilities that became apparent during the study in addition to the issues with Macros and local computer settings discussed earlier. It is not possible to cell-link qualitative responses to the answer sheet. However, Excel could be use to transform the responses into a coded format prior to inserting into SPSS. There is a requirement for a large email capacity if the survey is being administering to a large sample. Excel is unable to record the time that the respondent commences or completes the questionnaire. Finally, the systematic process of downloading each questionnaire from the inbox to the hard drive requires attention to detail in order to prevent the duplication of questionnaires.
Assessment of Microsoft Excel as a Data Collection Tool

The literature review provided a number of criteria against which to assess the benefits of using Microsoft Excel to design internet surveys. Whilst concern over access to the internet remains, the ubiquitous nature of Excel and its backwards compatibility, means that it is very likely that computer users will have access to this program. It also allows floppy disk and CD ROM versions of the survey to be distributed where internet access is not available. Respondent identity can be checked via the email address used to return the survey however is remains very difficult to ensure that only one version from each email account is submitted.

Start-up costs in terms of training and hard and soft ware costs are usually considerable with internets surveys (Donovan, 2000). If a pre-loaded version of Microsoft Office is available then it has no additional hardware or software cost. There may be additional cost based on the size of the email account of the researchers if large numbers of surveys are sent and received and if additional server capacity may have to be purchased. Therefore reduced cost is a particular strong point of Excel.

After successful download of the email plus attachment, the Excel designed surveys are completed locally therefore reducing greatly the problems caused by server and computer outages. Short-term server issues will also not pose a significant problem to the return of the completed survey. Whilst Excel allows a wide range of colour, multimedia and presentational benefits at the design stage, correct functioning of these can be affected by local computer setting (browser configuration, user preferences, monitor settings). There were a small number of instances when students did not enable macros and design features did not work as intended because of local settings. However, the email relationship allowed these issues to be cleared up simply. This ability to communicate with the respondent is a considerable strength of email based surveys (though it raises anonymity issues that can be managed) as it allows explanation of the questions and technical issues to be addressed. Excel surveys also appear to be accepted by the respondents There were no refusals to participate in the study presented here, which adds to the earlier findings by Tierney (2000).

A final critical area against which Excel should be assessed is its ability to incorporate a number of questionnaire design features to ensure that it allows accurate communication of answers from the desired sample. We found it straightforward to include a full range of scales types and to present these in a respondent friendly fashion. Cognitive burden was reduced by being able to include additional instruction, hide worksheets, and included skip patterns. A full range of stimulus material was used, allowing it to be used for a wide range of marketing research objectives and data quality was improved by randomising scale items and choice sets.

Conclusions and Recommendations

There are two overall conclusions that we draw from this research. First we found that using the internet provides us with considerable cost and speed benefits. The design process was straightforward, delivery and return of questionnaires was done promptly and there were considerable efficiencies at the data entry, cleaning and coding stages. We would suggest that researchers consider this medium when designing research. Secondly. Excel is a promising alternative to paper-based or commercially available software in circumstances where the
researcher is faced by a lack of financial resources and time constraints. The application is cost-efficient, easy to learn and the elimination of a manual data entry process saves times and increases accuracy. Importantly it also allows for a full range of survey design features to be incorporated. At this point in time, the application appears to be efficient for short to medium sized questionnaires and for a small to medium sized sample.

The use of Excel has the potential to produce a highly advanced questionnaire incorporating mathematical formulas and advanced macros into the design. The efficiency of more complicated questionnaire designs (with loops and skip patterns) or administration across a large sample remains to be determined.
6. References


