Australasian Journal of Market & Social Research

December 2011 • Volume 19 Number 2 • ISSN 1832 7362

CONTENTS

AMSRS conference review 2011……………………………………………………………………………………………………………………………Page 9

Using preferred, understood or effective scales? How scale presentations effect online survey data collection
Philip A. J Derham……………………………………………………………………………………………………………………………………Page 13

Handling “Don’t know” responses in parallel CATI and online surveys
Don Porritt and Benjamin R. Marx………………………………………………………………………………………………………………………Page 27

Assessing flat-lining response style bias in online research
Con Menictas, University of Newcastle
Paul Wang, University of Technology Sydney
Brian Fine, Australia Online Research……………………………………………………………………………………………………………Page 34

Grounded theory by Birks & Mills, book review
Ruth Rentschler………………………………………………………………………………………………………………………………………………Page 45

Australian MARKET & SOCIAL RESEARCH SOCIETY AMSRS
The Australian Market & Social Research Society Ltd (AMSRS) is a not-for-profit professional membership organisation established to serve those practising or interested in market, economic and social research.

We have a national membership of over 2,000 researchers, marketing executives, market research managers and buyers, field managers, recruiters, interviewers, academics and students.

Being a member benefits those people:
- involved in or interested in market, economic, advertising or social research
- employed in the marketing or research sections of an organisation
- conducting qualitative and quantitative research
- who supervise field operations, interview or recruit participants for research projects
- who teach or study in marketing or social and market research

Membership entitlements include:
- discounted member rates to AMSRS and affiliate events
- a copy of the Annual Directory & Yearbook, and 11 copies of Research News
- regular updates on state and national industry information, developments and events
- a membership card and membership certificate
- Full members are entitled to use the postnominals MMSRS (Member of the Market & Social Research Society), and vote in divisional and national elections of the AMSRS.

Note: Overseas and Student members receive different membership benefits so please check our website for full details.

All members are bound to observe the AMSRS Code of Professional Behaviour, which covers both ethics and standard conditions of conducting and reporting scientific research.

**Contact us at:**
Membership,
Australian Market & Social Research Society
Level 1, 3 Queen Street
Glebe NSW 2037

Phone: +61 (0) 2 9566 3100
Or check out Membership on our website at www.amsrs.com.au
Australasian Journal of Market & Social Research

The Australasian Journal of Market & Social Research (AJMSR) is the official journal of the Australian Market & Social Research Society (AMSRS). It is published twice a year in June and December.

All members of the AMSRS receive a complimentary subscription to AJMSR as part of their membership.

Submissions
Original papers are invited. There is an editorial policy to ensure that a mix of theoretical and practical papers is presented. Papers should be of some relevance to at least a segment of the population of practicing market or social researchers. Purely academic marketing papers are probably more suitable elsewhere. All papers are refereed by independent assessors.

Authors are required to supply their articles electronically in Microsoft Word format, doubled spaced with article references in Harvard referencing style. Diagrams and figures also need to be provided electronically. In writing papers, authors should consider the style of articles in previous issues.

From this December issue onwards all editions of AJMSR will be referenced by EBSCO, an on-line database of journal articles accessible by researchers worldwide. As a writer we need your permission to enable this to occur. Could you please indicate your agreement to your article being referenced in this way when you submit it for publication.

Contributors of paper will receive an electronic copy of the issue in which their article appears.

AJMSR is an open publication. All expressions of opinion are published on the basis that they are not to be regarded as expressing the official opinion of the Australian Market & Social Research Society. The AMSRS accepts no responsibility for the accuracy of any opinions or information contained in this publication and readers should rely on their own enquiries in making decisions.

All papers for submission should be forwarded to:
Dr. David Bednall, Associate Professor
School of Management and Marketing
Faculty of Business and Law
Deakin University (Burwood)
221 Burwood Highway
Burwood VIC 3125 Australia
Email: dbednall@deakin.edu.au

Contact details
All enquiries regarding AJMSR matters should be directed to:
Australasian Journal of Market & Social Research
Australian Market & Social Research Society
Level 1, 3 Queen Street
Glebe NSW 2037 Australia
Phone: +61 (0) 2 9566 3100
Fax: +61 (0) 2 9571 5944
Email: publications@amsrs.com.au

Copies of the articles in AJMR may be made for personal or classroom use, without charge and with the publisher’s consent. Copying for any other purpose must first be approved by the publisher.

AMSRS recognises the contribution of Associate Professor David Bednall and Deakin University.
Advice to Authors

This journal publishes refereed:
1. Research articles
2. Research Notes
3. Case Studies
and opinion pieces (non-refereed).

Submissions
Submissions are only accepted by electronic means. Please see www.amsrs.com.au

Length
Research Article submissions should generally not exceed 6,500 words of text, excluding figures, tables, and references. Appendices are not encouraged.

Research Notes should generally be under 4,000 words of text, excluding figures, tables, and references. Appendices are not encouraged. They may be shorter reports of single findings, useful extensions or replications of previous research, or other additions to the literature that warrant publication without the more extended theoretical development or detailed analysis characteristic of regular articles.

Case studies should generally be under 4,000 words of text, excluding figures, tables, references and appendices. Appendices may be included if they contain reports of studies or other data used to support the case.

Research Article structure
These guidelines are for regular research article submissions. Authors who want to use a different structure are advised to contact the editor before submission. Innovation is encouraged so long as there is a clear benefit to the readers of the article.

Subdivision - numbered sections
Divide your article into clearly defined and numbered sections. Subsections should be numbered 1.1 (then 1.1.1, 1.1.2, ...), 1.2, etc. (the abstract is not included in section numbering). Use this numbering also for internal cross-referencing: do not just refer to “the text”. Any subsection may be given a brief heading. Each heading should appear on its own separate line.

Introduction
State the objectives of the work and provide an adequate background, avoiding a detailed literature survey or a summary of the results.

Theory/Conceptual
A Theory section should extend, not repeat, the background to the article already dealt with in the Introduction and lay the foundation for the research. Empirical studies should conclude this section with hypotheses or research questions. Non-empirical papers are encouraged to expand this section to give adequate background and treatment of the issues.

Material and methods
Where the work is an empirical study, provide sufficient detail to allow another researcher to reproduce the study. Methods already published should be indicated by a reference. Only relevant modifications should be described.

Results and Discussion
Results should be clear and concise. Discussion should explore the significance of the results of the work, not repeat them. A combined Results and Discussion section is often appropriate. Avoid extensive citations and discussion of published literature.

Conclusions
The main conclusions of the study should be presented in a short Conclusions section. Implications for further research, limitations and implications of the research for industry practice must be given.

Appendices
Appendices are not encouraged.

Table and Figure Titles
These should appear above the Table or Figure in bold and be numbered consecutively. Supply titles separately, not embedded in the figure or table. Titles should be brief (preferably under 10 words).

Math formulae
Present formulae using the Word equation editor. Number consecutively any equations that have to be displayed separately from the text (if referred to explicitly in the text).

These guidelines have been extensively adapted from the submission guidelines for the Australasian Journal of Marketing, see http://www.elsevier.com/wps/find/journaldescription.authors/717598/authorinstructions and Public Opinion Quarterly, see http://www.oxfordjournals.org/our_journals/poq/for_authors/general.html
Footnotes
Footnotes should be avoided except in Tables and Figures. Indicate each footnote with a superscript lowercase letter.

Essential title page information
Title
Concise and informative. Titles are often used in information-retrieval systems. Avoid abbreviations and formulae.

Author names and affiliations
Where the family name may be ambiguous (e.g., a double name), please indicate this clearly. Present each author’s affiliation (where the actual work was done) after their names.

Corresponding author
Clearly indicate who will handle correspondence at all stages of refereeing and publication, also post-publication. Ensure that your e-mail address and the complete postal address are provided.

Abstract
A concise and factual abstract is required. The abstract should state briefly the purpose of the research, the principal results and major conclusions. An abstract is often presented separately from the article, so it must be able to stand alone. For this reason, References should be avoided, but if essential, then cite the author(s) and year(s). Also, non-standard or uncommon abbreviations should be avoided, but if essential they must be defined at their first mention in the abstract itself.

Keywords
Immediately after the abstract, provide a maximum of 6 keywords, using UK English spelling and avoiding general and plural terms and multiple concepts (avoid, for example, "and", "of"). These keywords will be used for indexing purposes.

Abbreviations
Define abbreviations that are not standard in this field in the text. Ensure consistency of abbreviations throughout the article.

Acknowledgements
Collate acknowledgements in a separate section at the end of the article after the references and do not, therefore, include them on the title page, as a footnote to the title or otherwise. List here those individuals who provided help during the research (e.g., providing language help, writing assistance or proof reading the article, etc.).

Format
The default is single spaced, A4 format with 2.54cm margins on each side. The use of colour is acceptable for graphics and figures, but should otherwise be used sparingly. Its use is intended to make each article easier to follow rather than to cater to aesthetics or fashion. It is anticipated that case studies will include colour photographs and by arrangement, videos.

Use of word processing software
Documents should be submitted in Microsoft Word format. Keep the layout of the text as simple as possible. Please do not use the word processor’s options to justify text or to hyphenate words. Please do not use multiple columns of text. However, do use bold face, italics, subscripts, superscripts etc. To avoid unnecessary errors you are strongly advised to use the Word "spell-check" and "grammar-check" functions. The AJMSR uses UK English as its preferred style for spelling.

When preparing tables, if you are using a table grid, use only one grid for each individual table and not a grid for each row. Only horizontal lines should be used in Tables. Figures should be embedded in the document. If graphics are used, please also provide a separate file with your submission. If you are not the copyright owner of material you have included, you must evidence your permission to use this material.

Artwork
Electronic artwork
• Make sure you use uniform lettering and sizing of your original artwork.
• Save text in illustrations as "graphics" or enclose the font.
• Only use the following fonts in your illustrations: Arial, Courier, Times, Symbol.
• Number the illustrations according to their sequence in the text.
• Use a logical naming convention for your artwork files.
• Provide captions to illustrations separately.
• Produce images near to the desired size of the printed version.
• Submit each figure as a separate file.
Please do not:
• Supply files that are optimised for screen use (like GIF, BMP, PICT, WPG); the resolution is too low;
• Supply files that are too low in resolution;
• Submit graphics that are disproportionately large for the content.

Colour artwork
Please make sure that artwork files are in an acceptable format (TIFF, EPS or MS Office files) and with the correct resolution.

References
The use of Endnote™ bibliographic software is recommended. The APA Sixth Edition is the standard used for referencing.

Citation in text
Please ensure that every reference cited in the text is also present in the reference list (and vice versa). Any references cited in the abstract must be referenced in full. Unpublished results and personal communications are not recommended in the reference list, but may be mentioned in the text. If these references are included in the reference list they should follow the standard reference style of the journal and should include a substitution of the publication date with either "Unpublished results" or "Personal communication" Citation of a reference as "in press" implies that the item has been accepted for publication.

Web references
As a minimum, the full URL should be given and the date when the reference was last accessed. Any further information, if known (DOI, author names, dates, reference to a source publication, etc.), should also be given. Web references should be included in the reference list.

Reference style
Text: Citations in the text should follow the referencing style used by the American Psychological Association. You are referred to the Publication Manual of the American Psychological Association, Sixth Edition, ISBN 978-1-4338-0561-5. Journal names should not be abbreviated. Use journal issue numbers only where these are necessary to identify the source of the article.

List: references should be arranged first alphabetically and then further sorted chronologically if necessary. More than one reference from the same author(s) in the same year must be identified by the letters "a", "b", "c", etc., placed after the year of publication.

Examples:
Reference to a journal publication:

Reference to a book:

Reference to a chapter in an edited book:

Submission checklist
The following list will be useful during the final checking of an article prior to sending it to the journal for review. Please consult this Guide for Authors for further details of any item:
• One Author designated as corresponding Author:
  • E-mail address
  • Full postal address
• All necessary files have been uploaded
• Keywords
• All figure and table titles
• All tables (including title, footnotes)
• Manuscript has been spellchecked and grammar-checked
• References are in the correct format
• All references mentioned in the Reference list are cited in the text, and vice versa
• Permission has been obtained for use of copyrighted material from other sources (including the Web)

The use of a copy editor to check the text for completeness and readability is advised.

If you have further queries, please contact the editor prior to submission.
The Industry Speaks Research!

Delightfully, this issue focuses on papers contributed by the industry or academics working collaboratively with the industry. It includes a summary of papers from the 2011 AMSRS Conference. (Many thanks to Peter Vitartas from Southern Cross University for organising the refereed stream at the Conference).

It is sometimes claimed that people from industry and companies will not contribute to this journal because of a reluctance to disclose their intellectual property to competitors. Perhaps with closed, proprietary systems a submission is unlikely. But with a range of important method issues, such as response rates and survey flat-lining, it is in everyone’s interest to share data and problems. It can help the industry make its practices more robust, allowing it to adhere to quality standards and reassure clients of the quality of information provided. Another possible issue for non-submission is a fear that others will criticise their methods and writing.

The editorial policy of this journal is that one academic and one industry reviewer be asked to review submissions. Reviewers are not told the identity of those submitting an article. "Blind" reviews like these can lead to some reviewers being rather more negative and scathing than they might otherwise be. It is an editor’s job to review the reviews and, as far as humanly possible, to moderate the outcomes to achieve both fairness and rigour. Usually the reviews contain many positive suggestions and ideas which those submitting find very useful. If the paper is rejected, ultimately, no one else has to know and the feedback will be useful. If the paper is accepted, the authors get independent certification of the quality of their work. Ultimately the published paper is also likely to be greatly improved by the review process. Thus, the Board of the journal is hopeful that there will be many further, quality submissions from industry in the years to come.

Associate Professor David Bednall
Editor
December 2011
AMSRS 2011 "Curiouser & Curiouser" conference - A Snapshot review of the curiosity

The 2011 AMSRS conference was held on 8th and 9th September in Sydney.

This paper reviews a selection of papers that examined curiosity during the conference. In so doing, three areas of curiosity are reported on – curiosity in the world around us, curiosity with new communication platforms and curiosity surrounding new research approaches.

Curiosity in the world around us
Sheila Keegan’s paper titled “Our minds are a changing. Should we be worried?” addressed curiosity from the perspective that the community is evolving. Sheila examined what this evolution means for the way we spend our time and relate to each other.

Sheila raised the possibility that our reliance on the internet is producing increased shallowness in our thinking. This is because our minds are responding to numerous stimuli quickly rather than focussing more deeply on select stimuli.

Frequent internet usage can also impact our emotional responses. The shallowness in our thinking can be replicated in the shallowness of our feelings. Consequently as a society, we can feel more socially fragmented despite the connections that social media provides.

Researchers, as members of society, may also be behaving this way. Clients are requesting greater speed of response. Researchers are therefore analysing the large amounts of data that is becoming available in the internet era faster, but more superficially as there is less time available. Hence researchers are prone to providing less insight.

The challenge for researchers is not to divorce themselves from the world around us. Rather it may be appropriate to report on the trite and trivial if this reflects how consumers are behaving. However this approach should be combined with the depth of thinking that can be reasonably expected from the market research profession.

Curiosity with new communication platforms
Jason Buchanan, Pete Cape and Keith Miller presented their paper “Connecting with the new consumer”.

This paper argues that people still seek the same information (e.g. news), but rather the form in which they prefer to receive information is changing, for example social media approaches to news dissemination. Therefore changes in behaviour may simply be a “communication platform” factor. Underlying human behaviour remains essentially the same.

This paper emulated Sheila Keegan’s view that consumers are multi-tasking and having less time to do individual tasks. 40% of respondents have done “other things” whilst “taking part” in market research. This paper questioned whether this is a problem if the process of participating in market research mimics how consumers behave and what they decide in reality.
The implications for researchers is that if respondent attention can only be gained in snippets, then maybe questionnaires can also only be administered in snippets. This means researchers will need to deal with increasingly incomplete data records. Alternatively, it may require substituting people to gain a whole record.

If researchers are to gain whole records for respondents, then surveys will need to be more engaging through the new media in order to gain 100% completion. For example, surveys may need to be games based rather than Q&A based. Avoiding biases in sample selection will then become a further consideration.

Curiosity with new communication platforms was continued with the paper presented by Sean Dunn & Ying Xin of Vision Critical titled “Unplugged – Research Goes Mobile”.

This paper examined how mobile surveys will be embraced more as Smartphone applications become more prevalent.

Mobile surveys have a number of intrinsic benefits. They can tap into people’s downtime. They can also be administered closer to the actual experience e.g. timing for a customer satisfaction survey.

The limitation of mobile surveys is that they cannot replicate other on-line surveys. This is because mobile surveys need to be more simple in design and shorter in length.

**Curiosity of new research approaches**

In her paper titled “The Caterpillar becomes the Butterfly: The inevitable metamorphosis of market research”, Erica van Lieven addresses how market research needs to evolve for a more data abundant world.

Erica postulates that given 80% of market research is currently dedicated to quantitative research, market research is likely to move beyond the survey to listening using the social media.

Erica provided a framework for research methodology to 2021. In so doing, Erica depicted a pathway of an industry that is methodologically curious. The market research industry has accepted mixed research approaches and is now embracing the rise of on-line panels and on-line communities.

As the research industry moves to 2021, it is likely researchers will become increasingly ‘invisible’. The industry will find new ways of listening. We will be observing in the social media without interference. The industry is also likely to become more attuned to tapping into the sub-conscious and emotions e.g. by using neuroscience, biometrics and eye tracking.

How does this compare to 2004? Van Lieven cites Roth (2004) who states that in 2004 “identifying communication opportunities outside traditional channels” was becoming more common practice. Today Van Lieven states “understanding and connecting with the customer is the number one business channel”. In otherwords, market research is moving from understanding communication opportunities to finding new ways of connecting with the customer.

As innovation accelerates in client organisations, it is imperative to remain connected with the consumer. This can require information from social media over and above
tracking studies. Consumer insight will continue to underpin the “raison d’être” of market and social research.

The challenge is that some clients are placing greater reliance on speed rather than quality of information. This is because innovation cycles in client organisations are becoming smaller. Allowing 12 weeks in the new product development cycle for market research is a luxury that cannot be afforded. Hence innovation in market research is required to help clients do things better, faster or cheaper in order to gain a competitive edge.

Curiosity in research approaches had a strong on-line focus at the conference. Sebastian Watson and Howard Parry-Husbands presented a paper on the topic “If they build it they will engage – understanding the value of collaborating with the audience to ensure engagement with online communities”. This rather long title makes the theme of the paper self-explanatory.

This paper highlights the importance of identifying influencers in the early stages of the on-line community to assist with shaping and building the community. The authors’ research showed that people are willing to engage in branded communities and that these communities will in turn deliver powerful advocacy for those brands and branded communities, if the purpose of these communities align with the passions of the community members.

Curiosity also extended to “Taking the conversation online; a researcher’s guide to the benefits and pitfalls of on-line qualitative research”. This paper presented by Jo Farquhar and Victoria Gamble, discussed how people are much more open to communicating on-line. Consequently people find it less daunting to participate in an on-line discussion. Furthermore many respondents are now incorporating social networking logic and design into their own social platforms, making the experience for the user and the moderator more intuitive and user friendly.

The key when running an on-line forum is for the researcher to make themselves as real as possible. To do this, the authors suggest the researcher uploads their own photo, uses a casual tone in addressing the forum and provides alternative contact methods (phone, email) for building their relationship.

During the conference, Ray Poynter provided an update to his “Handbook of online and social media research 2010”

Ray discussed how Neuroscience and biometrics are increasingly finding their niche in the evaluation of advertising and campaigns.

50% of market research revenue (ESOMAR 2010) comes from activities where respondents are not asked questions e.g. processing loyalty card data, web analytics. This is referred to as “Big Data”.

“Big data” can be stored in a brand-centric model – where all the data is held by the brand e.g. CRM data, loyalty card data. Alternatively it can be stored in a 3rd party warehouse e.g. Kantar’s X axis data base with profiles of greater than 600 million people, in order to help target advertising. The challenge is for market researchers to add the “why” to BigData’s “what”.

Ken Roberts, Elaine Wong and Darren Stein in their paper “feelings, nothing more than feelings: lessons for emotion based communication” provided an alternative view to the
merits of neuroscience. They claim that the development of neuroscience in predicting the response of consumers to stimuli assumes that neuroscience can unlock the discrete emotions that drive purchase behaviour.

The authors claim the idea that specific feelings are linked to one part of the brain and are therefore identifiable via neural imaging technology is not borne out of reality.

As an alternative method, Forethought has developed a feelings scale that is similar to a Likert scale showing visual emotional responses that respondents select from. Lessons from this feelings scale include:

- If you choose an emotion to focus on, then make sure it drives the desired consumer behaviour.
- Emotion is more important for acquisition than retention
- Communication can elicit different emotions in different consumer segments.
- Creative executions that attempt to explicitly activate an emotion do so less effectively than executions that implicitly attempt to execute the emotion e.g. a TVC which says “be happy” is unlikely to activate a happiness emotion. Rather more explicit emotional attempts activate a more cognitive appraisal.

David McCallum & Alastair Gordon’s article titled “Say it to my Face! – Applying Facial Imaging to understanding Consumer Emotional Response” was awarded “Best conference paper” and aptly concludes this review.

Of particular relevance within this paper is the distinction between facial imaging and facial coding. Facial imaging, the subject of this paper, is teaching a machine to understand a face and to estimate automatically what is going on. Facial coding is the method by which a human is trained in annotating the movement of facial muscles.

The key applications for facial imaging is presented as offering 4 benefits, namely

- Scalability – can be used with large or small samples.
- Non-intrusiveness – respondents need only to watch
- International applicability – can work across cultures
- Adaptability – easy to integrate with conventional studies

Key applications of facial imaging include measuring emotional response to advertising and sensory tests.

Facial imaging helps overcome some of the downsides of conventional research. Firstly conventional research is based on rational models of information and forces respondents to engage their rational brain. Secondly conventional research relies on recall and can be too slow to capture emotional responses. Finally conventional research is limited in taping into subtle, often unconscious reactions.

**Conclusion**

The 2011 AMSRS conference offered many perspectives around the theme “Curiouser and Curiouser”. The upshot of these perspectives was that change is accelerating; information is becoming more abundant and clients are under pressure to make decisions faster. In this race, social media platforms are emerging. The challenge for the researcher is to offer depth of insight as timeliness and platforms gravitate towards superficiality.
Using preferred, understood or effective scales? How scale presentations effect online survey data collection.

Philip A J Derham, Derham Marketing Research Pty. Ltd.

ABSTRACT

This paper reviews the impact of different scale presentations in online surveys on closed ended question completion levels, respondents' preferences for scale presentation types and the influence those preferences have on questions answered and on respondents' future survey intentions. The scale presentations assessed were word scales, number scales, and emoticon scales.

The rationale for the project was that marketing research seeks to obtain full and accurate answers from people invited to complete online self-completion surveys, and that survey design influences response and data quality.

Self-completion surveys often use Likert scales (expressed in words, numbers or images) to collect data about attitude, intention or feeling. Online surveys can use emoticons (moving imagery) to present scale answers in place of word or number tick-a-box scale presentation. As emoticons have been seen, intuitively, as attractive and useful variations to long lists of word or number scale presentations, it was hypothesised that the use of emoticons in place of word or number scales would strengthen online survey data collection.

Safe research practice requires confirmation of intuition, so five online surveys, each showing different presentations of scale formats, were undertaken in 2009 and in 2010.

The findings indicate that the most effective scale presentation was number scale presentations if more answers in the questions are sought but that number scales were least preferred and may be better avoided if the survey design aims include being attractive and interesting to respondents.

More respondents reported that they preferred word format scale presentations than emoticon or number scale presentations and respondents who completed word scale presentation format surveys were more likely to say they would complete similar such surveys in the future than those who completed number scale presentation format surveys.

The least effective of the scale presentations tested were emoticons (images and slider scales) and their use is not recommended.

INTRODUCTION

The marketing research industry has long been concerned about declining response rates to surveys (Goyder, McKenzie Leiper 1985, Malhotra 1999, Bednall, Shaw 2003, Hardie, Kotsomitis 2005, Groves 2006), about respondents not answering all questions in self-completion surveys (Bryman, Bell 2007), and by respondents “flat-lining” their answers to scale questions (Reynolds, Anderson and Sharp 2009).

Flat-lining is the term used when answer patterns to matrix questions suggest respondents merely moved down a column of statements selecting the same answer to each statement, without thinking afresh about each new statement. While such answers may reflect respondents’ views, if such answers were unconsidered movements through the questionnaire, those answers would raise concerns about the usefulness of the data captured. Better designed surveys have been posited as solutions to these concerns.

The question then is what make surveys better designed? In particular, given the primacy of online surveys in Australian marketing research (ESOMAR 2009, AMSRS/AMSRO 2010), what elements make better designed online surveys?

The online survey design element reviewed for this paper was the presentation of ranking or rating scales. These are often presented in a matrix, with questions in descending rows and the answers at the top of each column, as Figure 1 shows.

If scale questions could be better designed, the potential problem of flat-lining could be overcome, more questions would be fully answered, and online surveys would provide more effective data.

Scale type advantages and uses have been extensively discussed in marketing research literature and are key elements in textbooks (Boyd,
Westfall, Stasch 1977, Malhotra 1999, Malhotra, Hall, Shaw, Oppenheim 2006, Aaker, Kumar, Day, Lawley, Stewart 2007). Some texts offer advice on scale construction (Malhotra etc. 2006, Aaker etc. 2007), though that advice can be to pilot-test the scales before final use.

Some have recommended use of verbal scale formats (Malhotra etc. 2006) and others note use of smiley face and thermometer scale presentations without recommendation (Aaker etc. 2007) or with recommendation as best for use with children (Malhotra 1999, following Alwin 1997).

Visual scale presentations were proposed as good representations of multivariate data (Chernoff 1973) and visual stimuli were seen to be more effective in eliciting correct answers than direct questioning in some situations (Boyd, Westfall, Stasch 1977). Medical researchers have found the Chernoff faces to be the least preferred scale presentations for elderly people and reported graphic analogue scales were effective (Castle, Engberg 2004 and example in Figure 2), while others note graphic analogue scales findings could differ from those obtained by categorical scales (Funke, Reips 2006). Visual analogue scale presentations resulted in longer survey times and higher rates of missing data (Couper, Tourangeau, Conrad, Singer 2006).

Visual languages provide limited expressive value (Chrystal 2008) and the cultural specificity of at least some visual scale presentations can effect interpretation by respondents and researchers (Cape 2009). Hence, the effectiveness of visual scale presentations in marketing research was not clear, though the popular beliefs are that pictures are worth a thousand words (Barnard 1921) and that people find it easier to read pictures as these help convert data into food for the mind (Reichmann 1964).

My interest in examining this issue was piqued by a number of unsolicited comments received from online surveys, over some years. Some unsolicited comments about the emoticon visual scale presentations, which motivated this study, are overleaf.

Figure 1: Matrix question example.
“Great survey - I love the scales and happy faces - one of the best that I have completed! (sic).” “The clouds/glass/people are a cute touch :)” “Picture Scales are new Clever” “I prefer the picture scales - they make the survey fun!” “They are really good! Never seen them before, I've done lots of surveys and they are quite interesting!!!” “novel approach, good to see some innovation, and animated gifs” “I was hoping you will ask! Both of them very easy and pleasant to use. Also great that it is big and spaced out, it helps so much.” “Absolutely love the pictures on the scales.” “very cute and also clearly conveys the thought/emotion across.” “nice to see something different and pleasant for a change in a survey, it makes all the difference”.

Source: Derham Marketing Research’s online surveys, 2003 to 2009.

The particular scale presentation formats examined are word, number and visual image/slider scale presentations. The visual/slider scale presentations are generally called emoticons.

A possible (and so tested) advantage of the emoticons is that statements that are run one after another in matrix questions in word or number scale presentations must be asked as individual questions (Figure 10). In contrast, word and number scale presentations can be asked in long lists (Figures 8 and 9), and these lists of questions can be susceptible to flat-lining.

**Emoticon images**

Emoticon images are representations of scales and have been developed for use in online surveys. They are mostly used in scale questions and have been suggested as online survey design improvements. Online survey respondents are shown a question with an emoticon as the answer device and are asked to change the emoticon image to the one that best reflects their answer, using a slider next to the image to make the change.

The emoticon scale presentations were provided as a standard part of an online survey package and with an implicit, if un-evidenced, assumption that their use would improve online surveys. It seemed appropriate to test emoticon scale presentation impact on question answers and so on survey design.

Emoticons can be constructed from a range of images and the three used in the 2009 and 2010 surveys are shown in Figure 3. Each is shown at the midpoint neutral position, the position first seen by respondents. When respondents move the slider, the visible image changes. Only one image is visible at any one time, and the image visible depends entirely on the position of the slider.

**Figure 3: Emoticon images used, shown at the neutral position.**
In this study, each set of question and emoticon image answers was accompanied by written instructions asking respondents to move the slider to make the image the one which best represented their answer to the question. The images in Figures 4 to 6 show the range in each emoticon scale tested.

**Figure 4: Face and word emoticon image range.**

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
</table>

, or

**Figure 5: Glass full to empty emoticon image range.**

Or

**Figure 6: Sunny to stormy emoticon image range.**

The survey program recorded whether the respondents used the slider. If the respondent did not use the slider, the program recorded a “no answer” to the scale question when the respondent answered another, later question.

If the respondent wished to record a neutral answer, they had to move the slider and change the image to a positive or negative image and then return the slider to the neutral position, changing the image back to the neutral image. In contrast, when selecting a neutral answer in the word or number scales, respondents select the neutral answer box directly, without needing to first select a positive or a negative answer.

**METHOD**

The five surveys reported in this paper were undertaken by online survey. All were about banking but all were designed to identify whether scale presentation differences influenced question answer rates. Each survey was designed to have the same look, size and screen layout. Figure 7 shows that style with a question that was common to all surveys.

**Figure 7: Example of a screen style common to all surveys.**

The sample for each survey was drawn from the same source and all surveys were about banking. Every effort was made to ensure the email invitations and surveys themselves were the same, scale presentations excepted, and this comparability of survey enabled comparisons to be made.

The surveys used the electronic equivalent of the “tick-a-box” answer method for closed-ended questions, except when the emoticon format scale presentation was used. That format required respondents to touch and move the slider, to record answers.

All surveys included open-ended questions that allowed respondents to add in comments. By design, respondents did not have to answer a question to advance through their survey.

Examples of the four scale presentations used are shown in Figures 8 to 12. It was not necessary to read each question in these scale presentations as the discussion is about the presentation styles, not the questions.
The 2009 survey used the scale presentation example Formats 1 (words), 2 (numbers) and 3 (emojis). In 2010, the questions in one survey were in the Format 1 (words) style. Questions in the second survey were in the Format 2 (numbers) style, the third survey questions were in the Format 3 (emojis) style, and questions in the fourth survey tested the words, numbers and emojis formats with a “can’t say” interval added to each scale presentation.

Formats 1, 2 and 3 had the answer rows and emoji slider in horizontal positions. The horizontal slider position was deliberately adopted to match the horizontal style of the Format 1 and 2 word and number answers. Format 4’s emoji slider was positioned to the right of the image and required a vertical up and down movement to change the image (as this up and down movement reflects more common cultural practice than the horizontal slider - Casasanto, Dijkstra 2010).

Each potential respondent received one email invitation to participate in their survey and each was invited to participate in just one survey.

Each survey asked about financial institution use, and about the importance of, satisfaction with and likelihood of use of specified account types.

The 2009 survey was of 90 questions and included three sets of scale presentation Formats 1, 2 and 3. In the 2009 survey, participants were asked to assess scale presentations on a range of attributes and the formats for these attribute questions are shown in Figures 13 and 14. At the end of the 2009 survey, respondents answered questions about their self-perceptions, their demographics, whether they had enjoyed the survey and how likely they were to complete a similar survey in the future.

Each of the four 2010 surveys were about banking, and each was of 33 questions. The questions in each survey were the same and differed only in presentation of the scales formats. Each of the four 2010 surveys collected demographic details and, as in 2009, concluded by asking participants how they enjoyed the survey and how likely they were to complete similar surveys in the future.
Figure 11: Format 4: Word scales with the six points format, shown first in the Format 4, sequence. This presentation was followed by the emoticon scales with six points format, with the image shown in the initial neutral position (shown in Figure 12).

Figure 12: Format 4: Emoticon scales with the six points format, shown last in the Format 4, sequence.

Figure 13: Example scale presentation preference question.

Figure 14: Example scale presentation preference question.
RESULTS
Section 1 looks at question answer rates, Section 2 looks at respondents’ perceptions of different scale presentations, Section 3 looks at enjoyment from completing the survey and the impact of that on future survey intentions and Section 4 looks at the willingness to provide contact details.

SECTION 1 - QUESTION ANSWER RATES
Section 1a - Answer Rates For Individual Questions
The criterion used here for assessing scale presentation effectiveness is the proportions of questions left unanswered (the “no answer” level for closed ended questions). The no answers to the questions in the different scale formats are shown in Table 1 and the results to note follow.

Table 1: Questions not answered, 2009.

<table>
<thead>
<tr>
<th>Format 1 (words)</th>
<th>Format 2 (numbers)</th>
<th>Format 3 (emoticons)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4%</td>
<td>3%</td>
<td>24%</td>
</tr>
</tbody>
</table>

1. In 2009, as Table 1 shows, the no answer rate from the emoticon scale presentations was 24% while in contrast 4% of word scale presentations and 3% of number scale presentations were not answered. These are statistically significant differences (d. of f. 2, \( \chi^2 = 888.2, p < 0.01 \)) and show the way the scale is presented has an effect on the no answer levels obtained.

2. Furthermore, the no answer rates for Format 2 (number scale presentations) is significantly lower than for Format 1 (word scale presentations) (3% cf 4%, \( z = 4.53, p <0.001 \)). This analysis is based on the number of questions, not on the number of people completing the survey, hence the difference between the 3% to 4% is statistically significant.

3. The no answer rates for questions other than the scale questions in 2009 was 4%. The result for Format 3 (emoticons scale presentations) is significantly higher (24%) and the Format 2 (numbers at 3%) is significantly lower.

In the 2009 survey, the financial products asked about in the scale presentations were not the same, and so it was surmised the difference in no answer rates may have been related to the financial products enquired about, rather than to the scale presentations used. Accordingly, in 2010, this possible effect was eliminated. Each format tested the same financial products, in the same order, differing only by scale presentation.

Section 1b - No Answer Rate In The 2010 Surveys
Table 2 shows the 2010 surveys’ no answer responses by Format.

Table 2: 2010 No answers to questions by scale type.

<table>
<thead>
<tr>
<th>Format 1 words</th>
<th>Format 2 numbers</th>
<th>Format 3 words only</th>
<th>Format 4 words only</th>
<th>Format 3 words &amp; emoticons</th>
<th>Format 4 words &amp; emoticons</th>
<th>Format 3 emoticons only</th>
<th>Format 4 emoticons only</th>
</tr>
</thead>
<tbody>
<tr>
<td>3%</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
<td>7%</td>
<td>10%</td>
<td>12%</td>
<td>22%</td>
</tr>
</tbody>
</table>

Each Format is a separate survey and is independent of the others and so the findings of each can also be analysed independently.

The higher no answer rates are still associated with the emoticon scale presentation Formats 3 and 4. In each case, the Formats that included the emoticons had significantly higher recorded no answer levels.

In addition, the no answer rate to the Format 2 numbers presentation is lower than the no answer rate to the Format 1 words scale presentation. This difference is statistically significant, and indicates number scale presentation formats perform better than word scale presentation formats in collecting answers and minimising no answer levels.

Tests:
1: % Format 1 is significantly less than % Format 2 (2% cf 3%, \( z = 5.57, p < 0.001 \)).
2: Format 3 words and emoticons has significantly worse no answers to questions than Format 2 numbers. (7% cf 2%, \( z = 84, p < 0.001 \)). We can conclude that Format 3 words and emoticons has worse no answer rates than Format 1.
3: Format 3 words only with Format 3 emoticons only. (2% cf 12%, \( z = 23.05, p < 0.001 \)).
4: Format 3 words and Format 4 words and emoticons. Format 3 has a significantly less no answer rate than Format 4. (7% cf 10%, \( z = 8.98, p <0.001 \)).
Section 1c - Emoticons And The Way No Answers Are Registered In The 2010 Surveys

As noted earlier, the emoticon scale image first seen by a respondent appears as a neutral image.

To ensure respondents did not just leave the slider in that neutral position, instructions in the questions told the respondents to move the slider to ensure that an answer is recorded, even if they wished to choose the neutral position. Failing to move the slider would result in no answer being recorded.

After reviewing the data in Table 2, it was hypothesised that respondents who wanted to report a neutral answer may have left the emoticon image unchanged (despite instructions to move it, because open ended comments explaining why the emoticon scale format was quickest to use included statements such as “If a neutral response was required, no movement of the bar was necessary”). The data in Table 3 shows the recorded neutral answers and also the combined totals of the neutral and no answer responses.

The data in Table 3 shows fewer respondents shown the Formats 3 or 4 emoticons scale presentations recorded a neutral answer than respondents shown Formats 1 or 2 word or number scale presentations. Table 3’s data indicates that Formats 3 and 4 emoticon scale presentations under-report the proportion of respondents who would have recorded a neutral answer if they had been presented with word or number scale presentations.

The combined totals for the neutral and no answer shown in Table 3 suggests that the hypothesis, that respondents shown the emoticons scale presentations who wanted to record a neutral answer just left the slider untouched, is plausible.

The data was examined further and the recorded no answers were combined with the answers that were expressly recorded as neutral. The no answer and the expressly chosen neutral responses in the emoticon scale presentations Formats 3 and 4 were similar to those of Formats 1 and 2 though the no answer levels differed.

As the combined no answer and expressly neutral proportions were much the same, regardless of scale presentation, it could appear that, for analytical purposes, it might be appropriate to combine the no answer and the expressly neutral responses from Formats 3 and 4, if emoticons as presented, are used in future. However, the need to combine answers to get close to the “right” proportions seems, at best, methodologically awkward as it allows for reporting of neutral levels that are knowingly under-represented and so would seem a most inappropriate practice. The need for analytical purposes to combine the no answer and expressly neutral answers to get answer levels similar to those achieved by using word or number scale presentations questions the value of using the Formats 3 and 4 emoticon scale presentations.

Table 3 shows one other scale presentation finding. The additional scale interval of “can’t say” in Format 4 increased the proportion of neither positively or negatively engaged respondents by 15%, to 41%, and so the inclusion or omission of a “can’t say” option in scales merits future review.

Table 3: Neutral and no answer responses, 2009 and 2010 surveys.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral answers</td>
<td>27%</td>
<td>31%</td>
<td>26%</td>
<td>29%</td>
<td>13%</td>
<td>19%</td>
<td>14%</td>
</tr>
<tr>
<td>Neutral and “no answer” combined</td>
<td>31%</td>
<td>34%</td>
<td>29%</td>
<td>34%</td>
<td>37%</td>
<td>27%</td>
<td>27%</td>
</tr>
<tr>
<td>Neutral, “no answer” &amp; can’t say combined</td>
<td>A can’t say option was not offered with these Formats, so the responses in the row above apply.</td>
<td>41%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SECTION 2 - RESPONDENTS' OPINIONS ABOUT SCALE PRESENTATIONS

In the 2009 survey only, respondents were asked to express their opinions about the scale Formats tested (Formats 1, 2, and 3, shown in Figures 8, 9 and 10). Section 2a looks at preferences for specific aspects of scale presentations and Section 2b looks at individual comments made by respondents when given the opportunity to answer by more than just ticking a box.

2a - Preferences For Specific Aspects Of Scale Presentations

Table 4 shows the questions and summarises the responses respondents gave to the three formats used in the 2009 survey.

The data in Table 4, when analysed, revealed significant differences in responses depending on whether scale presentation Format 1 words, Format 2 numbers or Format 3 emoticons were used for ratings 1 to 6 ($\chi^2 = 13.77$, d. of f. 2, p < 0.01). These differences are summarised below.

- Respondents preferred Format 1 word scale presentations. The averages for all seven statements showed that 38% of respondents rated Format 1 words most highly, 31% rated Format 3 emoticons most highly, and only 22% rated Format 2 numbers most highly.
- One statement “Made you think most about each question before answering” could, on balance, be seen as a positive or as a negative statement. The open ended comments in the surveys, and Couper, Tourangeau, Conrad, Singer 2006, indicate this possibility. Accordingly, the summary rating was recalculated with that statement omitted. The results then showed 40% of respondents rated Format 1 word scales presentation most highly, 21% rated Format 2 number scale presentation most highly, and 30% rated the Format 3 emoticon scale presentation most highly. Essentially, the recalibration had generated little change from the initial summation.
- The analysis of individual statements showed that more respondents reported that the Format 1 word scale presentation was better, while, in contrast, on each scale attribute statement, the Format 2 numbers scale presentations had the lowest proportion of respondents agreeing with the statement in each scale attribute statement.
- The Format 3 emoticons scale presentation was marginally more preferred on the “appealing to use” measure (39%) and also caused more respondents to think most about each question before answering (36%, compared with 27% for the word and 25% for the number scale presentations).

Table 4: Attributes of scales presentation formats, 2009.

| Attribute measured: | Preferred: |  |  |  |  |  |
|---------------------|------------|------------|------------|------------|------------|
| Scale method that was: | Format 1 words | Format 2 numbers | Format 3 emoticons | None of these | No answer |
| 1 Most preferred for use | 43% | 23% | 30% | 0% | 3% |
| 2 Easiest to use | 40% | 21% | 29% | 3% | 7% |
| 3 Quickest to use | 37% | 23% | 31% | 3% | 6% |
| 4 Most appealing to use | 35% | 18% | 39% | 5% | 3% |
| 5 Best expressed your feelings about the question | 40% | 20% | 29% | 8% | 3% |
| 6 One which gave the truest indication of your answers | 45% | 23% | 23% | 4% | 6% |
| Average ratings 1 to 6 | 40% | 21% | 30% | 4% | 5% |
| 7 Made you think most about each question before answering | 27% | 25% | 36% | 8% | 4% |
| Average all 7 ratings | 38% | 22% | 31% | 4% | 5% |
• On average, 9% of respondents did not answer questions about the scale presentation formats or specified “none of these”, but only 3% did not answer or said “none of these” when asked their scale presentation preference. The low “no answer” or “none of these” answer level suggests respondents had very clear views of the scale presentations they preferred to use.
• Tests relating demographics and self-perceptions to the Format preferences were undertaken but no statistically significant differences between respondent characteristics and scale presentation Formats were found. These results can be obtained on application to the author.

In sum, Table 4’s data shows that respondents’ scale preference is mostly for Format 1 word scale presentations. Accordingly, it would appear to indicate the scale presentation style researchers would be advised to use. However, the data from Table 2 indicates that there is a significantly better level of questions answered when number scale presentations are used.

The conclusion from this is that preference and more effective completion may not be the same. The question researchers thus have to address is whether it is better to use emoticon scales because people find surveys containing them more appealing to complete; or to use word scales, which people prefer; or to use number scales, which are more effective in minimising no answer levels.

2b - Open Responses About The 2009 Scales Formats 1, 2 And 3
At the end of the scale attribute section, respondents were asked if they wished to make any comments about the scales they had used.

Most made no comment (73%). The coded answers were:
• 6% reported they liked the survey and its variety of scale presentations (“It was good to have a variety of scales, it made it more interesting”);
• 6% expressly liked the emoticons (“I just loved the picture ones, keep using them!” “Only that I really enjoyed using the picture scale, it was very innovative and so much fun to use for a change to the usual ones”);
• 6% noted expressly that Format 3 emoticon scales made them think more; and
• 4% noted Formats 1 or 2 scale presentations made them think more; though
• In contrast, 3% expressly did not like the emoticons or the sliders (“Don’t like them because it doesn’t convey your exact answer”, “The picture one could use more work. It was very limited in expressing opinions, and the scale was too short.”); and
• 6% noted problems with the sliders (“The scales were a little difficult to position correctly.”, and “slider was time consuming and hard to interpret exactly how I felt.”);
• 6% of respondents noted scale presentation Formats 1, 2 and 3 were all difficult to use;
• A further 5% suggested survey improvements, including a not applicable or can’t say option to the scales (“I’d like a not relevant or not applicable option.”).
• One insightful respondent noted “very intelligently done”.

SECTION 3 - SURVEY ENJOYMENT AND FUTURE SURVEY INTENTIONS
Making a survey enjoyable is useful if it encourages more complete answers or encourages respondents to do surveys in the future. The five surveys each concluded by asking if the respondent had enjoyed the survey, and then how likely they were to do similar surveys in the future (after Sargeant, 2006). Table 5 reports those findings.

Test: Enjoyment for Formats 1 and 2 (average) is less than enjoyment for Formats 3 and 4 (with emoticons, average) is significantly lower. (73.46% cf 89.5%, z = 8.9, p < 0.001)

While more 2010 Format 3 and 4 scale presentation survey respondents said they enjoyed their survey than did respondents to Format 1 or 2 surveys (Table 5), the Formats 3 and 4 surveys had the highest “no answer” levels (Table 3), indicating enjoyment did not ensure more complete answers.

The scale presentation formats shown to respondents in the 2010 surveys had the same levels of “intention to do similar surveys in the future” indicating scale presentation formats used do not influence future intention, merely make the current survey more or less enjoyable.

The slightly lower level of intention to undertake surveys in the future, reported in 2009, may have been influenced more by that survey’s greater length (90 questions) than by the scales it contained.
In summary, while significantly higher proportions of respondents who enjoyed the survey were drawn from those shown the Formats 3 and 4 emoticon scale presentations, this did not translate into any greater belief that they were more likely to do similar surveys in the future.

SECTION 4 - WILLINGNESS TO PROVIDE RE-CONTACT DETAILS

As more Format 3 and 4 respondents enjoyed their survey than the Format 1 and 2 respondents, it was hypothesised that enjoyment would lead higher levels of provision of email addresses for entry to the “thank you” competitions. The evidence (shown in Table 6) did not support this hypothesis.

DISCUSSION

1 KEY FINDINGS

The research program was undertaken to strengthen online survey design by identifying the impact of three different scale presentation formats (words, numbers and emoticons). The key findings from the research were that:

1. Survey respondents markedly preferred word scale presentations to number or emoticon scale presentations.
2. Emoticon scale presentations were generally seen as “cute” and surveys containing emoticon scale presentations were enjoyed by more respondents than surveys containing word or number scale presentations, but
3. This enjoyment was self-contained and did not lead to increased intentions to do similar surveys in the future, and
4. The emoticon scale presentations were significantly associated with increased no answer levels and lesser quality scale question data (lower levels of neutral answer than other scale presentations and higher no answer levels).
5. The negative impact of emoticon scale presentations on scale question data was self-contained to those scale questions and did not influence the answer levels of non-scale questions in the same survey.
6. Number scales, while least preferred by respondents, were the more effective in generating low no answer levels.
7. Almost all respondents will answer closed ended questions when they can.
8. The inclusion of a “can’t say” option in scale presentations seems an improvement that allows respondents who do not have an opinion to answer accordingly.
9. The visual appeal of the emoticon scale presentations was offset by respondent difficulty in answer, as the respondents had to give more thought to emoticon scale answers and the answers the emoticons offered may not reflect the answers respondents wanted to give.

Different conclusions can be drawn from these studies for respondents and for researchers.
2 CONCLUSIONS FOR RESPONDENTS
For respondents, the studies indicate that word scale presentations are preferred to number scale formats particularly and to emoticons scale formats. The appeal of emoticon scale presentations appears to be more with the visual element than the answering facility. This in turn suggests a respondent preference for surveys that are visually engaging as well as data collecting, but that this enjoyment is contained to the survey itself.

Respondents also appear unaware of the impact of different scale presentations on their answer levels to scale questions and answered number scale questions more fully than word scale questions but still expressed a preference for word scale presentations.

An additional and positive result of respondents’ unawareness of the impact of scales on their answers is that scale presentation formats found to be ineffective for data collection do not effect the quality of answers to other closed ended questions in the same survey (a benefit for researchers as well).

3 CONCLUSIONS FOR RESEARCHERS
The first conclusion for researchers is that choice of scale presentation format will influence the quality of the data collected.

When choosing a scale presentation format, researchers can expect word scales will be providing more respondents with the scale presentations they prefer, though use of number scales will increase the level of data collected.

Emoticon scales as detailed are inappropriate for online survey use, because of the difficulty respondents have in answering such scales, and because the lower levels of answers obtained would provide lower quality data.

Scale presentation format weaknesses or advantages appear limited to the data collected for those specific scale questions and appear not to influence the answer quality of other closed ended questions in the same survey.

Emoticon scale presentations are not a solution to the problem of flat-lining because they create other data collection problems. Other design solutions to resolve that problem should be sought.

There appears a need to add a “can’t say” option to scales to reflect the answer needs of respondents who, for whatever reason, cannot answer the question and so, in the absence of the “can’t say” option, are pushed to give an answer they really do not mean.

Respondent enjoyment with one survey appears not to influence future survey intention but it may be a selling point in online survey invitations to be able to say that more respondents have enjoyed similar such surveys.

The conclusions are that researchers may be better served overall by using word scale formats than number scale formats, and should definitely avoid using emoticon scale formats (as presented).

4 OTHER ISSUES FOR FUTURE RESEARCH
The research process and the data obtained from that indicate some areas for future research. These include the use of pictures and other visuals in online surveys to make them more appealing, the use of a slider without the emoticon images, and using a word or number scale with emoticon images instead of the slider, and the impact of adding a “can’t say” answer interval on scale presentations.

Some may wish to test emoticon images if shown as the scale intervals with “tick-the-box” answer formats, and some may wish to test emoticon format scale presentations which show all image options at once, or have the start positions other than neutral. It may be worth testing the use of images of real human faces showing different expressions in place of the cartoon characters used.

The findings from this research may encourage testing in other sectors than finance and suggest the need to review other online survey engagement techniques (such as ranking pre-set answers, moving boxes to different spots to record answers, box slider scales with number answers shown at the side, etc.) to determine the impact of those on the data collected.
For brevity, the examples that follow show the emoticon scale presentation format that did not include the “Can’t say” option, as they were exactly the same in presentation, except for the inclusion of the “can’t say” option, shown above.
REFERENCES


AMSRO and AMSRS. (2010). Australian market and social research industry survey 2010. PDF supplied by email July 21, 2010 and sent to participant companies.


Sergeant J. Doctor Strange Results, or how I learned to stop worrying and love non-response, AMSRS Conference, 2006, supplied privately.

ACKNOWLEDGEMENTS

The author notes with gratitude the particular and extensive assistance of Judy Milne-Pott, Sue Graham, Jeevan D’Almeida and Celia Derham in assessing these issues but notes that errors and omissions are his and his alone.
Handling “Don’t know” responses in parallel CATI and online surveys

Don Porritt, Senior Research Director, Taverner Research and Adjunct Associate Professor, Charles Sturt University
Benjamin R. Marx, Senior Research Officer, Independent Commission Against Corruption

Abstract
Survey respondents are more likely to select a “Don’t know” (DK) or other non-committal response option when it is explicitly offered to them compared with it merely being available as a valid response. In CATI surveys, such response codes are normally not read out, but can be accepted by interviewers. This approach is problematic for online surveys, especially if these are to be compared to results from CATI surveys. If the non-committal reply is presented on screen, it will attract more responses from respondents who are satisficing. However, insisting that respondents choose a definite reply will produce zero “don’t know” replies, and can result in both invalid answers and respondents dropping out of the survey part way through. A possible solution is to display a question online initially without non-committal options, and then to re-display with these options if the respondent clicks on “Next” without recording a reply. A survey was administered online (n=500) and by CATI (n=501) to well-matched samples of NSW residents, using this procedure for most Yes/No and all pre-coded rating scale items. For these items, this procedure resulted in markedly lower endorsement of non-committal replies in the online sample compared with the CATI sample. Online panel respondents almost always recorded a response even if there is no non-committal option shown on screen. It is likely that online panel members have been effectively trained to always select one of the offered responses before they click the “Next” option. If so, it will be difficult to produce equivalent use of DK responses when comparing CATI survey results with those from an online research panel. Suggestions for how to modify both methods to improve comparability are offered.

INTRODUCTION
There is now a substantial literature about method effects that produce differences in response distributions between surveys conducted by CATI and online. Couper (2000) outlines a range of reasons to be sceptical about the quality of data collected from internet survey panels. Reports of direct comparisons of results have also begun to appear in peer-reviewed journals (e.g., Berrens et al, 2003; Chang and Krosnick, 2008; Fricker et al 2005).

One issue that has not attracted much attention, however, is the difference in the rates of non-committal replies such as “Don’t know” (DK) or “Not applicable”.

Krosnick’s 1999 review of published research on survey research methods concluded, based on eight published studies that:

“Many more respondents say they have no opinion on an issue when this option is explicitly offered than when they must volunteer it on their own” (Krosnick, 1999, p. 557)

We have found no published work since that gives grounds to question that conclusion.

To reduce such “easy” responses, CATI interviewers normally do not read out non-committal response options, but can select them as valid response categories if specific respondents appear genuinely unable to select between the prompted replies. Not explicitly offering non-committal replies reduces the frequency of “satisficing” (giving the easiest response rather than the response that best expresses the respondent’s actual position after giving the item necessary thought; Krosnick and Presser, 2010).

Online surveys typically do not let a respondent progress to the next screen unless a reply has been recorded to all questions on the current screen. This parallels the requirement for interviewers in a CATI survey to record some response for an item before moving to the next item. Since some respondents genuinely have difficulty in answering particular questions, a DK option is normally displayed in online surveys, as there is no interviewer to decide that such a reply should be recorded if it is volunteered. However, this can result in relatively high rates of such replies being given, probably by respondents who are “satisficing” (Krosnick and Presser, 2010). This might be especially common with online respondents who are part of survey panels and are doing surveys as...
much for access to incentives as from an interest in giving their opinions.

An example of the size of this effect is provided by results published by Biotechnology Australia (Eureka, 2007). The survey had DK displayed on screen for the online survey but did not read DK out to CATI respondents. DK was more frequently selected by online participants for each of six items asking about the future impact of new technology developments on “our way of life”. The average percentage of DK replies was 21.5% (range 10% to 33%) for online respondents compared with 11.7% (range 4% to 18%) for CATI interviewees (see Table A2, p39 in Eureka, 2007 for the item by item data). The authors of the report predicted that there would be differences in CATI and online responses:

“In this survey, the DK response option was necessarily presented differently in the two survey versions for the majority of questions. In the CATI version, the ‘don’t know’ option was (historically) not read out to participants for the majority of questions, although it was accepted if offered voluntarily by the participant. In the online version of these questions, the ‘don’t know’ option was necessarily visually presented to participants in order for participants to have this response as an option. As a result, we would expect a greater proportion of ‘don’t know’ responses among participants in the online version and a subsequent decline in other responses.” (Eureka, 2007 p. 35)

After presenting the results summarised above, they remarked that:

“The impact of the differential presentation of the ‘don’t know’ response option is clear. There is a much higher incidence of ‘don’t know’ among online participants, which is the direct result of the presentation of this option in the online version and the lack of verbal presentation of this option in the CATI version. There is a corresponding reduction in other response options among online participants.” (Eureka, 2007 p. 39)

The effect was also evident for six Yes/No questions about awareness of biotechnology applications (average percent DK was 15.0 online and 0.8 for CATI, Eureka, 2007:40) and six Yes/No questions asking whether these applications were useful (average percent DK 19.2 for online and 3.7 for CATI, Eureka, 2007 p. 40).

These results are quite typical of Taverner’s own experience with parallel CATI and online surveys, and are consistent with the general principle that a response will be more common when included in a list of prompted replies than when it is not stated but can be recorded if volunteered. This extends beyond the evidence reviewed by Krosnick (1999) about the effect of offering a “don’t know” option (for example see Presser, 1990). Bishop et al (1988) confirmed the effect of offering or not offering a “don’t know” reply in both telephone and mail self complete versions of a survey in both US and German samples. In the same surveys Bishop et al (1988) showed that respondents were much more likely to indicate they took a middle position between two opposite options when an explicit “middle” category was offered.

From these results and a review of comparisons between open and closed versions of the same question, Krosnick (1999, p. 544) concluded that:

“That is, people generally ignore the opportunity to volunteer a response and simply select among those listed, even if the best answer is not included. Therefore, a closed-ended question can only be used effectively if its answer choices are comprehensive, and this is difficult to assure.”

Thus, the effect is not confined to online versus telephone administration, but is also evident within telephone surveys and within mail self-complete surveys and extends to other response options in addition to DK replies.

An explanation for this effect is that there is much less cognitive effort in endorsing explicit options than in volunteering a reply that is not explicitly offered. In particular any respondent who is inclined to take a lower-effort option when replying will be more likely to choose between the options offered than to volunteer that none of them apply.

In the typical CATI procedure, in effect the respondent is asked to choose between the prompted alternatives but allowed to provide a non-committing response. Interviewer briefing often emphasises that such replies should be probed and not accepted without question, and such options are rarely read out.
It appeared possible that this CATI process could be approximated online by initially presenting questions without a DK response option. If the respondent clicked on Next without recording a response, they would be taken to a new version of the question with a DK option displayed.

Based on the extensive evidence that not prompting DK reduces the rate at which it is recorded, we predicted that the alternative procedure would reduce the rate of DK responses in an online survey to be no higher than found in a CATI version of the same questionnaire in which DK is not read out but is available to the interviewer if the response is volunteered by a respondent.

**METHOD**

Taverner Research was contracted by the NSW Independent Commission Against Corruption (ICAC) to conduct parallel CATI and online general population surveys regarding perceptions of corruption-related topics in NSW. This provided an opportunity to test the effects of the suggested online procedure on the incidence of non-committal replies.

Samples of 501 (CATI) and 500 (online) completed interviews were obtained (ICAC, 2010). The online survey was conducted to test the relative validity of data obtained by the two methods. In Taverner’s proposal to conduct the study, the issue of the different rate of DK replies was explicitly raised, and a procedure proposed to reduce the effect. This procedure was implemented to test how effective it would be in achieving similar DK rates from the two data collection methodologies.

The CATI sample approached a stratified random sample of households with residential phone numbers and modified the rules for selecting respondents within multi-adult households to match the sample to age by gender targets for metropolitan and regional areas of NSW. The online sample was closely managed to also ensure matching to the general population targets by location within NSW, gender and age group. The resulting samples were well matched on these demographic variables and close to the population distribution. Although similar in highest completed level of education, both were higher than the population in the percentage with a university degree (see Appendix for the distributions by gender, age group and education).

In the online survey, twelve pre-coded questions that were answered Yes/No or with a short list of mutually exclusive pre-coded replies were subject to a two-stage procedure described below.

First, each item was displayed without any non-committal responses shown on screen. Respondents who went to the next screen without recording a reply were shown the question again and asked to record a reply, with the appropriate non-committal options now available.

Some thought was given to whether online respondents should be informed that they can gain access to a DK option by clicking “Next” without recording a reply. Since no parallel instruction is given in CATI surveys, it was decided not to alert respondents to the possibility. Thus to access DK, they would have to find enough difficulty in giving a reply from those that were prompted to click Next without recording a reply, and then discover that this gave them access to a DK option. Respondents were not alerted to the possibility of going to the next screen without recording a reply. 

**RESULTS**

The number and percentage of DK and other non-committal replies to the 12 pre-coded questions asked of all respondents can be found in Table 1.

**Table 1. Non-committal replies to 12 pre-coded questions by method**

<table>
<thead>
<tr>
<th>ITEM</th>
<th>CATI (n=501)</th>
<th>ONLINE (n=500)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NUMBER (%)</td>
<td>NUMBER (%)</td>
</tr>
<tr>
<td>A DK</td>
<td>28 (5.6)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>B DK</td>
<td>31 (6.2)</td>
<td>1 (0.2)</td>
</tr>
<tr>
<td>C DK</td>
<td>20 (4.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>D DK</td>
<td>16 (3.2)</td>
<td>1 (0.2)</td>
</tr>
<tr>
<td>E DK</td>
<td>25 (5.0)</td>
<td>5 (1.0)</td>
</tr>
<tr>
<td>F DK</td>
<td>10 (2.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Depends</td>
<td>16 (3.2)</td>
<td>1 (0.2)</td>
</tr>
<tr>
<td>G DK</td>
<td>39 (7.8)</td>
<td>2 (0.4)</td>
</tr>
<tr>
<td>H DK</td>
<td>8 (1.6)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>I DK</td>
<td>64 (12.8)</td>
<td>3 (0.0)</td>
</tr>
<tr>
<td>J DK</td>
<td>7 (1.4)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>K DK</td>
<td>22 (4.4)</td>
<td>1 (0.2)</td>
</tr>
<tr>
<td>L DK</td>
<td>17 (3.4)</td>
<td>6 (1.2)</td>
</tr>
<tr>
<td>MEAN</td>
<td>25.2 (5.1)</td>
<td>1.7 (0.3)</td>
</tr>
</tbody>
</table>
The online procedure produced a consistently lower incidence of non-committal replies to these questions. However, the responses to the different items are not statistically independent of each other, and thus conducting multiple significance tests is questionable.

It would be possible to simply count the number of items given a DK reply and test for the significance of the difference in the mean number or apply another non-parametric test of significance if the distribution is skewed. However, this assumes (rather than proves) that the tendency to give DK replies is an internally consistent individual difference variable.

To check whether the use of DK replies on different questions occurs independently or represents some underlying individual difference variable, a principal components analysis was used to identify any groupings of items for which giving a non-committal reply was related. This produced three factors on which summated scales could be based – one involving three items with the same set of 3 response codes (items A, B and C); another combining two binary items (items D and J); and another combining three items with four point scales (items F, H and K). The detailed results of this analysis are shown in the Appendix to this paper. Correlations between counts of non-committal responses on these scales and the four items that did not load on the three factors ranged from zero to +0.16. While some of these correlations are statistically significant, they indicate very little shared variance (<3%). Thus, for all practical purposes, scores on the three factor based scales and the frequency of DK responses on the remaining four items can be treated as statistically independent, and significance tests applied as if the results came from independent observations.

Given the highly skewed score distribution, the scores of the CATI and online sample were compared using contingency tables and chi-square tests of significance. Scores on the three scales were collapsed into two categories, those with no DK replies, and those with one or more. On the individual items that did not enter into scales, the frequency of DK replies was compared.

All the chi-square values were significant, with the CATI sample being much more likely to produce non-committal replies than the online sample. All expected frequencies were >5.0, so no continuity correction was required. The probability of obtaining seven such results by chance given that the items show little correlation is extremely remote, so we can reasonably conclude that the differences are real.

<table>
<thead>
<tr>
<th>ITEM</th>
<th>CATI (n=501)</th>
<th>ONLINE (n=500)</th>
<th>Chi Sq (1df)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NUMBER (%)</td>
<td>NUMBER (%)</td>
<td></td>
</tr>
<tr>
<td>Scale 1 DK&gt;0</td>
<td>56  11.2</td>
<td>1  0.2</td>
<td>56.16 (p&lt;.001)</td>
</tr>
<tr>
<td>Scale 2 DK&gt;0</td>
<td>21  4.2</td>
<td>1  0.2</td>
<td>18.55 (p&lt;.001)</td>
</tr>
<tr>
<td>Scale 3 DK&gt;0</td>
<td>45  9.0</td>
<td>2  0.4</td>
<td>41.19 (p&lt;.001)</td>
</tr>
<tr>
<td>E DK</td>
<td>25  5.0</td>
<td>5  1.0</td>
<td>13.70 (p&lt;.001)</td>
</tr>
<tr>
<td>G DK</td>
<td>39  7.8</td>
<td>2  0.4</td>
<td>34.74 (p&lt;.001)</td>
</tr>
<tr>
<td>I DK</td>
<td>64 12.8</td>
<td>3  0.0</td>
<td>59.39 (p&lt;.001)</td>
</tr>
<tr>
<td>L DK</td>
<td>17  3.4</td>
<td>6  1.2</td>
<td>5.36  (p&lt;.02)</td>
</tr>
</tbody>
</table>
DISCUSSION

Clearly the method adopted to reduce non-committal replies in the online self-completed questionnaire was “too effective”.

One possible explanation for this is that the online panel respondents would mostly have done other surveys which did not allow them to go to the next screen until they had selected a reply to the item(s) on the current screen. It could be that they had been effectively trained to nearly always click an answer and thus even those who were somewhat unsure did not discover that they could gain access to a version of the item with a DK option. As noted in describing the method, we did not alert online respondents to the possibility of not selecting an answer, and that they would then have access to a DK reply, as this is not made explicit in CATI surveys. It appears that very few discovered the possibility by attempting to go to the next screen without recording a reply.

Giving such an instruction might improve comparability with CATI, but could invite an increase in recourse to DK replies as an easy option for respondents who are satisficing. This might be decided by further research with parallel surveys using the method tested here, modified by alerting online respondents to the possibility of giving a DK reply by first clicking “Next” without selecting one of the listed replies. To achieve full comparability of process, CATI respondents would also need to be told that if they really cannot decide on the answer they should inform the interviewer. This could, however, increase the rate of DK replies in the CATI mode and reduce comparability of results with previous CATI surveys in tracking studies such as the one reported here.

A variation on this approach would be to warn respondents that if they reply “Don’t know” they will be asked to explain what they are unsure about. This could discourage selection of the option as an “easy answer” and so might mitigate the risk that reminding respondents of the possibility of giving a DK reply might increase the rate of such replies in the CATI mode.

Another option would to include a response option labelled “Other” for each item in the online version. Those who were simply trying to answer as quickly as possible would be unlikely to select this, as they would (if experienced panel members) know that a request to specify is likely to follow. Those who are genuinely having difficulty selecting from one of the options offered because they do not know how to choose between them would be quite likely to choose it. They could then be offered a question not shown to other respondents with options for DK, NA, and Other (Specify).

It could also be informative to compare the responses of “novice” online respondents (whether panel members doing their first survey, or respondents doing a survey online who are not members of a survey panel) to those of panel-member “veterans”.

Chang and Krosnick (2008) concluded that online administration results in higher quality data than CATI when both use randomly recruited respondents. This conclusion might not apply to online panel respondents who have been recruited in non-random ways, and are perhaps more motivated than respondents to CATI surveys or randomly recruited online respondents by the chance of gaining rewards for completing surveys and less motivated by intrinsic interest in the topics or by willingness to be helpful to researchers.

It is of course possible to “filter out” those who give a DK reply and only examine the distribution of replies from those willing to give a definite opinion. One question that arises is whether having an explicit DK option alters the relative distribution of other replies. Krosnick et al (2002) report evidence that the distribution of other replies to questions remain very similar when data collected without a DK option is compared to data from those who had such an option but did not use it. We have not examined this issue in this paper. However, if the incidence of giving DK replies is high, this can result in a loss of sample size that reduces the power of the sample to detect relationships between items. The evidence reviewed by Krosnick et al (2002) also suggested that not allowing a DK option does not reduce the strength of relationships with other items. Thus they advised that DK options should not be offered.

The data reported here do not provide any indication of what would result if the online respondents had been given an explicit DK option when first shown the items on screen. Data from some other items suggests that, at least in this sample, online panel respondents then give much higher rates of DK replies.

We have not explored demographic differences between those who gave more DK replies and those...
who gave fewer DK replies. As the factor analysis results showed that there were three independent factors plus four items not loading on any of the factors, it appears that some complex interaction between item characteristics and individual differences determines who makes more than one such response, and that this would not be determined by respondent characteristics alone. Further exploration of this issue was outside the scope of this paper.

CONCLUSION
The online procedure tested is not equivalent to the usual CATI procedure of not reading out a non-committal option while making it available for the interviewer to use if respondents appear genuinely unsure. While even more interviewer training and closer supervision could also reduce non-committal replies to CATI questions where such options are not read out, it appears that it is difficult to ensure the two methods produce similar levels of non-committal responses, whether these are offered explicitly or not.

Further research is needed to test whether results would be more comparable if both CATI and online respondents were told they can choose to say they do not know the answer to an item – in CATI by telling the interviewer, and online by clicking Next without selecting a response. However, we predict that while these procedures might produce more comparable results, they will also produce higher rates of DK responses and could destroy comparability with data collected by CATI without mentioning that DK replies can be given.

APPENDIX:
Sample Demographics

<table>
<thead>
<tr>
<th>ITEM</th>
<th>CATI (n=501)</th>
<th>ONLINE (n=500)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NUMBER (%)</td>
<td>NUMBER (%)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>255 (50.9)</td>
<td>247 (49.4)</td>
</tr>
<tr>
<td>Female</td>
<td>246 (49.1)</td>
<td>253 (50.6)</td>
</tr>
<tr>
<td>Age group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>58 (11.6)</td>
<td>58 (11.6)</td>
</tr>
<tr>
<td>25-44</td>
<td>187 (37.2)</td>
<td>186 (37.2)</td>
</tr>
<tr>
<td>45-64</td>
<td>163 (32.5)</td>
<td>165 (33.0)</td>
</tr>
<tr>
<td>65+</td>
<td>93 (18.6)</td>
<td>91 (18.2)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degree</td>
<td>150 (29.9)</td>
<td>143 (28.6)</td>
</tr>
<tr>
<td>VET</td>
<td>113 (22.6)</td>
<td>121 (24.2)</td>
</tr>
<tr>
<td>Upper 2ndry</td>
<td>92 (18.4)</td>
<td>96 (19.2)</td>
</tr>
<tr>
<td>Less</td>
<td>146 (29.1)</td>
<td>140 (28.0)</td>
</tr>
</tbody>
</table>

Factor analysis of DK replies.
Items were recoded so that DK had the value 1, and any other reply zero. A principal components analysis was conducted using SPSS with varimax rotation. Five components had eigen values greater than 1.0. These were (in order) 2.00, 1.39, 1.20, 1.10, and 1.03 accounting for up to 56% of the total variance. After testing rotations of five, four, three and two components, a solution using three (accounting for 38% of the total variance) was adopted. This is consistent with application of the scree test and produced the most interpretable structure.

The loadings on the rotated factors are shown in Table 3.
Table 3. Rotated factor loadings

<table>
<thead>
<tr>
<th>ITEM</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.75</td>
<td>-0.01</td>
<td>-0.04</td>
</tr>
<tr>
<td>B</td>
<td>0.75</td>
<td>-0.02</td>
<td>0.06</td>
</tr>
<tr>
<td>C</td>
<td>0.69</td>
<td>0.10</td>
<td>0.15</td>
</tr>
<tr>
<td>K</td>
<td>0.00</td>
<td>0.70</td>
<td>-0.04</td>
</tr>
<tr>
<td>H</td>
<td>0.05</td>
<td>0.66</td>
<td>0.14</td>
</tr>
<tr>
<td>F</td>
<td>0.10</td>
<td>0.58</td>
<td>0.03</td>
</tr>
<tr>
<td>J</td>
<td>0.02</td>
<td>-0.10</td>
<td>0.73</td>
</tr>
<tr>
<td>D</td>
<td>0.05</td>
<td>0.07</td>
<td>0.70</td>
</tr>
<tr>
<td>E</td>
<td>0.12</td>
<td>0.24</td>
<td>0.28</td>
</tr>
<tr>
<td>G</td>
<td>0.31</td>
<td>0.10</td>
<td>-0.30</td>
</tr>
<tr>
<td>I</td>
<td>0.28</td>
<td>0.31</td>
<td>0.02</td>
</tr>
<tr>
<td>L</td>
<td>-0.03</td>
<td>0.34</td>
<td>0.02</td>
</tr>
</tbody>
</table>

The factor structure appears to be determined in part by the format of the items. The first rotated factor involved three items with the same set of 3 response codes (items A, B and C). The second combined three items with four point scales (items F, H and K); and the third combined two binary items (items D and J).

Much of the correlation between items in the use of the DK response will be due to the variation between the two samples, as the DK rates were so low in the online sample. It appears that one sub-group of individuals tended to volunteer DK for the items with the same three response codes (A, B and C, all answered Major, Minor, None or DK); another for the two binary items (D and J, both answered Yes, No or DK) and another for the items with four point scales (F, H and K). While these groups overlap, they do not do so above chance levels. This produces the three factors shown above. There were items in the Yes/No/DK format and others with four point scales that did not have substantial loadings on the corresponding factor, so more is involved than the item format. It is beyond the scope of this paper to further explore the basis for such groupings.

REFERENCES


Eureka (2007) Community attitudes to biotechnology report on overall perceptions of biotechnology and general applications Canberra ACT: Biotechnology Australia.


Assessing Flat-Lining Response Style Bias in Online Research

Con Menictas*, University of Newcastle
Paul Wang, University of Technology Sydney
Brian Fine, Australia Online Research

Abstract
As online panels increase in number and more respondents view online panel membership as an alternative source of income, online research data quality is being called into question. Industry quality assurance standards attempt to increase the quality of online panel data, but these are only as effective as the degree of willingness of online panel providers to embrace the standards. One type of response style bias that is being increasingly observed in online research is known as flat-lining response style, where respondents exhibit low variability across rating scale items. This paper proposes one method for detecting flat-lining respondents and demonstrates the impact of flat-lining response style bias on market segmentation and brand equity modelling. In addition, the paper discusses the best practice regarding the recruitment of online panelists and recommends how to increase online panel data quality, via strict recruiting standards and better management of online panelists.

INTRODUCTION
As a profession, market and social research needs to be able to measure human attitudes and behaviour validly. Despite the rapid migration to online data collection, which has become the dominant methodology in Australia and other parts of the world, surprisingly there has been little research into possible biases in online research data. Most of the quality assurance standards focus on processes of managing a panel, such as ISO 26362. In Australia in particular, ISO 26362 was taken to a higher level with Quality Standards for Online Access Panels (QSOAP) Gold, which recognised the need for a Qualified Practicing Market Researcher (QPMR) to be directly involved in the end to end online research process. For an online panel company with ISO 26362 accreditation to be a qualified member of AMSRO, it also needs to meet ISO 20252 standard for market research.

The reason for these recommendations is that quality data cannot be achieved without careful panel management (Carleton 2006; Miller and Subbiah 2011). The pursuit of quality data also requires an understanding of the whole research process for online data, from online sample recruitment to careful management of the relationship, to weighting and modelling the responses. All of these processes need to deal with the issue of potential online data biases. In order for online sample data to represent the target population, we need to account for the following:

• individuals in the population who are not represented by online users;
• individuals in the online user population who will not join a panel;
• individuals in an online panel who do not usually participate in surveys; and
• individuals in an online panel who participate in surveys so frequently that it affects the way they respond to surveys.

Critics of online research recognise that online panelists are obtained via multiple phases and sources in the recruitment process (Hartmann 2011). Impending biases can be monitored and corrected for when the profile of potential online panel candidates is known at each phase. The correction of biases is mostly remedied at the demographic level via various weighting schemes.

Ideally, online panel respondent recruitment should be conducted via offline as well as online recruitment methods, in order to increase the likelihood that everyone in the target population has an equal chance of being selected. However, most panel providers do not recruit online panel membership via offline recruitment methods. They tend to recruit via online only, as it is far less costly. This causes their online panel to be skewed towards heavy online users when compared to an online panel that has been built using both online and offline recruitment methods.

* Corresponding author: Dr. Con Menictas. Email: con.menictas@newcastle.edu.au, Postal Address: Newcastle Business School, Faculty of Business and Law, University Drive, Callaghan NSW 2308 Australia.
A few of the main online panel biases that might exist in online panel data are as follows:

1. Panel recruitment bias, where door-to-door, CATI and online can all produce differing types of recruits;
2. Incentive bias, where the nature of the incentive can limit those that participate;
3. Questionnaire design bias, which is a bias caused by the questionnaire being difficult for respondents to complete; and
4. Acquiescence response bias (Podsakoff et al. 2003; Watson 1992), which is related to flat-lining, also known as straight-liners (Carleton 2006; Hartmann 2011; Herzog and Bachman 1981) in the case of rating scale data, or yeah-saying in the case of binary data.

In this paper we assess the impact of flat-lining response style bias on research outputs for banking performance data. We identify the profiles of flat-liners and illustrate how best practice in online panel recruitment and management can reduce the likelihood of encountering flat-lining respondents.

An easy way of detecting flat-lining respondents from online panel data is by examining within-subject variability. When there is little or no variation within a respondent, the respondents are deemed flat-liners. As an example, on a series of five Likert-type rating scales, ranging from one through seven, flat-lining respondents would produce ratings similar to the invariance response pattern observed in Table 1.

Table 1: Example of Flat Lining Respondents Using a 7-point Likert Type Rating Scale

<table>
<thead>
<tr>
<th>Respondent ID</th>
<th>Rating Scale 1</th>
<th>Rating Scale 2</th>
<th>Rating Scale 3</th>
<th>Rating Scale 4</th>
<th>Rating Scale 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale Direction</td>
<td>(+)</td>
<td>(-)</td>
<td>(+)</td>
<td>(+)</td>
<td>(+)</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

By contrast, non-flat-lining respondents would produce similar patterns to those observed in Table 2.

Table 2: Example of Non-Flat Lining Respondents Using a 7-point Likert Type Rating Scale

<table>
<thead>
<tr>
<th>Respondent ID</th>
<th>Rating Scale 1</th>
<th>Rating Scale 2</th>
<th>Rating Scale 3</th>
<th>Rating Scale 4</th>
<th>Rating Scale 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale Direction</td>
<td>(+)</td>
<td>(-)</td>
<td>(+)</td>
<td>(+)</td>
<td>(+)</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>3</td>
<td>7</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

The remainder of this paper is organised as follows. We first discuss the best practice regarding the recruitment of online panels and management of online panellists. We then assess the impact of flat-lining response style bias using the Erdem and Swait (1998) brand equity framework. Finally, we discuss the implications of this study for industry practice and avenues for further research.

2 RECRUITING RESPONDENTS FOR ONLINE PANELS

AMSRO endorses ISO 26362 and has established the QSOAP Gold standard for quality recruitment of online panellists, which is based on international standards and guidelines including ESOMAR 27 and ISO/DIS 26362 for internet access panels in market, opinion and social research. The scope of the Standard deals with online panels that are used exclusively for the purposes of market research and it provides for:

- ISO 26362 which deals with access panels in market, opinion and social research; and
- QSOAP Gold Standard which provides a standard reflective of ISO/DIS 26362 with some additional strengthening of the ISO criteria; and controls considered by AMSRO to reflect best industry practice (AMSRO 2008).

The Standard states that online panels which seek membership to QSOAP need to be actively managed, meaning that the online panel provider needs to maintain ongoing communication between themselves and the panellists. The Standard only applies to online panels used for market research.
In order to seek accreditation as a QSOAP approved online panel provider, the Standard requires fulfilment of the following clauses:

1. Clause 4.1: General Requirements;
2. Clause 4.2: Organisation and Responsibilities;
3. Clause 4.3: Recruitment of New Panel Members;
5. Clause 4.5: Access Panel Management;
6. Clause 4.6: Access Panel Usage;
7. Clause 4.7: Client Reporting;
8. Clause 4.8: Professional Rules of Conduct; and

We focus on two QSOAP Clauses for purposes of better validating online data and avoiding heavy online users:

1. 4.3.5 Offline Recruitment regarding offline versus online recruitment for online panels; and
2. 4.4.3 Profile of Panel Members regarding survey frequency.

QSOAP Clause 4.3.5 specifies that a minimum of 5% of the online panel’s recruitment should be conducted offline. The offline recruited panellists are to be identified in the online panel’s database and made available to the client when requested. Best practice suggests a minimum of 10% of the online panel should be recruited via offline. The reason that Clause 4.3.5 asks for a 5% minimum offline recruitment is to provide clients with a check and balance when they are buying online panel data, that the data are not skewed towards heavy online users.

QSOAP 4.4.3 specifies that among other things, participation rates of active online panellists shall be made available to clients. The clause aims to identify respondents that may be overexposed to online surveys.

Adhering to QSOAP helps avoid poorly recruited respondents with heavy internet usage. These respondents could well be members of multiple panels, doing an excessive number of surveys in a short time period with the aim of maximising incentives, all of which increase the potential for flat-lining response style bias.

We next discuss best practice regarding the management of online panels.

3 THE MANAGEMENT OF ONLINE PANELS

Clause 4.5 Access Panel Management suggests best practice of managing an online panel. The clause deals with record keeping of online panel organisation competencies and qualifications. It stipulates the responsibilities of the online panel to supply the client with this information when requested. The type and management of incentives need to be maintained in the online panel records for assessment of online panel quality.

Online panels subscribed to ISO 26362 and/or QSOAP Gold, have been audited on professional management processes with regard to record keeping and transparency to clients on the following aspects:

- over use of panellists which leads to flat-liners;
- removal of panellists found to be repeat offenders who:
  - provide low quality data such as flat-lining respondents;
  - provide low quality open-ended feedback;
  - speed through questionnaires in less than half the average time taken for the rest of the sample;
- ensuring that the online panel reflects the desired population;
- replenishing population groups that tend to:
  - unsubscribe;
  - drop off the online panel; or
  - do not participate in surveys over a 12 month period.

Analysis of online panel data should recognise the biases of poor research design. In order to illustrate the impact of the flat-lining response style biases, which is the focus of this study, we examine a recent banking case study. Our case study involves one of the top four banks in Australia. The study was conducted in 2010.

4 THE IMPACT OF FLAT-LINING RESPONSE STYLE BIAS

4.1 The Theoretical Framework
In order to contextualise our discussions to a real market situation, we use real online data collected for a banking client. The client is one of the top four banks in Australia and wanted to test the positioning
of the bank relative to the competition regarding brand equity. The study's focus was on daily transaction accounts for business customers.

Although there are a number of brand equity frameworks in the brand equity literature, this study used the Erdem and Swait (1998) brand equity framework as depicted in Figure 1.

**Figure 1: The Revised Erdem and Swait (1998) Brand Equity Framework**

The Revised Erdem and Swait (1998) brand equity framework uses signalling theory (Spence 1974). Signalling theory is derived from information economics where markets are characterised by imperfect and asymmetric information (Stigler 1961). Asymmetric information exists when one participating economic agent, e.g., a firm, knows more about their product than other agents, e.g., consumers. Imperfect information refers to a lack of complete information when evaluating product attributes (Nelson 1970).

Imperfect and asymmetric information leads to uncertainty, which in turn influences consumers' perceptions of brand attributes. Uncertainty about product quality also suggests that consumer beliefs may vary from person to person on the aspect of quality. This creates consumer perceived risk, which is something consumers like to avoid. Risk-averse consumers are not comfortable with ambiguous and uncertain product quality assessments. When the quality is uncertain, consumers are likely to search for more information. Erdem and Swait (1998) argued that consumers can use brands as a signal for quality. Brand credibility is hypothesised to be the key antecedent or mediator to brand quality, brand perceived risks and brand information costs. According to Erdem and Swait (1998), higher perceived quality, lower information costs, and lower perceived risks associated with credible brands can increase the expected utility of that brand.

We used seven Likert-type rating scales with increments one through ten \( \{1, 2, \ldots, 10\} \) for each of the seven constructs (Table 3) predicting the Expected Utility. A 10-point Likert type (1932) scale was used instead of an odd-numbered scale because 10-point rating scales provide greater dispersion and lesser positive skew than 5-point or 7-point rating scales (Krosnick and Leandre 1997; Valerie and Ritter 2007). Construct validity (Bagozzi et al. 1991) was established using confirmatory factor analysis and both procedural and statistical remedies were undertaken to minimise common method bias (Podsakoff et al. 2003; Spector 2006).

**Table 3: Erdem and Swait (1998) Brand Equity Constructs**

<table>
<thead>
<tr>
<th>Erdem and Swait (1998) Brand Equity Framework Constructs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Brand Investments</td>
</tr>
<tr>
<td>2. Consistency</td>
</tr>
<tr>
<td>3. Customer Service</td>
</tr>
<tr>
<td>4. Credibility</td>
</tr>
<tr>
<td>5. Perceived Quality</td>
</tr>
<tr>
<td>6. Perceived Risk</td>
</tr>
<tr>
<td>7. Information Costs</td>
</tr>
</tbody>
</table>

4.2 The Data and Analysis

The sample comprised 733 respondents that were customers of the bank and had an active savings account with the bank for at least 12 months. Respondents were solicited from the general population from a leading Australian online panel provider and incentivised accordingly to complete a 15 minute online survey.

In order to examine the likelihood of flat-lining, the means of each set of seven items for each brand equity construct were first derived for each respondent. The seven means were then used to compute the coefficient of variation (CV) (Equation 1) for each respondent as a standardized indicator of...
flat-lining responses. A high CV indicates variability in responses between items, whereas a near zero CV indicates flat-lining response behaviour. A distribution of the CV values is shown in Figure 2.

\[
CV = \frac{\sigma}{\mu}
\]

Equation 1: Coefficient of Variation

Figure 2: Distribution of CV

A Wards hierarchical cluster analysis was then conducted on the construct means to determine the optimal cluster solution. Table 4 and Figure 3 show that a two cluster or segment solution best separates the construct means.

Segment 1 exhibited a low CV mean of 0.1251 suggesting flat-lining respondents and Segment 2 exhibited a higher CV mean of 0.2343 suggesting non-flat-lining respondents (Figure 5).

Table 4: Duda Hart Indices

<table>
<thead>
<tr>
<th>Clusters</th>
<th>Je(2)/Je(1)</th>
<th>Pseudo T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5054</td>
<td>715.39</td>
</tr>
<tr>
<td>2</td>
<td>0.6294</td>
<td>266.14</td>
</tr>
<tr>
<td>3</td>
<td>0.6347</td>
<td>159.43</td>
</tr>
<tr>
<td>4</td>
<td>0.8174</td>
<td>80.21</td>
</tr>
<tr>
<td>5</td>
<td>0.805</td>
<td>51.84</td>
</tr>
</tbody>
</table>

Figure 3: Dendrogram for Clustering the Brand Equity Construct Means

As is seen from Figure 4, Segment 1 exhibits a uniformly higher mean for each of the seven brand equity constructs than Segment 2.

Figure 4: Brand Equity Construct Means by Segment 1 and Segment 2

Upon further examining the response patterns between the two segments, we noticed that Segment 1 comprised flat-lining responses (see Table 5 for an example compared to Table 6 for an example of non-flat-lining responses). A subsequent independent sample t-test confirmed that Segment 1 CV mean (12.51%) is significantly lower than Segment 2 (23.43%; \( t = 11.09 \); \( df = 731 \); \( p < 0.0001 \) (Figure 5).

This goes to show the impact of the flat-lining response bias on the segmentation, which is arguably the most popular market research and marketing tool in use today.
These respondents also tend towards lower incomes (<$75,000 p.a.) rather than higher incomes (>75,000 p.a.), with a high likelihood of internet connectivity. They consistently claimed a high level of brand or product knowledge, most likely because they were extremely active panellists (i.e., members belonging to 7+ online panels) most likely to chase survey incentives.

![Figure 5: Coefficient of Variation (CV) for Segment 1 and Segment 2](image)

Therefore, if we were to consider flat-lining online panellists to be “professional” respondents, it stands to reason that these respondents are undergoing a high frequency of online studies, which would explain their tendency to flat-line. Most likely for these respondents, they are rushing through online surveys and are likely to indicate high levels of brand or product knowledge so that they qualify for inclusion in an online study.

### Table 5: Example of Segment 1 Flat-Lining Brand Investment Construct Item Response Patterns

<table>
<thead>
<tr>
<th>Id</th>
<th>Bi1</th>
<th>Bi2</th>
<th>Bi3</th>
<th>Bi4</th>
<th>Bi5</th>
<th>Bi6</th>
<th>Bi7</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>13</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

### Table 6: Example of Segment 2 Non-Flat-Lining Brand Investment Construct Item Response Patterns

<table>
<thead>
<tr>
<th>Id</th>
<th>Bi1</th>
<th>Bi2</th>
<th>Bi3</th>
<th>Bi4</th>
<th>Bi5</th>
<th>Bi6</th>
<th>Bi7</th>
</tr>
</thead>
<tbody>
<tr>
<td>220</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>221</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>8</td>
<td>3</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>222</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>223</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>8</td>
</tr>
</tbody>
</table>

To shed more light on the two segments, we profiled each segment by conducting a Chi-Square test on profiling variables and reported those profiling variables that were significant (p < 0.05) in Table 7. As can be seen from Table 7, flat lining respondents in Segment 1 when compared to Segment 2 tend to be either young (<24 years) or old (65+ years) and female and are more likely not to be in full time employment.

### Table 7: Profile of Flat Lining Respondents v. Non-Flat Lining Acquiescent Respondents

<table>
<thead>
<tr>
<th>Respondent Profiles</th>
<th>Segment 1: Flat Lining Respondents n=279 (38.06%)</th>
<th>Segment 2: Non-Flat Lining Respondents n=454 (61.94%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tendency to flat line</td>
<td>High</td>
</tr>
<tr>
<td>2</td>
<td>Sex</td>
<td>Females</td>
</tr>
<tr>
<td>3</td>
<td>Age</td>
<td>Young and Old</td>
</tr>
<tr>
<td>4</td>
<td>Employment Full Time</td>
<td>Low</td>
</tr>
<tr>
<td>5</td>
<td>Income</td>
<td>Middle to Low</td>
</tr>
<tr>
<td>6</td>
<td>Internet Connection</td>
<td>High</td>
</tr>
<tr>
<td>7</td>
<td>Panel Membership</td>
<td>Membership of 5 to 8+</td>
</tr>
<tr>
<td>8</td>
<td>Claimed Bank Knowledge</td>
<td>Very High</td>
</tr>
</tbody>
</table>
Table 8 and Figure 4 above illustrate the over reporting of means of the brand equity constructs by some 23% when we do not remove flat-lining respondents, which would lead to significant biases in client reporting.

We next compare the Erdem and Swait (1998) brand equity framework via maximum likelihood structural equation models (SEM) for the complete sample as illustrated in Figure 6 and for the non-flat-line sample as illustrated in Figure 7. Table 9 shows the path coefficients in addition to the model fit statistics.

### Table 8: Comparison of Means for Complete Sample Relative to Non-Flat Lining Sample

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Complete Sample (A)</th>
<th>Segment 1 Flat Liners</th>
<th>Segment 2 Non-Flat Liners (B)</th>
<th>(A-B)/B*100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand Investments</td>
<td>6.3</td>
<td>7.5</td>
<td>5.6</td>
<td>13%</td>
</tr>
<tr>
<td>Consistency</td>
<td>6.0</td>
<td>7.5</td>
<td>5.1</td>
<td>18%</td>
</tr>
<tr>
<td>Credibility</td>
<td>6.3</td>
<td>8.2</td>
<td>5.2</td>
<td>23%</td>
</tr>
<tr>
<td>Perceived Quality</td>
<td>5.4</td>
<td>7.4</td>
<td>4.2</td>
<td>30%</td>
</tr>
<tr>
<td>Perceived Risk</td>
<td>5.9</td>
<td>7.8</td>
<td>4.7</td>
<td>25%</td>
</tr>
<tr>
<td>Information Costs</td>
<td>5.8</td>
<td>7.4</td>
<td>4.8</td>
<td>21%</td>
</tr>
<tr>
<td>Customer Service</td>
<td>5.4</td>
<td>7.4</td>
<td>4.2</td>
<td>29%</td>
</tr>
<tr>
<td>Sample Size</td>
<td>n=733</td>
<td>n=279</td>
<td>n=454</td>
<td>Mean = 23%</td>
</tr>
</tbody>
</table>

### Table 9: Maximum Likelihood Estimates for the Erdem and Swait (1998) Brand Equity SEM Framework to enable Comparisons of Total Sample to Non-Flat-Lining Sample

<table>
<thead>
<tr>
<th>ID</th>
<th>Structural Paths</th>
<th>Model 1 SEM with the Complete Sample</th>
<th>Model 2 SEM with the Flat-Lining Respondents Removed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gamma: Brand Investments → Credibility</td>
<td>Std'zed Estimate: 0.15 z-value: 4.98 p-value: 0.0000</td>
<td>Std'zed Estimate: 0.19 z-value: 4.18 p-value: 0.0000</td>
</tr>
<tr>
<td>2</td>
<td>Gamma: Consistency → Credibility</td>
<td>Std'zed Estimate: 0.20 z-value: 5.89 p-value: 0.0000</td>
<td>Std'zed Estimate: 0.22 z-value: 4.21 p-value: 0.0000</td>
</tr>
<tr>
<td>3</td>
<td>Gamma: Customer Service → Credibility</td>
<td>Std'zed Estimate: 0.67 z-value: 18.15 p-value: 0.0000</td>
<td>Std'zed Estimate: 0.60 z-value: 10.68 p-value: 0.0000</td>
</tr>
<tr>
<td>4</td>
<td>Beta: Credibility → Perceived Quality</td>
<td>Std'zed Estimate: 0.97 z-value: 30.15 p-value: 0.0000</td>
<td>Std'zed Estimate: 0.91 z-value: 14.93 p-value: 0.0000</td>
</tr>
<tr>
<td>5</td>
<td>Beta: Credibility → Perceived Risk</td>
<td>Std'zed Estimate: 0.74 z-value: 23.17 p-value: 0.0000</td>
<td>Std'zed Estimate: 0.78 z-value: 11.95 p-value: 0.0000</td>
</tr>
<tr>
<td>6</td>
<td>Beta: Credibility → Information Costs Saved</td>
<td>Std'zed Estimate: 0.98 z-value: 3.92 p-value: 0.0000</td>
<td>Std'zed Estimate: 0.51 z-value: 1.96 p-value: 0.0000</td>
</tr>
<tr>
<td>7</td>
<td>Beta: Perceived Risk → Information Costs Saved</td>
<td>Std'zed Estimate: -0.21 z-value: -0.85 p-value: 0.0000</td>
<td>Std'zed Estimate: -0.10 z-value: -0.37 p-value: 0.0000</td>
</tr>
<tr>
<td>8</td>
<td>Beta: Perceived Quality → Expected Utility</td>
<td>Std'zed Estimate: 0.30 z-value: 4.18 p-value: 0.0000</td>
<td>Std'zed Estimate: 0.20 z-value: 2.66 p-value: 0.0000</td>
</tr>
<tr>
<td>9</td>
<td>Beta: Perceived Risk → Expected Utility</td>
<td>Std'zed Estimate: 0.69 z-value: 7.15 p-value: 0.0000</td>
<td>Std'zed Estimate: 0.76 z-value: 7.54 p-value: 0.0000</td>
</tr>
<tr>
<td>10</td>
<td>Beta: Information Costs Saved → Expected Utility</td>
<td>Std'zed Estimate: -0.04 z-value: -0.92 p-value: 0.3576</td>
<td>Std'zed Estimate: -0.03 z-value: -0.66 p-value: 0.5619</td>
</tr>
</tbody>
</table>

Key Model Fit Statistics

<table>
<thead>
<tr>
<th>Model 1 SEM with the Complete Sample</th>
<th>Model 2 SEM with the Flat-Lining Respondents Removed</th>
</tr>
</thead>
<tbody>
<tr>
<td>df = 581</td>
<td>df = 581</td>
</tr>
<tr>
<td>Chi2 = 3448.12; P = 0.000</td>
<td>Chi2 = 2288.49; P = 0.000</td>
</tr>
<tr>
<td>CFI = 0.99</td>
<td>CFI = 0.96</td>
</tr>
<tr>
<td>NNFI(TLI) = 0.98</td>
<td>NNFI(TLI) = 0.96</td>
</tr>
<tr>
<td>RMSEA = 0.082</td>
<td>RMSEA = 0.081</td>
</tr>
<tr>
<td>SRMR = 0.29</td>
<td>SRMR = 0.31</td>
</tr>
<tr>
<td>PNFI = 0.91</td>
<td>PNFI = 0.88</td>
</tr>
</tbody>
</table>
Table 10: Squared Multiple Correlations (SMCs) of Endogenous Brand Equity Constructs

<table>
<thead>
<tr>
<th>Models</th>
<th>Credibility</th>
<th>Perceived Quality</th>
<th>Perceived Risk</th>
<th>Information Costs Saved</th>
<th>Expected Utility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete Sample</td>
<td>0.87</td>
<td>0.94</td>
<td>0.91</td>
<td>0.74</td>
<td>0.88</td>
</tr>
<tr>
<td>Flat-Lining</td>
<td>6.0</td>
<td>7.5</td>
<td>5.1</td>
<td>18%</td>
<td></td>
</tr>
<tr>
<td>Reduced Sample</td>
<td>0.68</td>
<td>0.85</td>
<td>0.79</td>
<td>0.42</td>
<td>0.82</td>
</tr>
<tr>
<td>Difference</td>
<td>0.19</td>
<td>0.09</td>
<td>0.12</td>
<td>0.32</td>
<td>0.06</td>
</tr>
<tr>
<td>Difference %</td>
<td>27.94%</td>
<td>10.59%</td>
<td>15.19%</td>
<td>76.19%</td>
<td>7.32%</td>
</tr>
</tbody>
</table>

It can be seen that the SMC for the information costs saved construct is alarmingly high for the flat-lining removed sample when compared to the complete sample. These construct items require higher cognitive effort to process, and in the case of flat-lining respondents who were seeking to speed through the rating scale items by expending minimum cognitive effort, they artificially inflated the SMC.

To a lesser extent, though still of concern, is the credibility construct which also requires higher
cognitive effort to be expended in the underlying rating scale items. In the case of the expected utility construct where the items are the easiest to cognitively process, the difference between the complete sample and the flat-lining removed sample is the smallest. Figure 8 provides a ranking of the cognitive difficulty of the construct. The end result of not controlling for the flat-lining response style bias is that the client would have been provided a rosier picture than reality, thereby misleading the client.

Figure 8: Descending Order of Construct Cognitive Difficulty

To mitigate the flat-lining response style bias in online research, we recommend that market research community increase the quality of online panels and enforce industry standards such as the AMSRO guidelines for quality recruitment and online panel management. To the best of our knowledge, there is little research on the topic in the published literature. We therefore urge the market research industry to heed our call for more attention in this important research area. If online panel recruitment and management standards are not properly adhered to by online panel providers, the credibility of the market research profession will suffer in the long run.

We believe that we have demonstrated the need for market researchers to think about how online sample data needs to be analysed to provide valid information to clients. Online sample data needs to represent the target population that is under study and account for response biases that distort research outputs.

We are not suggesting that all flat-lining respondents are not representative of the population, as there will be cases where respondents are invariant across any number of rating scales legitimately. However, in cases where the rating scales differ in their direction and flat-lining responses prevail, then it stands to reason that these particular respondents are not representative. Also, in cases where the rating scales do not differ directionally and respondents flat-line at the extreme ends of the scale, it stands also to reason that these respondents are clearly not representative.

The motives for flat-lining are many (Ben-Nun 2008). Earlier we alluded to flat-liners as respondents who are entering an online survey for the purpose of using survey incentives as a supplementary income stream, given their likely adverse financial situation (see Table 7). If this is the case, that is online surveys are viewed by particular respondents as an alternative income stream, then these respondents will be motivated to do as many surveys as possible. Accordingly, their likelihood of speeding through an online survey will be higher when compared to respondents that are not using online surveys as an alternative income stream. Another motive for flat-lining could well be one of avoiding arduous long surveys, where respondent fatigue is commonly encountered (Herzog and Bachman 1981).
Yet another possible reason for flat-lining is due to the cognitive load experienced by respondents who are required to complete a survey with random rotation designed to overcome order effects. In the case of rotations of rating scales across rather than within constructs, it means that respondents will find it difficult to rate these scales due to arbitrary construct representation, thus increasing cognitive burden. In this instance, the incidence of flat-lining will increase as respondents avoid cognitive load (Bradburn and Mason 1964).

Finally, incentive compatibility will also affect the likelihood of flat-lining because when incentives are large, the likelihood of consumers behaving rationally is high in both familiar and unfamiliar situations. However, when incentives are small, rationality may not hold (McFadden 2006). Typically in survey research, incentives are small in the form of either a low chance for a large prize or a high chance of a small prize. Therefore, the longer the survey and the smaller the incentive, thus becoming incentive incompatible, the higher the likelihood of flat-lining because of respondents trading off their effort-accuracy threshold versus their incentive (Payne et al. 1993).

The motives for flat-lining that we have discussed are by no means exhaustive. They however demonstrate that flat-lining, except in instances where respondents truly are invariant across rating scales that are of the same direction, are not representative of the population and hence biased in their responses.

In turn, we firstly recommend that quality procedures be built into the panel recruitment and management stages, and as a minimum that ISO 26362 or preferably QSOAP Gold panels be considered for use. Secondly, we recommend that both researchers and clients are educated as to the range of possible biases in online data and how to account for such biases in analysis and modelling. No matter how careful an online panel provider screens respondents, there exist respondents who will try to play the system. A quality online panel provider ensures that respondents take the time to undertake a survey and provide answers that make sense. Therefore, thirdly we recommend that respondents with problematic response patterns such as flat-liners or speeders i.e., those who race through an online survey, should be identified and dealt with appropriately.

One way of dissuading flat-liners or speeders, is by warning respondents that those who exhibit such behaviours will fail to be compensated and will be removed from the online panel. Another way of identifying flat-liners or speeders is by including a battery of psychological test items such as social desirability scales (Crowne and Marlowe 1960). Lastly, we recommend that researchers and analysts visually inspect their rating scale data and use measures such as CV, so that they may examine low and high CV values to identify respondents that are either completely or largely invariant. Once flat-lining respondents have been identified, the decision can be made as to whether these respondents should be removed from the dataset prior to analysis.

We hope that we have made apparent that flat-lining is a bias that limits the reliability and representativeness of online data. We also hope that we have demonstrated that flat-lining can be controlled by an end-to-end online data collection process that promotes best practice in online sample recruitment.

Two promising avenues for future research in the area of flat-lining include examining the usefulness of psychological test items such as social desirability scales (Crowne and Marlowe 1960), and also examining the impact of flat-lining on discrete choice modelling incorporating scales that detect likely truthfulness in survey response and whether truthfulness interacts with flat-lining.
6 REFERENCES


BOOK REVIEW

Grounded Theory: A Practical Guide

Melanie Birks and Jane Mills
Sage LA 2011.

Melanie Birks and Jane Mills are to be congratulated on compiling a book on grounded theory that reviews the field and provides excellent advice to the researcher. *Grounded Theory: A Practical Guide* provides the reader with an analysis and evaluation of approaches to grounded theory, over the decades, without getting stuck in the politics of one researcher’s approach over another’s. Qualitative research techniques, of which grounded theory is one, are often used as a means of engaging in research that seeks to drive social change. Birks and Mills provide a framework for researchers who wish to undertake research that follows a principled approach to making choices about important issues for researchers from diverse backgrounds and with varying levels of experience and or expertise.

As is often the case in qualitative research publications, the authors are academics in the health domain, both in Australian universities. They argue that there is a need for their book on grounded theory as they traverse the magnitude of publications in the field without complicating it further. In contrast, they clarify and simplify the domain, uncovering the general principles of grounded theory in the context of debate which is part of the tradition of qualitative research. The field is dominated by rivalries between different research paradigms, philosophical positions and methodological approaches. Birks and Mills manage to tread lightly around these sensitivities while providing the reader with guidance on how to understand grounded theory. The approach allows the reader to identify their own baseline position by reading about the perspectives on offer by other researchers on grounded theory over the decades.

The book is written for people seeking to undertake a study using grounded theory but who are finding the labyrinth of theories and approaches daunting and confusing. However, it is also valuable for those who are teaching qualitative research techniques to research students.

*Grounded Theory: A Practical Guide* has a number of strengths, including the structure and the writing style. The book is split into ten chapters, with a glossary, references, index and appendices. Multiple levels of detail are provided by experts who have written the boxed case studies and examples from various disciplines, providing a examples and issues of a practical nature. This provides the book with a richness and completeness that further benefits the reader.

The book is well written, easy to understand and well set out. Each chapter begins with learning objectives and concludes with a summary, critical questions and leads to appendices that suggest that the reader pursue additional examples of grounded theory.
Each chapter also provides boxed activities, short case studies and questions at the end of each chapter, for personal reflection and additional learning. The authors have illustrated their key concepts with useful figures and tables. Additional contributors are profiled in appendix B. The additional contributors work in a variety of departments in universities, from hospitality and retailing to social work and education, in various locations around the world. The spread of contributors and disciplines is no doubt one of the aspects of the book that makes it relevant to readers of different disciplines. They contribute the boxed case studies and examples.

The references are excellent, providing a rich list of options for further reading. I particularly found useful the references on diversity, ethnicity and Indigenous qualitative research approaches. References are current, topical but nonetheless range widely over the field. In summary, Birks and Mills provide a balanced approach to understanding grounded theory that is useful to the beginning researcher as well as the experienced researcher.

While this brief overview might suggest that the book is a text book, it is not. It is firmly based on the authors own research in grounded theory. This gives the book authenticity and authority. It contributes to their depth of knowledge of grounded theory and to their ability to explain the various roads that it has taken over the years.

The collective expertise and insights from the authors and the other contributors give the book its air of completeness and depth. The authors are authorities in their field, which is evident in their bios but also in their approach to the discussions on grounded theory. The collective expertise ensures that the book demonstrates an analytic grasp of the topic that is diverse, detailed and demonstrably authoritative without being didactic. Issues are brought to life by clear writing and good editing, accompanied by black and white illustrations.

Overall, the book is a valuable addition to a researcher’s book shelf. At 210 pages, it is a suitable length without being overwhelming for the reader. The international relevance of the book is reinforced by the examples from more than one discipline, figures, models and illustrations conveying aspects of grounded theory. The book concludes with a four page glossary and the ten page list of references, providing the reader with useful material up to the end.

The last chapter covers current debate in grounded theory, how the methods have evolved over 50 years, and potential frontiers for future theory development. Qualitative research itself is on far stronger foundations than was the case 50 years ago, with many articles and books covering what it is and how to engage in qualitative research. Grounded theory is one of those methods. Birks and Mills envisage readers at various stages in their research careers reading the book, and provide them with opportunities to position themselves in the agenda for social change, that is the underpinning of much grounded theory research.

Professor Ruth Rentschler
School of Management & Marketing
Deakin University