

Australasian Journal of Market & Social Research

June 2007 • Volume 15 Number 1 • ISSN 1832 7362



CONTENTS

Improve survey responses by pre-segmenting customer samples Philip Derham, Derham Marketing Research Pty. Ltd.	Page 5
Coffee, tea or replacement questionnaire? improving mail survey response rates from volunteers Mike Brennan, Jan Charbonneay and Andrew Hercus	Page 11
User-friendliness of answer formats – an empirical comparison Sara Dolnicar and Bettina Grün	Page 19
Pressure to innovate: Does validity suffer? Linda Brennan and Jim Camm	Page 29
Book Review: Handbook of Ethnography Reviewed by Michael Milgate	Page 43

AUSTRALASIAN JOURNAL OF MARKET & SOCIAL RESEARCH

Editor

Professor Lester W. Johnson
Melbourne Business School

Editorial Advisory Board

Susan Ellis
Melbourne Business School

Liane Ringham
Inside Story

Mark Jessop
Disability Services Commission

John Rossiter
University of Wollongong

V Kumar
University of Connecticut

Geoffrey Soutar
University of Western Australia

Peter Oppenheim
Deakin University

Don Stern
Washington State University

Pascale Quester
University of Adelaide

Jill Sweeney
University of Western Australia

AJMSR is published by the Australian Market & Social Research Society
Copyright 2007. ISSN 1832 7362

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
Web: 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 full members of the AMSRS receive a complimentary subscription to AJMSR as part of their membership.

It is also possible to subscribe to AJMSR:

Australian Orders

\$48.40 (inc GST) per single issue

\$82.50 (inc GST) per one year subscription

International Orders

\$53.00 (AUD) per single issue

\$94.50 (AUD) per one year subscription

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.

Contributors of papers will receive five copies of the issue in which their article appears.

Manuscripts should be typed and double spaced. Diagrams and figures should be provided electronically, and of a professional quality. In writing papers, authors should consider the style of articles in previous issues.

On acceptance of an article, authors will be asked to supply their article electronically, preferably in Microsoft Word format. However, conversions can be made from other packages.

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:

Professor Lester W. Johnson
Editor, AJMSR
Melbourne Business School
The University of Melbourne
200 Leicester Street
Carlton VIC 3053 Australia
Email: L.Johnson@mbs.edu

Subscriptions and advertising

All enquiries regarding subscriptions, advertising and other related 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

Advertising rates

Logo placement – outside front cover
(\$577.50 inc GST)

One page colour ad – inside front cover
(\$539.00 inc. GST)

One page colour ad – outside back cover page
(\$803.00 inc. GST)

One page black and white ad – inside
(\$269.00 inc. GST)

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 Professor Lester Johnson (Editor) and the Melbourne Business School.

AJMSR – Editorial

This issue of the journal contains several papers from practitioners and academics, something not seen very often in the pages of the *Australasian Journal of Market & Social Research*. The paper by Dolnicar and Grün empirically investigate the interesting question of user-friendliness of various question formats.

A second paper is by Derham (a practitioner) who examines how pre-survey segmentation can be used to improve overall response rates to self-completion surveys, and hence slightly reduce concern about non-response bias. A third paper co-authored by Brennan (an academic) and Camm (a practitioner) examines the trade-off between innovation in market research practice and subsequent validity of the research outcomes. This is a wide-ranging paper with lots of things for researchers to think about.

The fourth paper in this issue by Brennan, Charbonneay and Hercus (all academics) examines ways to improve mail survey responses from volunteers, and concludes that sending a replacement questionnaire substantially increases response rates compared to other methods considered.

In addition there is a book review by Milgate offering an extensive critique of the diverse and fascinating terrain as covered in the *Handbook of Ethnography*.

I trust that the issue will contain things of interest for all readers. Please feel free to submit a paper to the journal for consideration.

Lester W Johnson, Editor
Australasian Journal of Market & Social Research

Interested in becoming an AMSRS member?

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 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), vote in divisional and national elections of the AMSRS and receive the Society's Journal AJMSR.

Note: Student members receive limited membership benefits so please see 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.

JoinToday

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

Improve survey responses by pre-segmenting customer samples

**Philip Derham,
Derham Marketing Research Pty. Ltd.**

Summary

The overseas evidence suggests more, expensive and lengthy, contact is necessary to increase self completion survey response rates and to lessen the non-response bias. Commercial clients tend to use self completion surveys because of limited budgets and often require results in a reasonably short time, lessening opportunity to use the frequent contacts known to drive up response rates. The commercial options that follow are to accept a 10% to 15% self completion rate as reasonable and live with a concern about the non-response bias or take creative steps to increase response rates within the limits and work to lessen any non-response bias. This paper reports how pre-survey segmentation of prospective respondents can improve overall response rates and indicates non-response bias is less a concern because segment-weighted samples deliver slightly more accurate results. The conclusion is that one time contact self completion surveys can be undertaken cost-effectively and still be reliable guides to management decisions.

Self Completion Surveys

Self completion surveys are a staple data collection method used extensively in commercial market research, in non-commercial sector research and by academics in a range of disciplines (Hagar, Wilson, Pollak and Rooney 2003). As self completion surveys rely on prospective respondents choosing to complete and return surveys, there has been extensive research to identify techniques that improve response rates (Dillman 2000) and to identify reasons for non-response (Gendall, Hoek and Finn 2005). The response improvement techniques investigated initially concentrated on postal mail self completion surveys and have, more recently reviewed web self completion surveys (Sheehan 2001). These mostly American studies found that similar techniques in postal and web self completion surveys stimulate higher response rates. The techniques include more frequent contact with the prospective respondents, ensuring the survey is relevant to the prospective respondents and by offering appropriate incentives for participation (Kaplowitz, Hadlock and Levine 2004).

Single contact self completion surveys elicit response rates of 10% to 15% among general samples (Dillon, Madden and Firtle 1987; Malhotra 1999) and high-

er rates in specific sectors – 22% among logistics industry executives (Larson 2005) or 32% among students (Kaplowitz et al, 2004) and often more in other respondent groups. Higher response rates were seen to legitimise the results while lessening a possible non-respondent bias (Hager et al, 2003) that was discussed extensively in the American academic literature. The assumption that survey answers were reported as they were received – essentially raw and unweighted was un-stated but implicit in the articles reviewed.

In contrast, in commercial Australian market research, where self completion surveys are often used because of limited budgets, the extensive pre and post-survey contact recommended in the American literature is less often used. Similarly, the possible non-response bias in Australian self completion surveys is often managed by weighting results to already known sample universe characteristics (such as age, gender or area). The practical assumption that commercial Australian researchers make is that the other survey answers will also match the views of all people in the universe (within an expected sample variation). This weighting approach is more than Australian pragmatism. Australian and New Zealand studies suggest there are few who refuse

every survey opportunity and prospective respondents will complete relevant self completion surveys when they can (Gendall 2005). Hence, while weighting is seen as an appropriate answer to the non-response bias concern prevalent in American academic literature, the task of maximising response rates within limited budgets is one that Australian researchers need to address, to ensure the data adjustment made by weighting is as little as possible. In other words, Australian market researchers need to make their one time contact work harder. This article addresses that need.

Improved Sampling Without Increased Budget

The prompt to review response rates was partly driven by results from three, one time contact self completions web surveys among the same customer audience during one year. Those one time contacts generated response rates of 19%, 20% and 17%, but had quite different levels of respondent salience (Derham 2006). The same contact approach was used for each - a personally addressed email from the client's Chief Executive, which offered entry in a competition to win a prize if the survey were completed by a cut-off date. The first survey was a general survey of customer activity and in that 18% of the total sample reported involvement with financial planners. The second survey was specifically about financial planning and 47% of the respondents reported involvement with financial planning. The third survey was directed to those who had bought furniture in the last year and 74% of its respondents reported doing so, while only 15% in the first, general survey had reported furniture purchase. These findings suggested self-selected segmentation, motivated by the email subject line, may not change the overall response rate but could change the salience of the survey to customers and so change the composition of those who chose to respond.

Reliance on self-selection by salience segment is limiting, as the small proportions

of respondents with financial planning or furniture purchase salience in the first, general web survey indicate, and did not lead to higher response rates, merely more relevant samples. Hence, the need was to identify whether pre-existing segments could be used in sample design to improve the overall response rate, and if such segments would affect the non-response bias normally managed by weighting by age, gender and perhaps location.

The findings from earlier privately commissioned qualitative research among financial institution customers suggested customer engagement with financial institutions was variable and correlated more with the customer's value (profitability) to the financial institution than with age, gender or location. The financial institutions involved were able to pre-determine customers' financial value and so could draw samples by financial value, if required. If the level of engagement with the financial institution identified in the qualitative research also reflected propensity to respond to the financial institutions' customer self completion surveys (which was not qualitatively tested), the sampling frame could use pre-existing customer value segments to increase overall response rates, and better response rates could also enable more precise weighting - by value segment as well as age and gender - so further lessening any non-response bias.

The next step was to identify pre-existing segments of financial value to use to test the hypothesis that customer value affects response rates. The customer value segments were determined by the overall net positive or negative value of all customers' transaction costs, investments or loans value and use of fee-generating ancillary products. Using this formula, customers were divided into ten equal deciles. Customers in deciles 1 and 2 were the most profitable and least transactional customers. Customers in deciles 9 and 10 were the heaviest transactors and

Sample and response by decile combinations.

Decile Combinations	% Of Original Sample	% Of Survey Responses
1 to 3	36%	42%
4 to 6	32%	33%
7 to 10	31%	24%

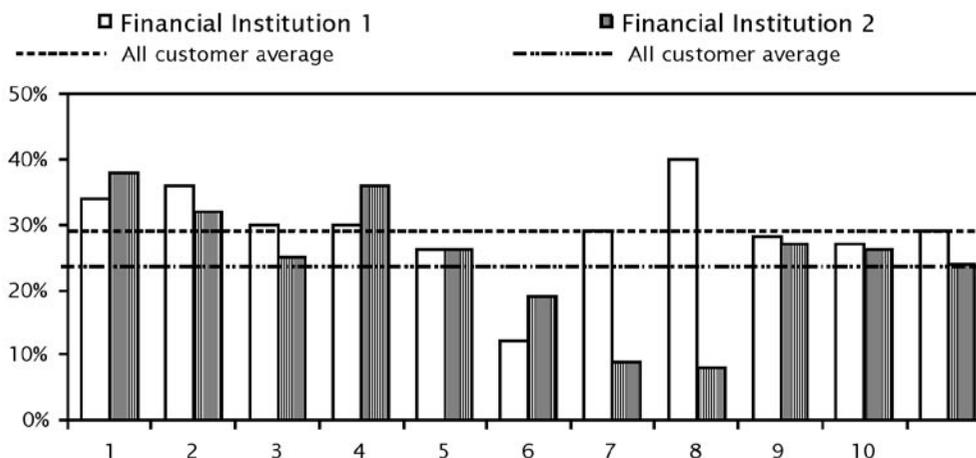
most costly/least profitable customers. Customers in deciles 6, 7 and 8 made little use of the financial institution and so cost little but also contributed little. Customers in the deciles 3, 4 and 5 were less valuable in terms of loans or investments but used the costly financial services less. The hypotheses were simple. If responsiveness to self completion surveys was independent of customer financial value to the financial institution, all segments would have similar response rates and if financial value was a determinant of response, response rates would vary by financial value decile.

Findings from a web self completion survey sent to customers in preset decile groups, with customers in different deciles receiving different survey links showed the financial value decile segments showed different response rates as the table above shows. This finding indicated the

financial value deciles may be an effective pre-segmentation tool.

A more extensive review of customers' financial value decile segments followed and results from two recent financial institution customer self completion surveys were analysed. The two separate financial institutions had both undertaken mail-out, mail-back self completion surveys. For economy, each had used samples of 2,000 randomly selected customers aged 18 years or older. Each randomly selected customer was mailed a survey pack containing a letter of invitation to participate, an attractive typeset A4 page size, double sided survey folded to DL size, and a postage-paid reply envelope. Both surveys offered prospective respondents an opportunity to be entered into a competition to win a prize if their response was received by the close date. Financial Institution 1 had a 28%

Responses by value decile



response rate and Financial Institution 2 a 22% response rate. The graph below shows that response rates varied markedly by value decile, supporting the hypothesis that customers' financial value to the financial institution was correlated with engagement and willingness to participate in self completion surveys.

The most profitable customers (those in deciles 1, 2, 3, and 4) responded more strongly than customers in the unprofitable deciles 6 and 7 and the differences were statistically significant at the 95% confidence level. The differences exceeded the 4% sample variation of Financial Institution 1 and greater than 5% for Financial Institution 2. Customers least active - those in deciles 6 and 7 - were also least likely to participate in self-completion research. Active transaction customers – those in deciles 9 and 10 - were no more likely than customers generally to participate in customer research. The responses from decile 8 customer responses differed by financial institution, and subsequent review of Financial Institution 1's decile 8 customers found most to be new customers, who, when the research was conducted, had had insufficient time to establish a value history.

These findings suggest response rates are linked to customers' financial value, that

regular transactional interaction indicates only average involvement and likelihood for self-completion survey response, and that low value and low interaction predict lower response rates than average.

Improved Responses Rates

In each project, the financial institution had a budget sufficient to allow for the printing and dispatch of a total out-bound sample of 2,000 customers, and to receive, keypunch and analyse about 500 responses. In both projects, 200 customers from each decile were randomly selected and invited to participate. Given these findings, it would seem that overall response rates can be improved if the sample is not evenly drawn by decile but is unevenly drawn so that fewer than the 10% of the sample is from value deciles whose customers are most likely to respond (deciles 1 to 4) and more than 10% of the customers sampled would be from deciles 6, 7 and 8, who are less likely to respond. The production costs would remain within the budget and the results strengthened.

Variable quota sampling by value decile should result in a more even number of responses in each decile, which in turn should lessen the possible non-response bias by more accurate weighting. The weighting would be by age and gender and by value decile which would ensure

Comparisons between weighted overall and weighted by decile

Satisfaction level	FI 1 – weighted by age and gender	FI 1 – weighted by age and gender and decile	Difference with decile weighting (var. 3% to 4%)	FI 2 – weighted by gender	FI 2 – weighted by gender and decile	Difference with decile weighting (var. 3% to 5%)
Very satisfied	58%	56%	2%	44%	40%	4%
Satisfied	40%	41%	1%	51%	54%	3%
Dissatisfied	2%	2%	0%	2%	3%	1%
Very dissatisfied	0%	0%	0%	0%	0%	0%
Not answered	0%	0%	0%	2%	3%	1%

Note: 0% may be up to 0.4% and the tables do thus add to 100%.

the results were not over-influenced by the numerically over-represented views of the more engaged customers (those in deciles 1 and 2) or under-represented the disengaged views of customers (those in deciles 6 and 7). This assumption was tested by analysis of results from a question common to both surveys – of overall satisfaction with the financial institution. The differences between the *customers weighted by all* results and the *customers weighted by all plus by decile* were small – at most 4% - and less than the variation that could be attributed to sampling. While the slightly more accurate results were from the additional weighting by decile, the minimal difference suggests the initial results were accurate and the non-response bias has been fairly addressed by weighting.

Conclusions

The examples presented were from the financial industry but indicate the potential value from customer database segmentation before survey research is undertaken, for sample design and for post-survey weighting to ensure more accurate reflection of the views of the sample and the universe sampled. Pre-segmentation will ensure the same budget produces better response rates and more accurately weighted data and be cost-effective marketing research. Variable quota sampling by segment cannot, of course, be undertaken the first time a segmented customer database is surveyed, though it can be used in subsequent follow-up surveys. Weighting by segment can be undertaken from the first segment-quota survey to ensure responses received better reflect the true population, and so is a recommended enhancement to survey practice.

REFERENCES

- Bednall, David H. and Shaw, Mike. (2003) 'Changing response rates in Australian Market Research'. *Australasian Journal of Market Research*, April 2003, Volume 11, No. 1, pp 31-41.
- Derham, Philip. (2006) 'Self completion surveys – relevance is necessary to strengthen reporting capacity!' *Quirks Marketing Research Review*, October 2006.
- Dillon, William R. Madden, Thomas J. Firtle, Neil H. (1987) *Marketing Research in a Marketing Environment*. St. Louis: Times Mirror/Mosby College Publishing.
- Dillman, Don. (2000) *Mail and Internet Surveys: The Tailored Design Method*. New York: John Wiley & Sons.
- Gendall, Philip. Hoek, Janet. Finn, Anna. (2005) 'The behaviour of mail survey non-respondents'. *Australasian Journal of Market & Social Research*, December 2005, Volume 13, Number 2, pp 39-50.
- Hager, Mark A. Wilson, Sarah. Pollak, Thomas H. Rooney, Patrick Michael. (2003) 'Response rates for mail surveys of non-profit organisations: A Review and empirical test'. *Non-profit and Voluntary Sector Quarterly*, Vol. 32, No. 2, June 2003, pp 252-267.
- Kaplowitz, Michael D. Hadlock, Timothy D. Levine, Ralph. (2004) 'A Comparison of Web and Mail Survey Response Rates'. *Public Opinion Quarterly*, Vol. 68 No. 1, pp 94-101.
- Larson, Paul D. (2005) 'A note on mail surveys and response rates in logistics research'. *Journal of Business Logistics*, 1/1/2005. http://findarticles.com/p/articles/mi_qa3705/is_200501/ai_n15716094
- Malhotra, Naresh K. (1999) *Marketing Research. An Applied Orientation*. New Jersey: Prentice Hall.
- Reichmann, W J. (1964) *Uses and Abuses of Statistics*. Harmondsworth, UK: Pelican.
- Sheehan, Kim. (2001) E-mail Survey Response Rates: A Review. *Journal of Computer-mediated Communication*, Volume Six, Issue Two, January 2001, <http://jcmc.indiana.edu/vol6/issue2/sheehan.html>



Coffee, tea or replacement questionnaire? Improving mail survey response rates from volunteers

**Mike Brennan, Jan Charbonneay - Massey University
Andrew Hercus - Christchurch College of Education**

Abstract

Research has demonstrated that multiple mail-outs and incentives (preferably monetary) are effective at increasing response rates in mail surveys. However, when surveying volunteers it may be neither appropriate nor desirable to use more than one reminder or monetary incentives. This study examines the effectiveness of using a teabag, coffee sachet or both as a pre-paid incentive in two surveys involving 2860 volunteers from two sporting codes and two health charities. While teabags and coffee sachets, either alone or in combination, were found to be ineffective, sending a replacement questionnaire rather than a reminder letter alone substantially increased the response rates for all volunteer groups.

Introduction

Mail surveys are an important market research tool and are particularly suitable for covering large geographic areas or when research requires respondents to view graphic material or read detailed descriptions. They are also useful when the survey is long, or complicated, or when time is not a crucial consideration. However, as with all survey methods, respectable response rates are needed to minimise the possibility of non-response bias.

Techniques for improving mail survey response rates have been well researched. Virtually every aspect of the mail survey procedure has been examined, such as the envelope (e.g. size, colour, franked versus stamped, type of stamp, type of address, handwritten versus typed address, logo), the cover letter (e.g. tone, signature, paper type, honorific, content), questionnaire (e.g. colour, size, topic, layout), contact procedures (e.g. number of contacts, letter versus postcard versus telephone call, sending replacement questionnaire or not), and the use of incentives – either promised or pre-paid (see Dillman 1978; Kanuk & Berenson 1975; Linsky 1975; Duncan 1979; Yu & Cooper 1983; Harvey 1987; Fox, Crask & Kim 1988; Dillman 1991; Brennan 1992a; Church 1993; Dillman 2000).

While numerous survey elements have been found to make small differences to response rates, many have proven ineffective. However, the use of multiple attempts to contact non-respondents and the use of a pre-paid monetary incentive have consistently been found effective in generating higher mail survey response rates (Kanuk & Berenson 1975; Linsky 1975; Yu & Cooper 1983; Harvey 1987; Dillman 2000).

For surveys of volunteers, however, neither of these procedures is particularly desirable. For volunteer organisations one simply would not want to risk damaging the goodwill of the respondents or of their organisation by persisting with attempts to elicit respondent participation. Furthermore, the use of a cash incentive also seems inappropriate when surveying volunteers. Using a cash incentive raises the legitimate concern of the risk of insulting or alienating volunteers by this apparent misuse of valuable funds. The challenge, then, is to find an alternative incentive that is low cost, easy to include with a cover letter, and seems appropriate for the target sample. For surveys of volunteers, what is needed are techniques that will generate timely responses to the survey, eliminating the need for more than one follow-up contact, while also

generating a respectable response rate in a cost-effective manner. One possible way of doing this is to send a replacement questionnaire rather than just a letter with the follow-up mail-out, and to use a non-monetary pre-paid incentive.

In a survey where two follow-up mail-outs are used, the usual procedure is to send out a questionnaire plus cover letter in the first mail-out, a reminder letter only with the first follow-up mail-out, and a replacement questionnaire with second follow-up, in case the original has been discarded or mislaid. While several studies have examined whether it is better to send a replacement questionnaire with each mail-out when three mail-outs are used, the findings are inconsistent (see Futrell & Lamb 1974; Herberlein & Baumgartner 1978; Brennan 1992a, 2004, 2005; Brennan & Charbonneau 2005). There also appears to be no specific data for surveys of volunteers. Thus for surveys where only two mail-outs are possible or desirable due to the unique characteristics of the survey population, such as for volunteers, it would seem sensible to determine whether it is sufficient to send a reminder letter only with the first follow-up, or whether a replacement questionnaire should be sent as well.

The challenge is to find a suitable incentive for volunteers. A wide range of pre-paid incentives have been examined in various studies reported in the literature, including stamps, book marks, coins, scratch cards, lottery tickets, phone cards, tea bags, coffee sachets, chocolates, pens and even books (see Kanuk & Berenson 1975; Linsky 1975; Dillman 1978, 1991, 2000; Yu & Cooper 1983; Harvey 1987; Brennan 1992a,b; Church 1993; Gendall, Hoek & Brennan 1998; Ryu, Couper & Marans 2005; Brennan & Charbonneau 2005). Unfortunately, some of these studies were conducted many years ago, and in very different economic and social climates, and often involved unique survey populations. It is not surprising then, that the findings for

some incentives tested in different studies are conflicting. As a consequence, these previous studies need to be taken as sources of ideas rather than as indicators of what works in any specific situation today.

Of the incentives listed above, some seem unsuitable for volunteers due to the suggestion of payment (e.g. phone cards, cash, stamps) or gambling (lottery, scratch & win), while others are more difficult to send via the post (e.g. pens, books, chocolates). On the face of it, using tea bags or coffee sachets as “tokens of appreciation” would seem to be appropriate as incentives in a survey of volunteers, as they have connotations of “caring and sharing” consistent with the volunteering experience. While tea bags have been found previously to be ineffective in surveys of the general public (Gendall, Hoek & Brennan 1988), this does not necessarily mean that they would not work for a survey of volunteers. A possible reason for the poor results in previous studies is that they used only tea bags, yet a substantial proportion of the population drink coffee. It would therefore seem sensible to test whether it is more effective to use either tea or coffee as the incentive, or whether one should send both so respondents have a choice. Because they are inexpensive and easily sent by mail, and seem to fit the profile of a suitable incentive for this group, further investigation of tea bags and coffee sachets seems warranted.

Thus the purpose of the studies reported here was two-fold:

- i. To examine the effectiveness of using a teabag or coffee sachet (or both) as a pre-paid incentive; and
- ii. To examine the effectiveness of using a replacement questionnaire with the first reminder.

Method

Two mail surveys were conducted between October and December, 2005, using identical 6 page questionnaires and

Table 1 Research design for Survey 1

Wave	Treatment					
	1 %	2 %	3 %	4 %	5 %	6 %
1	Letter + Q'naire	Letter + Q'naire + Teabag	Letter + Q'naire + Coffee	Letter + Q'naire + Teabag + Coffee	Letter + Q'naire	Letter + Q'naire
2	Letter	Letter	Letter	Letter	Letter + Q'naire	Letter + Teabag + Coffee

cover letters. The survey asked respondents for standard demographic information as well as questions about their volunteer service, what motivated them to volunteer with the specific organisation (30 items), volunteering in general (30 items), and their satisfaction with the organisation (18 items). The cover letter was on University letterhead and explained that the survey was being conducted on behalf of the organisation and that respondent names were provided by the organisation. University ethics approval was obtained and respondents given an assurance of confidentiality. A copy of the letter is provided in Appendix A.

In Survey 1, the sample was 1872 volunteers (coaches, referees and administrators) from a large sporting organisation. The final sample size, adjusted for GNA (Gone-No Address) and ineligible

was 1779. Six treatment groups, each with an initial sample size of 312, were used to test the effects of a combination of treatments in the first and second mail-outs (see Table 1). The first mail-out included a teabag, coffee sachet, both or neither, attached to the cover letter (plus a questionnaire and reply-paid envelope). The second mail-out was sent 14 days after the first. This comprised a letter only (Treatments 1-4), a letter plus replacement questionnaire but no incentive (Treatment 5), or a letter with both a tea and coffee sachet attached but no questionnaire (Treatment 6).

In Survey 2, the same questionnaire and cover letter used in Survey 1 was sent to two further groups: volunteers from two health charities (N = 658, adj. N = 647) and a different sports organisation (N =

Table 2 Research design for Survey 2

Wave	Treatment		
	1 %	2 %	3 %
1	Letter + Q'naire	Letter + Q'naire	Letter + Q'naire
2	Letter	Letter + Q'naire	Letter + Teabag

331, adj. N = 324). The standard procedure was used in the first mail-out (cover letter, questionnaire, reply-paid envelope, no incentives), with three treatments tested in the second mail-out 14 days later: Treatment 1 - control (cover letter only); Treatment 2 (cover letter and replacement questionnaire); Treatment 3 (cover letter and teabag, no questionnaire).

In the following discussion, responses to the first mail-out will be referred to as Wave 1, and responses to the second mail-out (first reminder) as Wave 2.

Results and Discussion

Survey 1

The results for Survey 1 are reported in Table 3. The response rates for the first mail-out (Wave 1) ranged between 20% and 31%, with the best result actually from one of the control groups (no incentive). The lowest response was from Treatment 1 (20%), which was the overall control group. However, this result appears to be an anomaly, since two other groups (Treatments 5 and 6) were also controls for the first mail-out, and these both achieved substantially higher response rates than Treatment 1 (31% and 27% respectively).

While both the teabag (Treatment 2) and the coffee sachet (Treatment 3) achieved a slightly higher response rate than two controls (Treatments 5 and 6) in Wave 1, sending both tea and coffee (Treatment 4) produced a lower response rate than sending no incentive, which seems somewhat strange. However, since the differences between these control and treatment groups were not statistically significant, these difference are likely due to chance, and one can conclude that both the teabags and coffee sachets were ineffective.

The response rates to the follow-up mail-out (Wave 2), ranged from 22% to 38%, with the best result produced by the treatment using a replacement questionnaire (Treatment 5). Treatment 5 produced a significantly higher response in Wave 2 than all of the other treatments, demonstrating clearly the effect of sending a replacement questionnaire rather than simply sending a reminder letter (T5:T1, $\chi^2 = 8.63$, $df = 1$, $p < .01$; T5:T2, $\chi^2 = 4.54$, $df = 1$, $p < .05$; T5:T3, $\chi^2 = 12.21$, $df = 1$, $p < .001$; T5:T4, $\chi^2 = 7.55$, $df = 1$, $p < .01$; T5:T6, $\chi^2 = 9.81$, $df = 1$, $p < .001$).

Overall, the final response rates after two mail-outs ranged from 39% - 56%. The

Table 3 Response rates to Wave 1 and Wave 2 for Survey 1

Incentive	Treatment						Average / Total
	1	2	3	4	5	6	
Wave 1	LQ	LQT	LQC	LQTC	LQ	LQ	% / Total
Wave 2	L	L	L	L	LQ	LTC	
Wave	%	%	%	%	%	%	
1	20.0	29.4	31.0	24.5	31.0	26.8	27.1
2	25.1	28.2	22.2	25.7	38.0	24.0	27.1
1+2	41.0	49.5	46.1	43.9	57.2	44.6	47.0
N (Wave 1)	300	293	297	294	297	298	1779
N (Wave 2)	239	206	203	222	205	217	1292

Note: L = Letter, Q = Questionnaire, T = Teabag, C = Coffee sachet Response rate reported for Wave 2 is based on the sample size for Wave 2, not the overall sample size

most effective by far was again Treatment 5, which used the default option for the first mail-out (questionnaire, reply paid envelope and cover letter, no tea or coffee incentive) coupled with a second mail-out to non-responders that included a replacement questionnaire, reply-paid envelope and cover letter (no incentive). Treatment 5 produced significantly higher response rates overall than all of the treatments (T5:T1, $\chi^2 = 15.74$, $df = 1$, $p < .001$; T5:T2, $\chi^2 = 3.56$, $df = 1$, $p < .05$; T5:T3, $\chi^2 = 7.34$, $df = 1$, $p < .01$; T5:T4, $\chi^2 = 10.55$, $df = 1$, $p < .001$; T5:T6, $\chi^2 = 9.46$, $df = 1$, $p < .001$). Thus for volunteers, in this sporting code at least, the tea and/or coffee incentives had a negligible effect on response rate, and so are not a cost-effective option. However, sending a replacement questionnaire was an effective way of increasing response rates.

Survey 2

The results for Survey 2 are reported in Table 4. These results are for Wave 2 only, as only non-respondents to Wave 1 were assigned to treatment groups, so the experiment was only conducted in Wave 2. As in Survey 1, the teabag incentive had no effect on response rates for either the Health volunteer group or the Sports volunteer group (Health: T1:T3 $\chi^2 = .393$, $df = 1$, $p > .5$; Sport: T1:T3 $\chi^2 = .002$,

$df = 1$, $p > .9$). However, consistent with the findings in Survey 1, the inclusion of a replacement questionnaire substantially increased response rates for both groups compared with the control (T1: letter only), although this increase was not statistically significant for Health. The results for Health were 46% for control (letter only) versus 57% for replacement questionnaire (T1:T2 $\chi^2 = 2.12$, $df = 1$, $p > .1$). For Sports the results were 18% for control versus 36% for replacement questionnaire (T1:T2 $\chi^2 = 5.47$, $df = 1$, $p < .05$).

A notable feature of the results in Survey 2 is the difference in response rate achieved for the two different groups of volunteers. The health group produced much higher response rates (41% to 58%) than the sports group (18% to 36%). However, the response rates for the sports group are very similar to those achieved by the sports group in Survey 1 (22% to 38%). In spite of these differences in response rate, the effect of sending a replacement questionnaire was dramatic for both groups in Survey 2, where it produced an increase in response rate of 23% ((57.6-46.6)/46.6) for the Health volunteers and 100% ((35.8-17.9)/17.9) for the Sports volunteers. By comparison, the increase in response in Survey 1 due to the replacement questionnaire was around 50%.

Table 4 Response rates to first reminder in Survey 2

Incentive	Treatment			Average
	1	2	3	
Wave 1	LQ	LQ	LQ	
Wave 2	L	LQ	LT	
	%	%	%	%
Health	46.1	57.1	41.4	48.1
N	89	84	87	260
Sport	17.9	35.8	17.6	23.8
N	67	67	68	202

Note: L = Letter, Q = Questionnaire, T = Teabag

Response rate to first mail-out was 58.6 % for Health and 37.7 % for Sport

Conclusions

The two studies reported in this paper indicate that both teabags and coffee sachets, either on their own or in combination, are not effective incentives in surveys of volunteers in both the health and sports sectors. This finding is consistent with findings reported for surveys of the general public (Gendall, Hoek and Brennan, 1998), and suggests that this result is likely to be generalisable across samples, at least in New Zealand. Of course, this conclusion does not preclude trials in other countries. The search for a suitable incentive for volunteer surveys continues.

However, what this study does demonstrate is that a significant increase in response rate can be achieved by sending a replacement questionnaire with the second mail-out, rather than simply sending a reminder letter. Again, this result

confirms similar findings reported elsewhere (Futrell and Lamb, 1974; Brennan, 2004; 2005; Brennan and Charbonneau, 2005), which suggests that sending a replacement questionnaire is best practice. This study also demonstrates that it is possible to achieve quite respectable response rates, in the order of 36%-58%, using just two mail-outs. A key question that still needs to be addressed, however, is whether a response rate of this magnitude is sufficient to alleviate the effects of non-response bias.

While this study is of course limited in terms of the groups of volunteers selected, and the types of incentives tested, and to a single country (New Zealand), any methodological research of this type requires replication and extension to determine whether the results are generalisable. The studies reported here make a contribution to this process.

REFERENCES

- Brennan, M. (1992a) Techniques for improving mail survey response rates. *Marketing Bulletin* 3, 24-37.
- Brennan, M. (1992b) The effect of a monetary incentive on mail survey response rates: New data. *Journal of the Market Research Society* April, 34 (2), 173-177.
- Brennan, M. (2004) A test of two procedures for increasing responses to mail surveys. *Marketing Bulletin* 15, Research Note 3, 1- 9.
- Brennan, M. (2005). The effect of a prize draw on response rates in a mail and mixed-mode (mail and web) survey. *Proceedings of Australian and New Zealand Marketing Conference*. Freemantle; University of Western Australia, CD Rom.
- Brennan, M. & Charbonneau, J. (2005) The colour purple: The effect of questionnaire colour on mail survey response rates. *Marketing Bulletin* 16, Research Note 5, 1 – 7.
- Church, A. (1993) Estimating the effect of incentives on mail survey response rates: A meta-analysis. *Public Opinion Quarterly* 57, 62-79.
- Dillman, D. (1978) *Mail and Telephone Surveys: The Total Design Method*. New York: John Wiley & Sons.
- Dillman, D. (1991) The design and administration of mail surveys. *Annual Review of Sociology*, 17, 225-249.
- Dillman, D. (2000) *Mail and Internet Surveys: The Tailored Design Method*. New York: John Wiley & Sons.
- Duncan, W. (1979) Mail questionnaires in survey research: A review of response inducement techniques. *Journal of Management*, 5, Winter, 39-55.
- Fox, R. J., Crask, M.R. & Kim, J. (1988) Mail survey response rate: A meta-analysis of selected techniques for improving response. *Public Opinion Quarterly*, 52, 467-491.
- Futrell, C. M. & Lamb, C.W. (1974) Effect on mail survey return rates on including questionnaires with follow-up letters. *Perceptual and Motor Skills*, 52, 11-15.
- Gendall, P., Hoek, J. & Brennan, M. (1998) The tea bag experiment: More evidence on incentives in mail surveys. *Journal of the Market Research Society* 40 (4), 347-351.
- Harvey, L. (1987) Factors affecting response rates to mailed questionnaires: A comprehensive literature review. *Journal of the Market Research Society* 29 (3), 341-353.
- Heberlein, T.A. & Baumgartner, R. (1978) Factors affecting response rates to mailed questionnaires: a quantitative analysis of the published literature. *American Sociological Review*, 43 (4), 447-462.
- Kanuk, L. & Berenson, C. (1975) Mail surveys and response rates: A literature review. *Journal of Marketing Research* 12, 440-453.
- Linsky, A. (1975) Stimulating responses to mailed questionnaires: A review. *Public Opinion Quarterly* 39, 82-101.
- Ryu, E., Couper, M. & Marans, R. (2006) Survey incentives: cash vs in-kind; face-to-face vs mail; response rate vs nonresponse error. *International Journal of Public Opinion Research* 18 (1), 89–106.
- Yu, J. & Cooper, H. (1983) A quantitative review of research design effects on response rates to questionnaires. *Journal of Marketing Research* 20, 36-44.



User-friendliness of answer formats – an empirical comparison

Sara Dolnicar, School of Management and Marketing, Marketing Research Innovation Centre (mric), University of Wollongong

Bettina Grün, Institut für Statistik und Wahrscheinlichkeitstheorie, Technische Universität Wien, Marketing Research Innovation Centre (mric)

Acknowledgements: This research was supported by the Australian Research Council (through grants DP0557257 and LX0559628), the Austrian Science Foundation (through grant P17382), the University of Wollongong through internal research grant schemes, and the Austrian Academy of Sciences (ÖAW) through a DOC-FFORTE scholarship for Bettina Grün. This study was presented at ANZMAC 2006.

Abstract

Response rates in market research are dropping (Bednell & Shaw 2003; Hardy & Kotsomitis 2005). One way to address this problem is to make questionnaires more user-friendly. A key aspect of user-friendliness is the way in which respondents are asked to answer a question. Although much research has been undertaken in comparing reliability and validity of alternative answer formats, only very few studies have compared user-friendliness of answer formats. This study contributes to this field of research by conducting an empirical study comparing five different answer formats: binary (yes-no), 3-point, 7-point, continuous and percentage. Results indicate that the binary, 3-point and 7-point scales were preferred by respondents. These three answer formats were also perceived as equally pleasant and equally capable of capturing the respondents' opinions. The binary format required the least time to be administered.

Introduction

The questionnaire is the dominant measurement instrument in market research. Consequently, a large number of studies have been conducted in the past to determine how questionnaires can best be designed to measure what they are supposed to measure and to do so in a reliable manner that is practically feasible. One aspect of questionnaire design is the answer format. The answer format is the way in which respondents are asked to answer a question. Consequently answer formats have a major impact on the validity of questions. But they also impact the user-friendliness and efficiency of the survey.

The most comprehensive review of research investigating the effects of the number of answer options conducted in the field of marketing (Cox, 1980) concludes that two and three-point scales are generally inadequate and recommends the use of a 7-point scale. This

recommendation appears to have influenced marketing research practice significantly, which is currently dominated by the use of 5 or 7-point ordinal answer formats (Van der Eijk, 2001).

This is surprising and concerning, given that most of the prior work comparing answer formats has led to contradictory conclusions with respect to which answer format is optimal, typically using criteria such as reliability and validity. With respect to reliability the majority of studies conclude that the answer scale is not related to reliability (Bendig, 1954; Peabody, 1962; Komorita, 1963; Komorita and Graham, 1965; Matell and Jacoby, 1971; Jacoby and Matell, 1971; Remington, Tyrer, Newson-Smith and Cicchetti, 1979; Preston and Colman, 2000), although a number of studies come to the opposite conclusion (Symonds, 1924; Nunnally, 1967; Jones, 1968; Oaster, 1989; Finn, 1972; Ramsay, 1973).

Similarly studies investigating validity also do not lead to clear conclusions. Three studies result in the finding that no difference in validity can be found in dependence of the answer format used (Matell and Jacoby, 1971; Jacoby and Matell, 1971; Preston and Colman, 2000). Two studies come to the opposite conclusion claiming higher validity of answer formats with higher numbers of scale points (Loken, Pirie, Virnig, Hinkle and Salmon, 1987; Hancock and Klockars, 1991).

Later, Chang (1994) demonstrated methodological flaws in some of the reliability and validity comparisons that failed to decompose systematic method variance and trait variance. As a consequence, more answer options appeared to be more reliable although this was only a consequence of the restriction of range effect (Nunnally, 1970; Cohen, 1983; Martin, 1973; 1978). Chang found criterion related validity to be independent of the number of answer options and reliability values were in fact better when an answer format with fewer scale points was used.

The reliability and validity comparisons consequently do not provide guidance to market researchers as to which answer format they should be using. Consequently, other criteria need to be investigated, one of which is user-friendliness. This area of research has not received much attention in the past. Jones (1968) investigated the matter very early. Later Preston and Colman (2000), Dolnicar (2003) and Dolnicar and Grün (in press) made contributions to the area. Jones (1968) concludes that respondents prefer multiple categories in answer formats. Preston and Colman (2000) find that individuals can better express their feelings if a larger number of answer options is provided. The perceived speed of questionnaire completion, however, increases with fewer answer options. These results are of high practical importance given that fatigue effects resulting from long questionnaires are known to have negative consequences on

data quality (Herzog and Bachman, 1981; Johnson, Lehmann and Horne, 1990). The study conducted by Dolnicar (2003) supports the latter findings. Results from her empirical study suggest that the use of the binary format in the context of brand image measurement leads to a significant increase in efficiency with respondents requiring 30% less time to complete a binary questionnaire as opposed to multi-category questionnaires. Furthermore, respondents expressed that the multi-category answer format was significantly more difficult to answer than the binary one. Dolnicar and Grün (in press) empirically compared binary, 7-point and continuous answer formats for two different constructs (behavioural intentions and attitudes) and conclude that the binary format is quicker and perceived as quicker by respondents. No differences could be detected with respect to the perceived pleasantness, the perceived ability of respondents to express their feelings and perceived simplicity of the questionnaire.

A few other authors comment on the speed aspect which is closely related to the economic efficiency of survey research: Payne (1951), Dillman (1978), Bradburn and Sudman (1979), Churchill (1979) and Peterson et al. (1982) warn researchers from using too many answer categories. Komorita and Graham (1965, p. 989) make a very bold recommendation based on the efficiency argument: "the major implication is that, because of simplicity and convenience in administration and scoring, all inventories and scales ought to use a dichotomous, two-point scoring scheme."

Another concern when choosing an answer format is the susceptibility to response styles. In this context Cronbach (1950, p. 21) stated that "Since response sets are a nuisance, test designers should avoid forms of items which response sets infest". Cronbach consequently recommends the use of binary answer formats to avoid these effects. Since Cronbach's recom-

mendation a number of studies have been undertaken into response styles, mainly those that manifest themselves on multi-category answer formats. A number of identification and correction techniques were recommended (Cunningham, Cunningham and Green, 1977; Greenleaf, 1992a; 1992b; Heide and Gronhaug, 1992; Watson, 1992; Van de Vijver and Poortinga, 2002; Welkenhuysen-Gybels, Billiet and Cambre, 2003). However, none of them actually solve the problem at its source because they make assumptions about the nature of data contamination and undertake a correction which in itself can potentially distort the data. A recent study (Lee, Soutar, Louviere and Daly, 2006) used best worst scaling to avoid answer formats; the authors conclude that – for certain kinds of items - best worst scaling is a very efficient way to address the problem at its source rather than having to manipulate data ex post to eliminate distortions due to response styles.

Research Questions

The aim of the present study is to investigate respondents' preferences for certain answer formats. We use a choice-based approach in which respondents have the opportunity to first assess alternative answer formats and then make an informed decision which one they would prefer to use rather than imposing an answer format on them.

The following propositions underlie our study.

- Proposition #1: Respondents have clear preferences for certain answer formats.
- Proposition #2: Respondents prefer multi-category answer formats because they are widely used and respondents are familiar with them.
- Proposition #3: Women have different preferences for answer formats than men.
- Proposition #4: Cultural background affects answer format preference.
- Propositions #3 and #4 emerge from the response style literature which demonstrates strong effects of socio-demographics on the use of answer formats.
- Proposition #5: Answer formats differ in the time needed to complete the survey.
- Proposition #6: Respondents perceive different answer format as different in terms of pleasantness of the questionnaire.
- Proposition #7: Respondents feel that they can express their feelings better using some answer formats than others.

Data

Data was collected among first year marketing students at the University of Wollongong. The topic of the survey was water recycling. Two different questionnaires were handed out: one included behavioural intentions with respect to recycled water and the other brand image evaluation. Both surveys also included a shortened version of the scale known as New Environmental Paradigm (Dunlap, Van Liere, Mertig and Jones, 2000) which measures environmental attitudes. The questionnaire type (behavioural intentions or brand image evaluation) was randomly assigned to students. The students were informed that they could choose their favourite answer format out of five different formats. The five answer formats were binary (yes – no), 3-point scale, 7-point scale, metric (respondents were asked to mark a point on a line) and percentage (respondents were asked to indicate a percentage). The answer formats were presented on a separate sheet together with an exemplary question. The respondents were asked to (1) study the options, (2) indicate their favourite answer for-

mat and (3) complete the corresponding questionnaire. In addition, the following information was collected from students: the actual beginning and end time of completing the questionnaire, language spoken with their parents, perceived simplicity ("Simple"), perceived pleasantness ("Pleasant"), perceived speed ("Quick") and perceived ability to express their feelings ("Express feelings"). The responses on the perception of scale were recorded in the same way for all questionnaire versions, namely using a five-point bipolar ordinal scale. The survey resulted in 263 completed questionnaires. Of these 263, 136 (52%) students had completed the brand image questionnaire and 127 the version including behavioural intentions (48%). 155 respondents (59%) were female and 107 (41%) male. For 188 (72%) respondents English is the main language they speak with their parents, while it is Chinese for 40 (15%) of the respondents. In addition there are only nine respondents speaking a former Yugoslavian language, two speaking Greek and one speaking Italian. The remaining 20 (8%) of the respondents indicated that they speak another language at home.

All computations and graphics for the empirical analysis have been done using the R statistical software package (R Development Core Team, 2006).

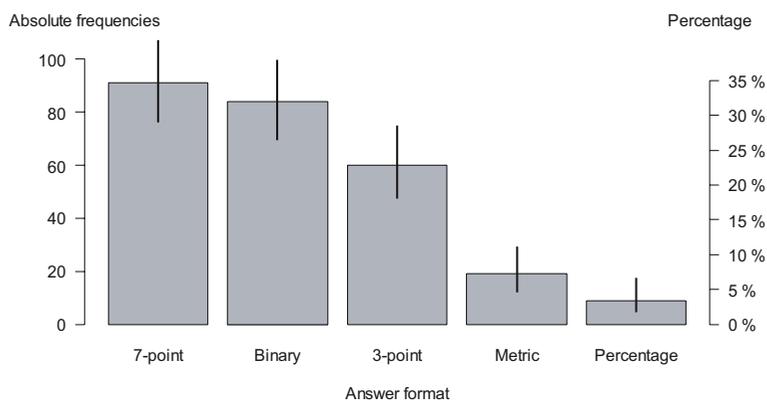
Results

Proposition #1: Respondents have clear preferences for certain answer formats

Proposition #2: Respondents prefer multi-category answer formats because they are widely used and respondents are familiar with them.

The first research question is if a favourite answer format exists. To investigate this question the absolute and relative frequencies of answer format choice are determined together with the 95% confidence intervals given the proportion for each answer format. The results are depicted in Figure 1. The results indicate that no single most popular answer format exists. However, multi-category answer formats (binary, 3-point, 7-point) are generally preferred to formats where the answer is recorded on a nearly continuous scale, such as the metric and percentage scale. While the 7-point scale is the most popular, the binary scale is chosen almost as frequently. The 3-point scale is significantly less popular than the 7-point scale with respect to a significance level of 5%, while the null hypothesis of equal proportions can not be rejected for a comparison of binary with 3-point scale. Percentage is the least popular scale, but the difference to the metric scale is not significant.

Figure 1: Absolute (relative) frequency of answer format choice with 95% confidence interval



Proposition #1 (clear preferences exist) is supported. Although no single most preferred answer format can be determined, the two multi-category and the binary formats are clearly preferred to the metric versions. Proposition #2 (respondents prefer multi-category answer formats) is not supported: binary answer format is equally popular as multi-category formats are.

Proposition #3: Women have different preferences for answer formats than men.

Proposition #4: Cultural background affects answer format preference.

The relative frequencies of the answer format choices cross-tabulated with gender and language are given in Figure 2. Relative frequencies are used as the absolute values are not directly comparable between the groups as they differ in overall size.

Fisher's exact test is used to investigate if there is a significant association between choice of answer format and one of the concomitant variables. The null hypothesis of independence can not be rejected for gender (p -value = 0.44). For the analysis of language only those respondents with language English (188 respondents)

and Chinese (40 respondents) are included. The null hypothesis of independence between answer format and language can not be rejected at a significance level of 0.05 (p -value = 0.06), but the p -value is rather low and the power of the test is decreased as the Chinese sample is rather small including only 15% of the respondents. Figure 2 indicates that the Chinese respondents tend to choose the 3-point scale more often while in return the 7-point scale is less popular.

Proposition #3 (gender effect) and #4 (culture effect) are not supported.

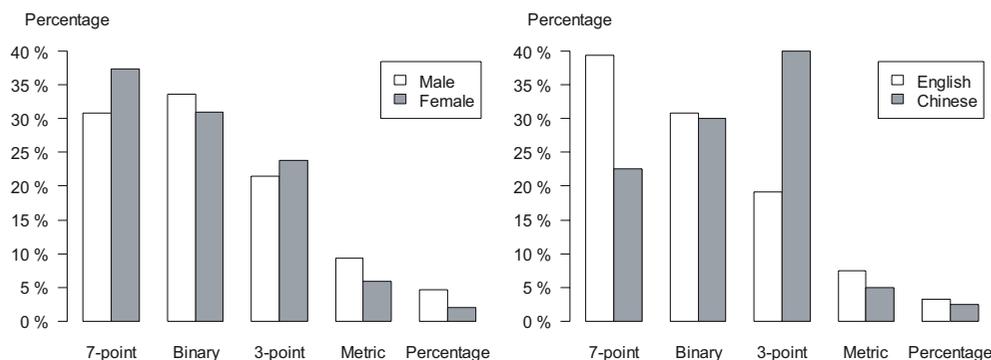
Proposition #5: Answer formats differ in the time needed to complete the survey.

Proposition #6: Respondents perceive different answer format as different in terms of pleasantness of the questionnaire.

Proposition #7: Respondents feel that they can express their feelings better using some answer formats than others.

The duration and respondents' perception of alternative answer formats is analysed using ANOVA with linear models. In the analysis of duration only those respon-

Figure 2: Relative frequencies of answer format choice depending on gender and language



dents (97%) were included where end minus begin time gave a positive number assuming that the respondents needed at least one minute to complete the questionnaire. For the analysis of respondents' perception of alternative answer formats equal distances between the five answer categories of the perception question are assumed. Analyses of variance are conducted: duration and perception are the dependent variables and different answer formats and questionnaire type are the independent variables for the computation. The logarithm of the duration variable is used because the original duration variable is distributed skewed to the right. Results are provided in Table 1.

The ANOVA for duration indicates that there is a significant influence of questionnaire type and answer format, but no interaction effect. With respect to respondents' perceptions of alternative answer formats the interaction effect is not significant for any of the variables. This means

that the influence of the answer formats on the evaluation of the questionnaires does not depend on the questionnaire type. This can be seen as a reassurance that the conclusions drawn with respect to answer format can be generalized to other questionnaires.

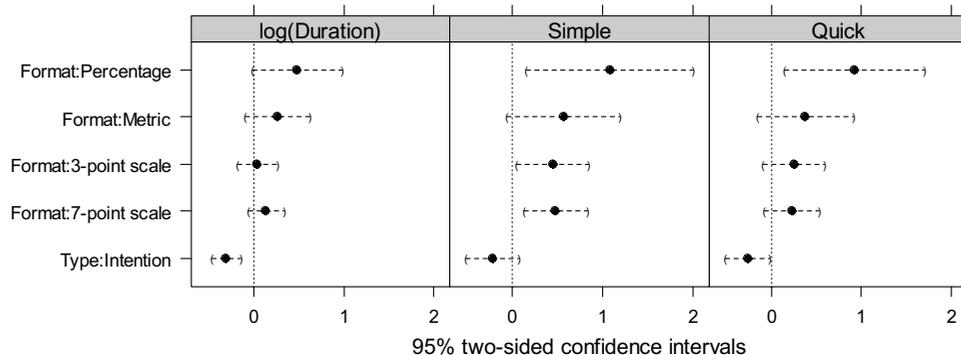
The influence of the questionnaire type is always significant, while the answer format leads to a significant difference only for the items "Simple" and "Quick". It can consequently be concluded that all answer formats are perceived as equally "Pleasant" and suitable to "Express feelings". Simultaneous confidence intervals are derived for the coefficients of the relevant linear models without interaction effect using Dunnett's method (Dunnett, 1955) and are given in Figure 3. The intervals are plotted as horizontal lines where the limits of the intervals are given by round brackets and the estimates by a point. The binary scale is used as base category for the answer format. It can

Table 1: ANOVA for duration and perception of scale depending on answer format and questionnaire type

	log(Duration)			Simple			Pleasant		
	Df	F value	P value	Df	F value	P value	Df	F value	P value
Type	1	23.16	<0.001	1	4.62	0.033	1	7.47	0.007
Format	4	2.62	0.036	4	5.12	<0.001	4	0.74	0.566
Type:Format	4	1.31	0.266	4	1.34	0.257	4	0.15	0.961
Residuals	226			244			244		

	Quick			Express feelings		
	Df	F value	P value	Df	F value	P value
Type	1	8.09	0.005	1	6.63	0.011
Format	4	3.31	0.012	4	0.20	0.938
Type:Format	4	1.04	0.385	4	0.96	0.432
Residuals	244			244		

Figure 3: Simultaneous confidence intervals for duration and perception of scale



be seen that with respect to duration or perception the questionnaire including the behavioural intentions always has lower or equal scores. With respect to simplicity the percentage, 3-point and 7-point scale are perceived as significantly less simple than the binary scale. With respect to quickness a significant difference to the binary scale can only be determined for the percentage scale.

Proposition #5 (answer formats differ in speed) is supported. The continuous answer formats took longer to complete. Proposition #6 (answer formats differ in perceived pleasantness) and Proposition #7 (answer formats differ in respondents' perceived ability to express feelings) is not supported: no differences in perceived ability to express feelings can be determined.

Conclusions and Future Work

Respondents have different preferences with respect to answer formats. No single one answer format is preferred by the majority of respondents, but a clear preference of ordinal and binary answer formats was determined as opposed to metric formats. The answer format preference does not depend on gender and dependence on the cultural background could not be assessed well enough in this study given the small number of non-English-speaking respondents. In terms of user-friend-

liness, the continuous answer formats were found to take longer to complete, all formats are perceived as equally pleasant and as providing the same opportunity to express feelings. With respect to simplicity and quickness the respondents choosing the binary scale gave the most favourable evaluation. These evaluations could potentially be blurred by the fact that the answer format was not imposed on the respondents, but that they actually chose it themselves and thus might be evaluating the answer format they chose as more favourable than they would have evaluated the same scale if imposed on them.

The implications for marketing researchers from this study are that certain answer formats are indeed preferred by respondents and can thus be used to make the market research experience as pleasant as possible and, as a consequence increase data quality and the willingness of respondents to participate in survey research. Interestingly, binary format is among the most preferred, which opens up the opportunity of not only offering respondents a pleasant scale but also benefiting from the quickness of this format and saving fieldwork cost. Optimally, the suitability and preference of answer formats should be pre-tested for the construct under study before the actual fieldwork phase is undertaken.

REFERENCES

- Bednell, D. H. B., and Shaw, M. (2003) Changing Response Rates in Australian Market Research, *Australasian Journal of Market Research*, 11(1), 31-41.
- Bendig, A. W. (1954) Reliability and the Number of Rating Scale Categories. *Journal of Applied Psychology* 38(1), 38-40.
- Bradburn, N., and Sudman, S. (1979) *Improving Interview Method and Questionnaire Design*. San Francisco: Jossey-Bass
- Churchill, G. A., Jr. (1979) A Paradigm for Developing Better Measures of Marketing Constructs. *Journal of Marketing Research* 16(1), 64-73.
- Chang, L. (1994) A Psychometric Evaluation of Four-point and Six-point Likert-type Scales in Relation to Reliability and Validity. *Applied Psychological Measurement* 18, 205-215.
- Cox, E. P. (1980) The Optimal Number of Response Alternatives for a Scale: A Review. *Journal of Marketing Research* 17 (4), 407-422.
- Cohen, J. (1983) The Cost of Dichotomization. *Applied Psychological Measurement* 7 (3), 249-253.
- Cronbach, L.J. (1950) Further Evidence on Response Sets and Test Design. *Educational and Psychological Measurement* 10, 3-31.
- Cunningham, W. H., Cunningham, I. C., and Green, R. T. (1977) The Ipsative Process to Reduce Response Set Bias. *Public Opinion Quarterly*, 41(3), 379-384.
- Dillman, D.A. (1978) *Mail and Telephone Surveys: The Total Design Methods*, John Wiley, New York.
- Dolnicar, S. (2003) *Simplifying Three-Way Questionnaires - Do the Advantages of Binary Answer Categories Compensate for the Loss of Information?* ANZMAC CD Proceedings.
- Dolnicar, S., and Grün, B., in press. How Constrained a Response: A Comparison of Binary, Ordinal and Metric Answer Formats. *Journal of Retailing and Consumer Services*.
- Dunnett, C. W. (1955) A Multiple Comparison Procedure for Comparing Several Treatments with a Control. *Journal of the American Statistical Association* 50(272), 1096-1121.
- Dunlap, R. E., Van Liere, K. D., Mertig A. G., and Jones R. E. 2000. Measuring Endorsement of the New Ecological Paradigm: A Revised NEP Scale. *Journal of Social Issues* 56(3), 425-442.
- Finn, R. H. (1972) Effects of Some Variations in Rating Scale Characteristics on the Means and Reliabilities of Ratings. *Educational and Psychological Measurement* 32 (2), 255-265.
- Greenleaf, E. A. (1992a) Improving Rating-Scale Measures by Detecting and Correcting Bias Components in Some Response Styles. *Journal of Marketing Research*, 29(2), 176-188.
- Greenleaf, E. A. (1992b) Measuring Extreme Response Style. *Public Opinion Quarterly*, 56(3), 328-351.
- Hancock, G. R., and Klockars, A. J. (1991) The Effect of Scale Manipulations on Validity: Targeting Frequency Rating Scales for Anticipated Performance Levels. *Applied Ergonomics* 22 (3), 147-154.
- Hardy, T., and Kotsomitis, N. (2005) *The F Word in Market Research: High Impact Public Relations in the 21st Century*. CD Proceedings of the Australian Market & Social Research Society National Conference 2005.
- Heide, M., and Gronhaug, K. (1992) The Impact of Response Styles in Surveys: A Simulation Study. *Journal of the Market Research Society*, 34(3), 215-223.
- Herzog, A. R., and Bachman, J. G. (1981) Effects of Questionnaire Length on Response Quality. *Public Opinion Quarterly*, 45(4), 549-559.
- Jacoby, J. and Matell, M. S. (1971) Three-Point Likert Scales Are Good Enough. *Journal of Marketing Research* 8, 495-500.
- Johnson, M. D., Lehmann, D. R., and Horne, D. R. (1990) The Effects of Fatigue on Judgments of Interproduct Similarity, *International Journal of Research in Marketing*, 7, 35-43.
- Jones, R. R. (1968) *Differences in Response Consistency and Subjects' Preferences for Three Personality Inventory Response Formats*. Proceedings of the 76th Annual Convention of the American Psychological Association, 247-248.

- Komorita, S.S. (1963) Attitude Content, Intensity, and the Neutral Point on a Likert scale. *Journal of Social Psychology* 61, 327-334.
- Komorita, S.S. and Graham, W. K. (1965) Number of Scale Points and the Reliability of Scales. *Educational and Psychological Measurement* 25(4), 987-995.
- Lee, J. A., Soutar, G. N., Louviere, J., and Daly, T. M. (2006) *An Examination of the Relationship between Values and Holiday Benefits across Cultures using Ratings Scales and Best-Worst Scaling*. ANZMAC CD Proceedings.
- Loken, B., Pirie, P., Virnig, K. A., Hinkle, R. L., and Salmon, C. T. (1987) The Use of 0-10 Scales in Telephone Surveys. *Journal of the Market Research Society* 29 (3), 353-362.
- Martin, W. S. (1973) The Effects of Scaling on the Correlation Coefficient: A Test of Validity. *Journal of Marketing Research* 10(3), 316-318.
- Martin, W. S. (1978) Effects of Scaling on the Correlation Coefficient: Additional Considerations. *Journal of Marketing Research* 15(2), 304-308.
- Matell, M. S. and Jacoby, J. (1971) Is There an Optimal Number of Alternatives for Likert Scale Items? Study I: Reliability and Validity. *Educational and Psychological Measurement* 31, 657-674.
- Nunnally, J. C. (1967) *Psychometric Theory*. New York: McGraw-Hill, 1st edition.
- Nunnally, J. C. (1970) *Introduction to Psychological Measurement*. McGraw-Hill, New York.
- Payne, S. L. (1951) *The Art of Asking Questions*. Princeton, NJ: Princeton University Press.
- Peabody, D. (1962) Two Components in Bipolar Scales: Direction and Extremeness. *Psychological Review* 69(2), 65-73.
- Peterson, C. R., Semmel, A., von Baeyer, C., Abramson, L. Y., Metalsky, B. I. and Seligman, M. E. P. (1982) The Attributional Style Questionnaire. *Cognitive Therapy and Research* 6 (3), 287-300.
- Preston, C. C., and Colman, A. M. (2000) Optimal Number of Response Categories in Rating Scales: Reliability, Validity, Discriminating Power, and Respondent Preferences. *Acta Psychologica* 104, 1-15.
- Oaster, T. R. F. (1989) Number of Alternatives per Choice Point and Stability of Likert-type Scales. *Perceptual and Motor Skills* 68, 549-550.
- R Development Core Team (2006) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Ramsay, J. O. (1973) The Effect of Number of Categories in Rating Scales on Precision of Estimation of Scale Values. *Psychometrika* 37, 513-532.
- Remington, M., Tyrer, P. J., Newson-Smith, J. and Cicchetti, D.V. (1979) Comparative Reliability of Categorical and Analogue Rating Scales in the Assessment of Psychiatric Symptomatology. *Psychological Medicine* 9, 765-770.
- Symonds, P.M. (1924) On the Loss of Reliability in Ratings Due to Coarseness of the Scale. *Journal of Experimental Psychology* 7, 456-461.
- Van de Vijver, F. J. R. and Poortinga, Y. H. (2002) Structural Equivalence in Multilevel Research. *Journal of Cross-Cultural Psychology*, 33(2), 141-156.
- Van der Eijk, C. (2001) Measuring Agreement in Ordered Rating Scales. *Quality & Quantity* 35, 325-341.
- Watson, D. (1992) Correcting for Acquiescent Response Bias in the Absence of a Balanced Scale: An Application to Class Consciousness. *Sociological Methods and Research*, 21(1), 52-88.
- Welkenhuysen-Gybels, J., Billiet, J. and Cambre, B. (2003) Adjustment for Acquiescence in the Assessment of the Construct Equivalence of Likert-type Score Items. *Journal of Cross-Cultural Psychology*, 34(6), 702-722.

Interested in becoming an AMSRS member?

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 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), vote in divisional and national elections of the AMSRS and receive the Society's Journal AJMSR.

Note: Student members receive limited membership benefits so please see 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.

JoinToday

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

Pressure to innovate: Does validity suffer?

Linda Brennan

**Associate Professor in Marketing, Swinburne University
John Street Hawthorn Vic 3122
lbrennan@swin.edu.au**

Jim Camm

**Associate Director, Automotive Practice Group
ACNielsen Melbourne**

Introduction

Imagination is a necessary precondition for innovation and subsequent advancement in any field. We accept this as self-evident. But it is legitimate to ask the question whether or not all innovation is 'good' for the field of market research, or can innovation also start to lead us astray?

The pressure to innovate

There is, no doubt, a pressure from clients to innovate. The problems of marketing are complex and often appear to be intractable, relating as they do, to the behaviour of customers. Therefore, there will always be some unfinished business, some unturned stone or a yet-to-be-explored angle awaiting every brand and marketing manager. Furthermore, there is constant pressure on these same marketers to get an edge on their competitors. Competitive advantage does not come about by doing more of what the competition is already doing. It comes about by being out there and taking the lead in the market, being creative, taking risks, and dealing with the consequences in a complex and rapidly changing business environment (Hult, Ketchen, and Slater 2005; Vanderkaay 2005). Indeed, there are more success stories in marketing that relate to risk taking and imagination than there are relating to in-depth analysis and understanding (Temporal 2006). However, there is no 'correct' way to take risks or be creative. Each organisation must make its own decisions. Hence a key value of research is

to provide information that assists in the creative process – or stops it in its tracks before it becomes too risky.

On the supplier side, there is the desire to differentiate and position their offerings in order to improve margins. It is a truism in marketing that undifferentiated, commodity-like offerings have the lowest margins (cf. Wolfe 1977; Clemons, II-Horn, and Hitt 2002; Hill 1990). Related to the complexity of the marketplace is the observation that it is often the case that 'out-of-the-box', validated methodologies do not match the specific needs presented by a new marketing problem.

Of course, on both sides, there is also a desire to reduce costs. However, one important risk of extreme cost reduction strategies is the standardisation of the service, which in turn can also encourage the application of a methodology to an inappropriate situation.

All these factors combine to create demand for something new and different. Via the mechanism of competition, the research community supplies this demand by offering innovative methodologies. Barker (2002 p 152) put it succinctly:

'...the world of commercial qualitative research is awash with apparently new techniques, new models, new approaches. The self image of the sector is of creative developers, inspired gurus,

bearers of always new wisdom and unique insight to the tables of brand managers and advertisers. The commercial pressures to innovate (occasionally for the sake of it) are indeed very great – client research briefs often literally beg for new ways of approaching problems. We know our competitors will be proposing living with a sample of families for two weeks in order to gain those hidden nuggets of insight, which will provide the leverage that brand of cat food really needs; so we need to go one step further, one step wackier.'

Research can be boring (McAnena 2004). However, the quest for imagination in research has been with us for a long time. In 1961 Blankenship (p 34) wrote: *'Creativity in research provides no defense (sic) for use of inadequate research standards.'*

Market research innovates across disciplinary boundaries

There are many sources of innovation in market research. First, there is the transfer of methodology developed in other fields into marketing research. Second, there is transfer of knowledge. For example, concepts in psychology are adapted into market research. Finally, there are extensions of existing techniques within the domain of market research itself.

On the whole however, market research is a theory borrowing discipline, much like the better documented case for marketing (Murray, Evers, and Janda 1995; Deshpandé 1983). Our practitioners bring their theories from a wide variety of disciplinary backgrounds; ranging from applied statistics to anthropology. While there is value in the variety of approaches to solving problems offered by this diversity, there are dangers that researchers will adopt ideas from others without a full understanding of the theoretical underpinnings.

Within market research, therefore, there is a large number of methodologies and approaches that have been borrowed from other domains and used with new populations, situations and contexts. In addition, both these borrowed methodologies and those developed within market research are 'tinkered with'. By this we mean that practitioners change the methodology without considering the effect it will have on the validity of the methodology. For example, ethnography is a trendy 'method' that is applied as a term without much thought as to the ontological and epistemological foundations, nor the methodologies that were originally validated in the discipline. Participant-observation, for example, does not mean simply observing; it means actively participating, reflecting and observing, writing about the observation and being assimilated into the community under consideration. Most clients won't pay for this kind of research so it rarely takes place; although there is much non-participant 'observation' that is passed off as 'ethnography.'

These borrowings and tinkering helped shift the market research discipline into a credible and scientific status within the domain of business decision-making. However, we argue that the market research discipline got lost on the way to true scientific status.

This is because:

1. There is pressure to innovate, and
2. Opportunities exist to borrow and tinker with valid methodologies which leads to the proliferation of methodologies that have not been validated or revalidated for the new populations, situations and contexts they are used for. This in turn leads to the generation of research findings that might or might not be correct.

Ways that established methods can be invalidated

The argument rests squarely on what we mean by 'valid'. Therefore, we now review the basic concepts of validity. Figure 1 illustrates in a novel way the many types of validity as summarised from the typology provided in Appendix A. The concepts of validity and reliability hold a central place in measurement science. In this paper, for the purpose of having an in-depth look at one issue at a time and to minimise potential uncertainty, we put reliability to one side and discuss only validity. We note in passing, however, that there is confusion in the literature with many authors apparently using the terms interchangeably. For now, a working definition of reliability is 'that a measure will yield the same result if repeated under the same conditions' and a working definition of validity is 'that a measure measures what it purports to measure'. An early example of a validation issue in market research is provided by Blankenship (1961):

'Many years ago a researcher didn't believe the answers he was getting on direct questions as to the publications people read. Everyone reported 'prestige' publications; few admitted to readership of the low prestige type. He got a very different picture when he had solicitors collect magazines from the same homes "for charitable purposes." What people said they read, on direct question, and what they actually read, on the basis of magazines in the home, were quite different things' (p 35).

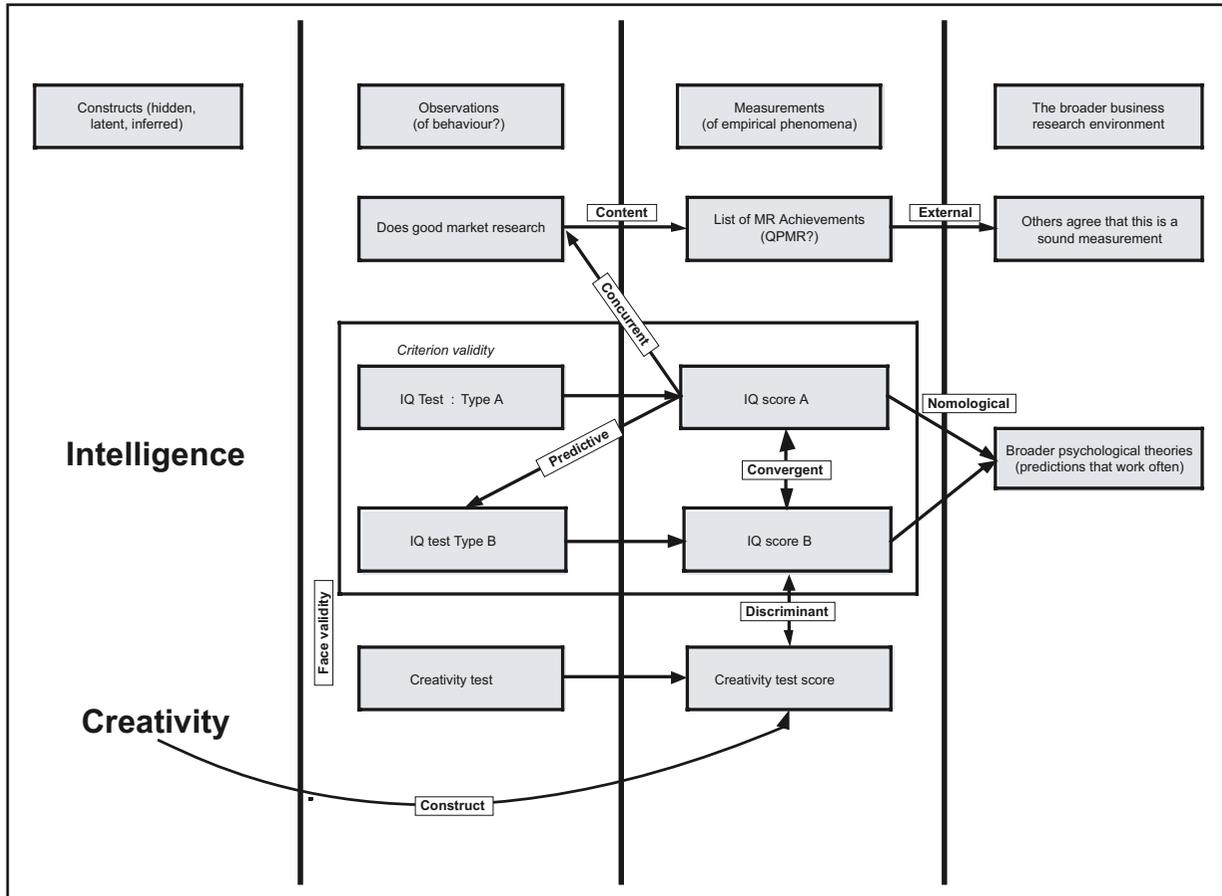
Quite simply, the methodology of direct questioning was found to be invalid. It did not measure what it purported to measure. It does have what is known as 'face' validity but fails the test of content validity.

Face validity exists when the measure 'looks as if' (on the face of it) it should measure a particular attribute. This is

usually easy for others to agree with, so it is also called consensus validity. Both terms are often used synonymously with content validity. However, face validity does not imply expert acceptance of the measure's validity as does content validity. Interestingly; there is a large volume of research based on the assumption that face validity has been established before all measurement takes place. However, because it is such a base assumption, there is no way to test face validity outside of the context. We argue that the words face validity have two meanings. First, a face belongs to an individual idea. For example in Figure 1 – the construct of 'intelligence' is a single idea, which may or may not be uni-dimensional. Second, 'on the face of it' means simply that it 'looks as if' it would be OK to measure a phenomenon. However, we never ask - who does it look that way to? In the establishment of face validity, an important but often missing question is - whose face? Whose perspective are we looking from? For example, few people in Melbourne would argue that a 24 degree day with sunshine is a 'nice day'. We could 'on the face of it' establish 24 degrees and sunshine as two separate variables of the construct 'nice day.' However, if we do not know which temperature measure we are using (Fahrenheit or Celsius) then the 'nice day' could be very different depending on whether you are located in the US which uses Fahrenheit or somewhere which uses Celsius. Furthermore, artists might prefer clouds and brooding skylines, pilots might be concerned about invisible turbulences and drought stricken farmers might prefer lots of rain. Thus, perspective is important in the establishment of validity.

The usual method of establishing face validity in the social sciences is often a relatively cursory process (Rossiter 2002). Most validation is done with university students and academics (Basil, Brown, and Bocarnea 2002). These people are not representative of the general population. Market researchers are usually adamant

Figure 1: Relationships between ‘Validities’



about random sampling but if face validity is established with people who are not representative of the population then the first step is flawed.

Figure 1 illustrates that only face and content validity can be established in the absence of other measures because all other forms are defined in terms of their relationship to other variables or measures. That is, validity of one measure is based on the supposed validity of another measure. We argue that if the first step is flawed, then so is the rest of the process.

In order to describe these relationships, imagine a hypothetical group of market researchers who have both Intelligence

and Creativity (well don't we all?). In the diagram these are shown under the heading 'Constructs (hidden, latent, inferred)'. In the next column we have 'Observable Behaviour', which includes performing two different IQ tests and a creativity test. In addition, we note that these market researchers do 'good' market research. In the next column we show measurements that could arise relating to each of the behaviours. For instance, an IQ test score arises from undertaking an IQ test and we could assemble list of market research achievements for each researcher who does good market research (cf. the QPMR accreditation requirements). The rightmost column represents two aspects of the broader environment. Let us imagine that these

market researchers first do a test that purports to measure their creativity and obtain a creativity test score:

Construct validity is the extent to which the test score is a measure of what we understand by 'creativity'. The creativity test is made up of a large number of individual test items.

Discriminant validity is the extent to which the IQ and Creativity tests differentiate between the two constructs 'creativity' and 'IQ'. That is to say if we believe that a person could have high IQ and low creativity, or vice versa, then the tests will be able to show this:

Convergent validity is the extent to which the two IQ tests provide the same results. That is to say if some of our market researchers scored highly on one IQ test but obtained a low score on the other we have cause to doubt that the tests are measuring the same thing. Let us assume that there is a belief that a high IQ score indicates that the researcher will perform good market research:

Predictive validity is the extent to which the test score for IQ predicts how well researchers conduct market research. It might well be the case that IQ does not correlate at all with good quality market research and it is found that creativity is a far better predictor. The question then arises of how we know whether a researcher is doing good work? The measure we have suggested is to assemble a list of market research achievements. Researchers with longer lists of achievements would then be said to have done better research than those with shorter lists of achievements. Of course, for a younger researcher another criterion for 'good' would have to be found as the length of the list could simply be due to length of tenure. One could assume that length of tenure in the market research industry is related to capability – very few researchers last long if they are not good at the job:

Concurrent validity is the extent to which the test score could, for instance, differentiate senior managers from factory workers. It is similar to Predictive validity, but does not imply future activity. Predictive validity and Concurrent validity are usually considered to be variants of Criterion validity. In our researchers' situation, an IQ test would potentially differentiate between levels of verbal and/or numerical skills. This could help us decide whether or not we want to be quantitative analysts or qualitative specialists (or both).

Criterion validity is the measurement of how well one variable (or set of variables) predicts or correlates with an outcome based on information from other variables (includes concurrent, predictive and diagnostic validity). In this case our IQ tests are related to each other but may be used to determine different elements of IQ depending on the needs of the person testing. One variable (e.g. numerical skills) might be sufficient in some circumstances and can be used as 'valid' alternative to a battery of tests.

Content validity is the extent to which a test is made up of a representative sample from the domain of interest. We note that the issue of circularity, which is inherent in many aspects of validity, now arises. How do we know that the items we have chosen correlate with 'good market research' unless we already have an idea in mind for what constitutes 'good' market research? This is a vexing question. It is partly answered by the concept of 'face validity', which is not shown as a set of relationships on the diagram because it should be applied in all cases:

Face validity is the extent that the measurement measures the quality that it purports to measure on the basis of a common sense assessment. Our hypothetical group of market researchers are only a small subset of all market researchers. Having established that the 'list of

achievements' is a valid measure of quality of market research, we can now ask the question whether or not the idea of using this type of measure is applicable for the broader market research community, or even for other professions. For instance, is 'list of achievements' a valid measure for the quality of social work, police work or sales?

External validity is the extent to which the results can be generalised to other populations, situations or conditions. Finally, we return to our IQ scores. It is reasonable to ask whether the ideas we have developed – that creativity is different to IQ and that IQ, not creativity, is a good predictor of quality of market research as measured by lists of achievements – fits in with broader psychological and social theories.

Nomological validity is the extent to which the pattern of results for a theoretical network is consistent with broader theoretical networks. For instance, it could be the case that there is strong evidence in another domain, say, sales, where it has been shown that IQ does not predict performance at all and that lists of self-reported achievements are poor indicators of sales ability.

In Appendix A we provide a more detailed review of the validity literature, and introduce the classifications 'Formative Validity' and 'Prognostic Validity' as organising concepts for the various measures of validity.

So if we know all that, where do things go wrong?

Now that we have established some of the concepts relating to validity, we turn now to three specific examples of how methodologies become invalid.

Example 1: Scales

The workhorse of market research is the questionnaire, and one of the most important aspects of a questionnaire is

the use of scales. There is established methodology for devising and validating scales (cf. Celsi et al. 1992; Salzberger 2000; Rossiter 2002; Churchill 1984; O'Connor 2005; Gerbing 1988; Gerbing and Anderson 1988). We ask the question, how often is this established methodology used to devise scales? The design and construction of questionnaires is prefaced on face validity being established 'up front.' Face validity can be established by asking the population of interest to confirm the researchers' ideas of 'what looks as if' it might be correct and to make adjustments to the scales as necessary. Let us now assume that a 'validated' scale is pressed into service for a new population, situation or context. Very few clients are willing to pay for the required experimentation to re-validate the scale and very few suppliers are willing to admit that they are not sure if their proprietary instruments are valid in the new context. Thus, we inevitably have a drift in the certainty of the outcomes. Unless you test it within the population, situation and context you cannot be sure that it has external validity and 'really' works. Nomological validity implies repeated measures. For example, the SERVQUAL scale (Parasuraman, Zeithaml, and Berry 1985) is still being used consistently to evaluate 'service quality' in different contexts (cf. Donnelly et al. 2006; Badri, Abdulla, and Al-Madani 2005; Gounaris 2005) although issues with its discriminant validity were identified as early as 1990 (Carman).

Example 2: Projective Techniques

Projective techniques are used extensively in qualitative research. We have no issue with the use of stimulus to promote discussion and to help respondents express their feelings and emotions. However, projective techniques are often taken further; whereby images chosen by respondents are interpreted by the market researcher. This interpretation by the market researcher introduces the full range of validation issues. A detailed

discussion is beyond the scope of this paper, but let us just provide two simple cases. First, it is a commonly held belief that colour can be validly interpreted. For instance, red is often thought to mean aggression whereas green means tranquillity and furthermore that these attributes can be inferred when respondents associate these colours with products or services. The issue here is whether the interpretation is valid. But psychologists and psychiatrists who attribute meaning to psychological responses to colour cannot agree on which colour means which, and in a cross cultural context colour has entirely different meanings depending on your cultural background (Jacobs et al. 1991; Aslam 2006). The temptation for an innovative market researcher would be to take a validated methodology based on interpretation of colour and apply it in a different culture.

Similarly, after more than 50 years of usage, there is still debate about the interpretation of Thematic Apperception Test (TAT) pictures (Hibbard 2003; Ackerman et al. 2001; Lilienfeld, Wood, and Garb 2000; Cramer 1999). If the experts can't agree on what it all means, how could the 'face' who is being asked to review the pictures? We note that many research companies have proprietary projection 'kits' that are applied across populations, situations and contexts. If the experts can't agree on TAT after 50 years, how valid might some instances of proprietary projections be?

Example 3: Constructs

Some methods have developed entire mythologies that have grown around them without recent peer review and challenge. A case in point is Geert Hofstede's now famous cross cultural dimensions (cf. Hofstede and Bond 1988; Hofstede 1994; Hofstede 2001). Hofstede analysed a large (secondary) data base of employee values scores collected by IBM between 1967 and 1973. The data was from more than 70 countries, from which he first used the 40 largest only and afterwards extended the analysis to 50 countries and 3 regions. Since 2001, scores are listed for 74 countries and regions, partly based on replications and extensions of the IBM study on different international populations. As of mid-June 2006, the search term 'Hofstede' produced over 45,000 citations in scholar.google.com – while not the most authoritative source, a clear indication of the volume of researchers who are using Hofstede's concepts. One of the dimensions identified by Hofstede and Bond (1988) from an original study of Chinese Cultural Values is something that is now called 'long term orientation' (Merkin 2004). The constructs in this dimension are 'measured' by the items shown in Table 1.

You can see that most of the words used in these 'measures' have very little to do with: 1) time, or 2) long term outlook, or 3) orientation (taking a relative position). They may have more construct validity if compared with the origi-

Table 1: Measures of long term orientation

DIMENSION LABEL (CONSTRUCT)		
Long term orientation		Short term orientation
Scale Items	Persistence (perseverance)	Personal steadiness and stability
	Ordering of relationships by status	Protecting your face
	Thrift	Respect for tradition
	Having a sense of shame	Reciprocation of greetings, favours and gifts

nal Chinese Values Scale as put forward by Bond (but they don't). However, the use of these items to 'measure' whether or not an individual approaches their life from a long term, future orientated, perspective would appear to be inconsistent with dictionary definitions of the words. Notwithstanding the ambiguity of meaning, the scale continues to be used apace even after Fang (2003) made it very clear that these concepts were being fundamentally misused. Any basic face validity test would have identified the problem: that after borrowing the methodology from the original Chinese that it now makes little sense. Put another way, the original items may have been valid, but they have been applied uncritically throughout the world in new populations, situations and contexts. No doubt, the use of passive students or disinterested employees in their corridors resulted in uncritical 'evaluations' of these items.

Why is the problem of validity such an issue in market research?

We have provided three examples of how invalid methodology can infiltrate market research. This leads to the question of why validity needs to be an important issue in market research, more than when compared to other sciences or engineering.

Validity arises as a problem in market research because there is often a lack of linkage back to the 'real world.' In order to explain this, consider a comparison with the science of bridge building. In bridge building, methodology and design principles are constantly validated by virtue of bridges having to pass the test of not falling down. Think of an imaginative young engineer devising an imaginative new approach for calculating resonant frequencies in a bridge. A wrong calculation can result in vibrations being amplified in the bridge structure, which is undesirable. In the real world, the brutal check on the validity of this imagi-

native approach is whether or not the bridge vibrates. But in market research, we can propose a solution to a problem, have it implemented, observe some outcomes and then move on without having to live with the consequences. It is very difficult to determine whether that hoped-for increase in satisfaction and subsequent sales did not occur because the methodology findings were invalid, or whether uncontrolled factors in the marketplace intervened.

Another example, from everyday life, is the 'theorising' that occurs for the latest 'whodunit' in the news. When the details are sketchy, unfettered imagination amongst armchair criminologists is free to roam far and wide. However, as time progresses and more 'facts' are brought to bear there is a process of pruning the wild thicket of theories that flourished when the news first breaks. For those of us old enough: think back to the Azaria Chamberlain case and the wild theories that that event generated (Bryson 2005; Linnell 2004; Brown 2004). This sort of theorising about social events is not so different to the exploratory or discovery stage in science. Evidence accumulates and ideas have to fit the ever increasing body of evidence. Furthermore, in many sciences we have the luxury of not just accumulating new evidence, but of being able to go back to the laboratory and design specific experiments that can differentiate between competing theories (or hypotheses).

The point we wish to make is that whether it be bridge building, theorising on whodunits or conscious laboratory experiments, there is a process of continuously confronting the methodology with the evidence of the real world. But for market research, it is often the case of delivering results and moving on. This fact places a huge onus on the researcher to ensure that methodology is validated beforehand. It explains why there can be an accumulation of methodologies of doubtful usefulness.

These days we are fortunately not too often subjected to unfettered imagination in bridge building. And while armchair theorising about titillating social events probably does not do much harm, the same is not true in market research. We have an obligation to ensure that imaginative new methodologies and approaches are valid and reliable. We don't want to see imagination or innovation being used in a way that damages the status of our profession by having poorly validated methods and measures. At the very least, market research suppliers (and their clients) should ensure that they 'cover off' the validity of their research by firstly accepting that scales can be very reliable and yet not valid (as per the Hofstede example described above). Furthermore, such 'reliability' should not be used as a proxy for validity.

Secondly, we suggest that there are two overarching types of validity to be considered before commencing research. The first of these is 'formative validity' which we argue is a necessary (but not sufficient) first step in establishing the validity of the overall research. Formative validity consists of face validity and content validity. These types of validity must be con-

sidered in the formation of the project and come before any data is collected.

The second type of validity assessment proposed is that of 'prognostic validity.' Prognostic validity consists of those types of validity which are valuable to the assessment of the predictions and scenarios that will be created from the research outcomes. We argue that prognostic validity encompasses assessment of the predictive, descriptive and reflective measures of validity AFTER the data has been collected. It will incorporate an understanding of design, methods, data collection and sources of error. It may also include some tests of reliability where these tests contribute to an understanding of the prognostic value of the data.

Assessment of formative and prognostic validity will ensure that the market research conducted contributes the most economic and social value to the sponsoring organisation. In this sense, information supplied by research is building a bridge from a known past (research) to an unknown future (decision making based on research). The bridge you build should stand the test of time.

Appendix A : Typology of validity in marketing research

Validity comes before reliability – something should be valid before it can be considered reliable

Note: This list is meant to be illustrative not exhaustive

Term	Description	Authors
Validity	The measure measures what it purports to measure	(Nunnally 1967)
Formative validity		
	This is the term we propose - it encompasses the 'measures' of validity which come before the collection of data – therefore is an integral part of project design	<i>Brennan and Camm (this paper)</i>
Content validity	The extent to which an instrument is made up of a representative sample from the domain of interest. Often used synonymously with face validity.	(Sireci 1998) <i>Thorndike and Hagen 1971</i> <i>(Cronbach and Meehl 1955)</i> <i>(Diamantopoulos 2005)</i>
Face validity	Exists when the measure 'looks as if' is should indicate a particular variable. This is usually when others agree so it is also called consensus validity. This is often used synonymously with content validity. However, face validity does not imply expert acceptance of the measure's validity.	<i>(Turner 1979)</i> <i>(Heeler and Ray 1972)</i> <i>(Diamantopoulos 2005)</i>
Prognostic validity		
	This is the term we propose which encompasses assessment of the predictive, descriptive and reflective measures of validity AFTER the data has been collected. It will incorporate an understanding of design, methods, data collection and sources of error. It may also include some tests of reliability where these tests contribute to an understanding of the prognostic value of the data.	<i>Brennan and Camm (this paper)</i>
Construct validity	The extent to which a test may be said to measure the theoretical idea (construct).	<i>(Goode)</i> <i>(Churchill 1979)</i> <i>(Redmond and Griffith 2003)</i> <i>(Heeler and Ray 1972)</i> <i>(Peter 1981)</i>
Concurrent validity	The extent to which particular measures correlate with other criterion measures concurrently – that is there is concurrent variation in the outcomes	<i>(Heeler and Ray 1972)</i> <i>(Cronbach and Meehl 1955)</i>
Predictive validity	The extent to which particular measures predict other criterion measures That is you can use one measure to predict the outcomes of another. Predictive and concurrent validities are variations of Criterion validity	<i>(Heeler and Ray 1972)</i> <i>(Cronbach and Meehl 1955)</i>

Convergent validity	The extent to which the measure correlates or converges with similar measures of the same construct.	(Campbell and Fiske 1959) (Heeler and Ray 1972) (Diamantopoulos 2005) (Rossiter 2002) (Reichardt and Coleman 1995)
Discriminant validity	Antonymous with convergent validity it is the extent to which a measure differentiates between constructs. Also called divergent validity.	(Heeler and Ray 1972) (Rossiter 2002) (Diamantopoulos 2005) (Reichardt and Coleman 1995)
Criterion validity	The measure of how well one variable (or set of variables) predicts or correlates with an outcome based on information from other variables (Includes concurrent, predictive and diagnostic validity)	(Nunnally 1967) (Redmond and Griffith 2003) (Diamantopoulos 2005)
Nomological validity (law-like)	The degree to which predictions from a formal theoretical network containing the concept under consideration are confirmed. Thus, there are formal hypotheses derived from theory	(Venkatraman 1989) (Diamantopoulos 2005) (Peter 1981) (Cronbach & Meehl, 1955)
Diagnostic validity	Used mainly in the health sciences. Does the test outcome consistently relate to the diagnosis of the client/patient condition?	(Redmond and Griffith 2003)
Substantive validity	A subset of construct validity, it is the extent to which the scale items have a theoretical linkage to the construct.	(Schwab 1980) (Garver and Mentzer 1999)
Internal validity	When any differences between outcomes are attributed solely to the effect under investigation. The role of theory in this type of validity is to differentiate as something different from other constructs.	(Redmond and Griffith 2003) (Peter 1981)
External validity	The extent to which results can be generalised to other populations, situations or conditions.	(Redmond and Griffith 2003) (Peter 1981)

REFERENCES

- Ackerman, Steven J., Mark J. Hilsenroth, Amanda J. Clemence, Robin Weatherill, and J. Christopher Fowler. (2001). Convergent validity of Rorschach and TAT scales of object relations. *Journal of Personality Assessment* 77 (2):295-306.
- Aslam, Mubeen. (2006). Are you selling the right colour? A cross cultural review of colour as a marketing cue. *Journal of Marketing Communications* 12 (1):15-30.
- Badri, Masood A., Mohamed Abdulla, and Abdelwahab Al-Madani. (2005). Information technology center service quality: Assessment and application of SERVQUAL. *International Journal of Quality & Reliability Management* 22 (8):819-848.
- Barker, Andy. (2002). Nature, nurture and inertia: Where do qualitative researchers get their ideas from? *Qualitative Market Research* 5 (2):152-153.
- Basil, Michael D., William J. Brown, and Mihai C. Bocarnea. (2002). Differences in univariate values versus multivariate relationships: Findings from a study of Diana, Princess of Wales. *Human Communication Research* 28 (4):501-514.
- Blankenship, A. B. (1961). Creativity in consumer research. *Journal of Marketing* 25 (6):34-38.
- Brown, Jenny. (2004). Lindy rejects bizarre claims. *Woman's Day (Australia)*, 18.
- Bryson, John. (2005). Evil angles. *Bulletin with Newsweek*, 28.
- Campbell, D T, and Donald W. Fiske. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* 56:81-105.
- Carman, James M. (1990). Consumer perceptions of service quality: An assessment of the SERVQUAL dimensions. *Journal of Retailing* 66 (1 Spring):33-35.
- Celsi, Richard, Simeon Chow, Jerry Olson, and Beth Walker. (1992). The construct validity of intrinsic sources of personal relevance: An intra-individual source of felt involvement. *Journal of Business Research* 25 (2 September):165-185.
- Churchill, G A Jr, Peter, J P. (1984). Research design effects on the reliability of rating scales: a meta analysis. *Journal of Marketing Research* 21 (November):360-375.
- Churchill, Gilbert A Jr. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research* 16 (February):64-73.
- Clemons, Eric K., Hann Il-Horn, and Lorin M. Hitt. (2002). Price dispersion and differentiation in online travel: An empirical investigation. *Management Science* 48 (4):534-549.
- Cramer, Phebe. (1999). Future directions for the thematic apperception test. *Journal of Personality Assessment* 72 (1):74-92.
- Cronbach, Lee J, and Paul E Meehl. (1955). Construct validity in psychological tests. *Psychological Bulletin* (52):281-302.
- Deshpandé, Rohit. (1983). "Paradigms Lost": On theory and method in research in marketing. *Journal of Marketing* 47 (Fall):101-110.
- Diamantopoulos, Adamantios. (2005). The C-OAR-SE procedure for scale development in marketing: A comment. *International Journal of Research in Marketing* 22:1-9.
- Donnelly, Mike, Neil J. Kerr, Russell Rimmer, and Edward M. Shiu. (2006). Assessing the quality of police services using SERVQUAL. *Policing: An International Journal of Police Strategies & Management* 29 (1):92-105.
- Fang, Tony. (2003). A critique of Hofstede's fifth national culture dimension. *International Journal of Cross Cultural Management: CCM* 3 (3):347-368.
- Garver, Michael S., and John T. Mentzer. (1999). Logistics research methods: Employing structural equation modeling to test for construct validity. *Journal of Business Logistics* 20 (1):33-57.
- Gerbing, David W, and James C Anderson. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research* 25:361-370.
- Goode, Sigi. (2006). *On the theoretical justification of the organisational size construct in information systems research*. Australian National University [accessed 24 Feb 2006].

- Gounaris, Spiros. (2005). Measuring service quality in b2b services: An evaluation of the SERVQUAL scale vis-à-vis the INDSERV. *Journal of Services Marketing* 19 (6):421-435.
- Heeler, Roger M, and Michael L Ray. (1972). Measure validation in marketing. *Journal of Marketing Research* 9 (November):361-370.
- Hibbard, Stephen. (2003). A critique of Lilienfeld et al.'s (2000) "The Scientific Status of Projective Techniques". *Journal of Personality Assessment* 80 (3):260-271.
- Hill, Nigel. (1990). Commodity products and stalemated industries: Is there a role for marketing? *Journal of Marketing Management* 5 (3):259-281.
- Hofstede, G H. (2001). *Culture's consequences*. 2nd ed. Thousand Oaks, CA: Sage.
- Hofstede, Geert. (1994). Business cultures. *The Unesco Courier* Apr (4):12.
- Hofstede, Geert, and Michael Bond. (1988). The Confucius connection: From cultural roots to economic growth. *Organizational Dynamics* 16 (4):4-21.
- Hult, G. Tomas M., Jr David J. Ketchen, and Stanley F. Slater. (2005). Market orientation and performance: An integration of disparate approaches. *Strategic Management Journal* 26 (12):1173-1181.
- Jacobs, Laurence, Charles Keown, Reginald Worthley, and Kyung-Il Ghymn. (1991). Cross-cultural colour comparisons: Global marketers beware! *International Marketing Review* 8 (3):21-30.
- Lilienfeld, Scott O., James M. Wood, and Howard N. Garb. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest* 1 (2):27-66.
- Linnell, Garry. (2004). A sorry affair. *Bulletin with Newsweek* 122 (6430):6.
- McAnena, Fiona. (2004). Taking insight from dull data. *Marketing (UK)*:33-.
- Merkin, Rebecca S. (2004). Cultural long-term orientation and facework strategies. *Atlantic Journal of Communication* 12 (3):163-176.
- Murray, Jeff B., Deborah J. Evers, and Swinder Janda. (1995). Marketing, theory borrowing, and critical reflection. *Journal of Macromarketing* 15 (2):92-106.
- Nunnally, Jim C. (1967). *Psychometric theory*. New York: McGraw Hill.
- O'Connor, Thomas. R. (2005). Types of scales and indexes. *Encyclopedia of Social Measurement* 3.
- Parasuraman, A., Valarie A. Zeithaml, and Leonard L. Berry. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing* 49 (4 Fall):41-45.
- Peter, J Paul. (1981). Construct validity. *Journal of Marketing Research* 18 (May):133-145.
- Redmond, Elizabeth C, and Christopher J Griffith. (2003). A comparison and evaluation of research methods used in consumer food safety studies. *International Journal of Consumer Studies* 27 (1 January):17-33.
- Reichardt, Charles S, and S C Coleman. (1995). The criteria for convergent and discriminant validity in a multitrait-multimethod matrix. *Multivariate Behavioral Research* 30 (4):513-538.
- Rossiter, John R. (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing* 19 (4):305-336.
- Salzberger, Thomas. (2000). An alternative way of establishing measurement in marketing research- its implications for scale development and validity. Paper presented at ANZMAC 2000, at Griffith University.
- Schwab, Donald P. (1980). Construct validity in organizational behavior. In *Research in organizational behavior*, edited by B. Staw and L. L. Cummings. Greenwich, Connecticut: JAI Press.
- Sireci, Stephen G. (1998). The construct of content validity. *Social Indicators Research* 45 (1 - 3):83-117.
- Temporal, Paul. (2006). *Asia's star brands*. Singapore: John Wiley and Sons.
- Turner, Stephen P. (1979). The concept of face validity. *Quality and Quantity* 13 (1):85-90.
- Vanderkaay, Sharon. (2005). The business of selling knowledge: Part I. *Canadian Architect* 50 (2):27.
- Venkatraman, N. (1989). Strategic orientation of business enterprises: The construct, dimensionality, and measurement. *Management Science* 35 (8):942-962.
- Wolfe, A. R. (1977). Commodity marketing and generic promotion. *European Journal of Marketing* 11 (7):532-547.

Interested in becoming an AMSRS member?

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 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), vote in divisional and national elections of the AMSRS and receive the Society's Journal AJMSR.

Note: Student members receive limited membership benefits so please see 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.

JoinToday

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

Handbook of Ethnography

By Paul Atkinson, Amanda Coffey, Sara Delamont, John Lofland, and Lyn Lofland (eds)(2002) (London: Sage), 507 pages.

Reviewed by Michael Milgate

Ethnography is large. To see the multitudes it contains, start with its Oxford English definition: "Ethnography is the scientific description of nations or races of men, their customs, habits, and differences." This has the advantage of being clear and straightforward, but if we take a group of 46 highly respected ethnographers: the contributors to Sage's well-conceived *Handbook of Ethnography*, for example, we find there is not a single word in that definition that all agree should be there. What is more, it excludes the two defining elements of their craft that these 46 come closest to agreeing on: Ethnography is rooted in the first-hand experience of the research setting, and ethnography is committed to interpreting the point of view of those under study. Even these last two statements, though widely accepted, are contentious oversimplifications requiring careful qualification with weasel words such as rooted and committed to allow exception and ambiguity their rightful place. Ethnographers believe in the virtues of concrete description but vague definition. There is reason for this. As the editors of the *Handbook*: Paul Atkinson, Amanda Coffey, Sara Delamont, John Lofland, and Lyn Lofland, argue, "While theoretical fashions can come and go, the products of ethnographic research remain extraordinarily durable." Ethnographic data do not exist to substantiate or disprove theories. The data, the descriptions and interpretations, are not the means to something else: They are the end. The definitions and theories and paradigms are the means: They exist to help us understand the data. "The enduring value of the ethnographic tradition is grounded in its attention to the singular and the concrete," despite the contention within that tradition about general principles.

It should be no surprise, then, that ethnography persists and continues to grow in influence-not only in anthropology, where it was born, but in sociology and in applied disciplines such as education, medicine, and, of course, management-despite an almost complete lack of agreement among ethnographers about seemingly fundamental and definitional issues. Consider the debates within ethnography that its dictionary definition exposes and that the *Handbook* usefully illuminates. Again, "Ethnography is the scientific description of nations or races of men, their customs, habits, and differences." What is wrong with this? First, there is no agreement among ethnographers that ethnography is "the" scientific description of anything. Any particular ethnography is only "a" scientific description (if, indeed, it is that) and one not necessarily better than others that are always possible. The standards of quality for evaluating ethnographies are contested. Furthermore, there is no agreement that ethnography is, or should be, "scientific." Some argue that cultural anthropology would be better placed among the humanities than the social sciences. Others disagree. Some would regard the accounts of participant observing journalists such as Hunter S. Thompson (1966), Tracy Kidder (1981), or David Lipsky (2003) as authentically ethnographic. Others would preserve a distinction between ethnography and reportage. Still others, however, would deny the distinction between ethnography and fiction. The intellectual location of ethnography is contested.

On we go. There is no agreement about what it means to claim that ethnography is about "description." Is ethnography a method of describing or is it the product of description? In other words, is ethnography *Writing Culture* (Clifford and Marcus, 1986) or the *Tales of the Field* (Van Maanen, 1988) themselves? If we can see that it may be both, then

we must recognise that it can also be neither: Ethnography sometimes signifies neither product nor process but simply presumption or prestige, as when it is borrowed by the practitioners of other forms of qualitative research for the (supposed) aura of rigor and legitimacy it affords. The use of the label ethnography is contested. That aside, there is, anyway, no agreement among those who call themselves ethnographers that ethnography should primarily be about description rather than, say, political action. The goals of ethnography are contested as well.

There is no agreement that the important boundaries when doing ethnography are those “of nations or races.” The usefulness of bounding culture in national or racial terms is hotly debated. Organisations, work groups, communities, ethnic groups, regions, religious groups, classes, and professional groups all form the legitimate subject of ethnography, but there is no consensus about how best to bound a particular group under study. The boundaries in ethnography are contested. There is certainly no agreement that ethnography is the study of men or that terms and concepts can ever be neutral with regard to gender. Can we escape our gendered and racialised perspectives to come to understand others different from us? Can we, indeed, ever overcome the idiosyncrasies of our biographies sufficiently to understand anyone, even ourselves? If we can never do so completely, how should we proceed? If we can never do so at all, what should we do instead? There is no consensus among ethnographers. The very possibility of ethnography is contested. Finally, there is no agreement that the ethnography of a group of people should be about “their customs, habits, and differences.” Ethnography is concerned with culture, but what is culture? If there are just about as many definitions of ethnography as there are ethnographers, there are more definitions of culture still. Is culture about customs and habits or is it about artifacts or practices or ideas? Is it about elements or is it about patterns and propensities? Is it about what people share or can it include how they differ? Is it social? Is it divisible? There is no consensus. The subject matter of ethnography is contested.

The immense diversity of perspectives that ethnographers take toward their craft creates the need for a *Handbook of Ethnography* to help students, and the rest of us make sense of the differences within ethnographic practice and between ethnography and adjoining disciplines such as ethnomethodology, phenomenology, and grounded theory. At the same time as it creates need, though, it also creates problems for those who would meet that need. Difficult choices face the editors about what to include and what not to include, how to clarify but not oversimplify. Their task is more difficult even than this, however, because ethnography is large in other ways. Ethnographers tend to identify more with their method and their perspective on it than with the substantive areas they are studying. The area of organisational ethnography, for example, has been influenced by ostensibly medical ethnographies such as those by Becker, Greer, et al. (1961), Goffman (1961), and Bosk (1992); by ethnographies of deviance such as those by Whyte (1955) and Adler (1985); by ethnographies of education such as that by Willis (1977); by ethnographies in the science and technology studies tradition such as those by Latour and Woolgar (1979) and Rabinow (1996). Meanwhile, ethnographies of work and organisation such as those by Dalton (1959), Van Maanen (1975), Burawoy (1979), Hochschild (1983), Barley (1986), Smith (1990), Kunda (1992), and Vaughn (1996) have had wide influence outside of the disciplines of management and business.

In addition to reviewing, and exhibiting, the theoretical and methodological debates in the field and also the major substantive areas of ethnographic inquiry, the Handbook further provides a number of excellent introductory chapters on the various origins of ethnography in social anthropology (largely British), cultural anthropology (largely

American), and sociology (largely Chicagoan). All in all, the editors, facing an impossible task and wisely rejecting the ambition of being either definitive or comprehensive, have succeeded in assembling a volume of quality and breadth that deserves a place on the bookshelf of graduate students interested in ethnography as well as degreed ethnographers who want a better understanding of the history and scope of ethnography and of the current and perennial debates surrounding it.

The book is divided into three parts. The first part contains chapters about where ethnography came from and where it did not go: its origins and its differences from related disciplines. The chapters by Deegan on the Chicago school, by Rock on symbolic interactionism, by Faubion on cultural anthropology, and by Macdonald on social anthropology deserve particular praise for providing a primer of these fields, and of the classic ethnographies that serve as their milestones, that should be required reading for any doctoral student of organisational ethnography. Also worthy of special mention are the chapters by Pollner and Emerson about what ethnographers might learn from ethnomethodology and by Charmaz and Mitchell about what the differences are between ethnography and grounded theory.

The second part contains chapters about various key substantive domains of ethnographic exploration. Of particular interest is Vicki Smith's excellent review of ethnographies of work, but ethnographers of work may learn more from the chapters by Bloor on the ethnography of health and medicine; by Gordon, Holland, and Lahelma on the ethnography of educational settings; by Hobbs on the ethnography of deviance; and by Hess on the ethnography of science and technology. If the received wisdom is true that "the only way to learn how to do ethnography is to read the ethnographies," then the value of these chapters is in the advice they give about where to read, the maps they give of the literature. They form the heart of the Handbook.

The third and final part of the Handbook contains chapters about different methods, perspectives, realities, and fantasies of ethnography. There are several extremely good chapters in this section. The chapter by Wellin and Fine about ethnography as work stands out in particular. Picking up a theme also discussed by Smith in her chapter, they discuss just how difficult the craft of ethnography becomes when it is set in the bureaucracy of academia. They point out how many of the most influential ethnographies in the field of work and organisations are based on dissertation fieldwork and how few of us manage to achieve a second major ethnographic study in our career. Spencer's chapter on ethnography after postmodernism is also exceptional for managing to explain in the clear, plain language worthy of an ethnographer what has been the lasting influence of the crisis of representation, the literary turn, the postmodern, post-structuralist, experimental moment on ethnographic practice and writing. Also deserving of praise are the chapters by Emerson, Fretz, and Shaw on participant observation and field notes, by Heyl on interviewing, and by Fielding giving an introduction to computer-aided qualitative data analysis. Overall, this third part of the volume gives a broad overview of the key issues and debates along with pointers to further reading.

It is in this third part of the *Handbook of Ethnography*, though, where the overlap between it and what should be seen as a companion volume, Sage's *The Handbook of Qualitative Research*, edited by Norman Denzin and Yvonna Lincoln, is most evident. To have these two enormous volumes (the *Handbook of Ethnography* has more than 500 pages; the third edition of the *Handbook of Qualitative Research* has more than 1,600, and both books are set in two columns of small type) is a testament to the

growing influence of ethnography and other forms of qualitative research. Although ethnographic practice features heavily in the *Handbook of Qualitative Research*, the two handbooks are very different in their approach and in their goals. Denzin and Lincoln are explicit that their agenda for their handbook is “to show how the discourses of qualitative research can be used to help create and imagine a free democratic society,” and each of the chapters takes up that project in some way. Denzin and Lincoln propose a model of the seven moments of the history of qualitative research, and the contributions to the *Handbook of Qualitative Research* are very much focused on the later moments: the crisis of representation, the postmodern period of experimental writing, the postexperimental moment, and the future. The focus is on new and different paradigms, theories, strategies, and practices of inquiry. In contrast, the editors of the *Handbook of Ethnography* explicitly reject in their introduction Denzin and Lincoln’s model of moments and emphasise the importance of classic ethnographic design that they claim endures while fashion waxes and wanes. Thus, their focus on origins and on those ethnographic texts that have had enduring influence. As a result, the *Handbook of Qualitative Research* can be read (by the ambitious or extremely exam-anxious doctoral student, perhaps) as a massively expanded third part of the *Handbook of Ethnography*. Taken together, even better than when considered separately, they reveal the tensions and contradictions of ethnography but also the power and the fascination it holds.

By itself, the *Handbook of Ethnography* covers a broad terrain. One consequence of this is that not every one of its chapters is directly relevant to organisational scholars. Yet even those that seem at first glance to be parochially anthropological—such as the chapter on Orientalism by Marcus, may be metaphorical food for thought. How might organisational ethnographies be compromised by the interests of business schools to produce useful managerial knowledge in the same way that anthropological ethnographies were compromised by colonial interests? The more that we organisational ethnographers know of the historical and current debates in the broader ethnographic world, the less we have to reargue them ourselves from scratch. That alone makes the appearance of the *Handbook of Ethnography* welcome. It would benefit from a much better index. The index to a book of this size and type is critical and sorely lacking in completeness here, poor Gideon Kunda, to give one example, is mentioned at least three times in the book but does not appear at all in the index. A consolidated bibliography would also enhance the usefulness of the volume as a reference. Some chapters would have benefited from stronger editorial consistency in forcing the authors to focus on ethnographies themselves and not on theoretical fashions. Overall, however, this is an immensely useful book that I recommend strongly to students of organisational ethnography, in whatever stage of career.

REFERENCES

- Adler, P. (1985). *Wheeling and dealing: An ethnography of an upper-level drug dealing and smuggling community*. New York: Columbia University Press.
- Barley, S. R. (1986). Technology as an occasion for structuring: Evidence from observation of CT scanners and the social order of radiology departments. *Administrative Science Quarterly*, 31, 78-101.
- Becker, H., Greer, B., et al. (1961). *Boys in white: Student culture in medical school*. Chicago: University of Chicago Press.
- Bosk, C. L. (1992). *All God's mistakes: Genetic counseling in a pediatric hospital*. Chicago, University of Chicago Press.
- Burawoy, M. (1979). *Manufacturing consent: Changes in the labor process under monopoly capitalism*. Chicago: University of Chicago Press.
- Clifford, J., and Marcus, G. E. (1986). *Writing culture: The poetics and politics of ethnography*. Berkeley: University of California Press.
- Dalton, M. (1959). *Men who manage: Fusions of feeling and theory in administration*. New York: John Wiley.
- Denzin, N. K., and Lincoln, Y. S. (2005). *The handbook of qualitative research* (3rd ed), Thousand Oaks: Sage.
- Goffman, E. (1961). *Asylums: Essays on the social situation of mental patients and other inmates*. New York: Doubleday.
- Hochschild, A. R. (1983). *The managed heart*. Berkeley: University of California Press.
- Kidder, T. (1981). *The soul of a new machine*. New York: Avon.
- Kunda, G. (1992). *Engineering culture: Control and commitment in a high-tech corporation*. Philadelphia: Temple University Press.
- Latour, B., and Woolgar, S. (1979). *Laboratory life: The social construction of scientific facts*. Princeton: Princeton University Press.
- Lipsky, D. (2003). *Absolutely American: Four years at West Point*. Boston: Houghton Mifflin.
- Rabinow, P. (1996). *Making PCR*. Chicago: University of Chicago Press.
- Smith, V. (1990). *Managing in the corporate interest: Control and resistance in an American bank*. Berkeley: University of California Press.
- Thompson, H. S. (1966). *Hell's Angels*. Harmondsworth, UK: Penguin.
- Van Maanen, J. (1975). Police socialization: A longitudinal examination of job attitudes in an urban police department. *Administrative Science Quarterly*, 20, 207-228.
- Van Maanen, J. (1988). *Tales of the field: On writing ethnography*. Chicago: University of Chicago Press.
- Vaughn, D. (1996). *The Challenger launch decision: Risky technology, culture, and deviance at NASA*. Chicago: University of Chicago Press.
- Whyte, W. F. (1955). *Street corner society: The social structure of an Italian slum*. Chicago: University of Chicago Press.
- Willis, P. (1977). *Learning to labor: How working class kids get working class jobs*. New York: Columbia University Press.

