



**asc**  
association for survey computing  
southampton 2007

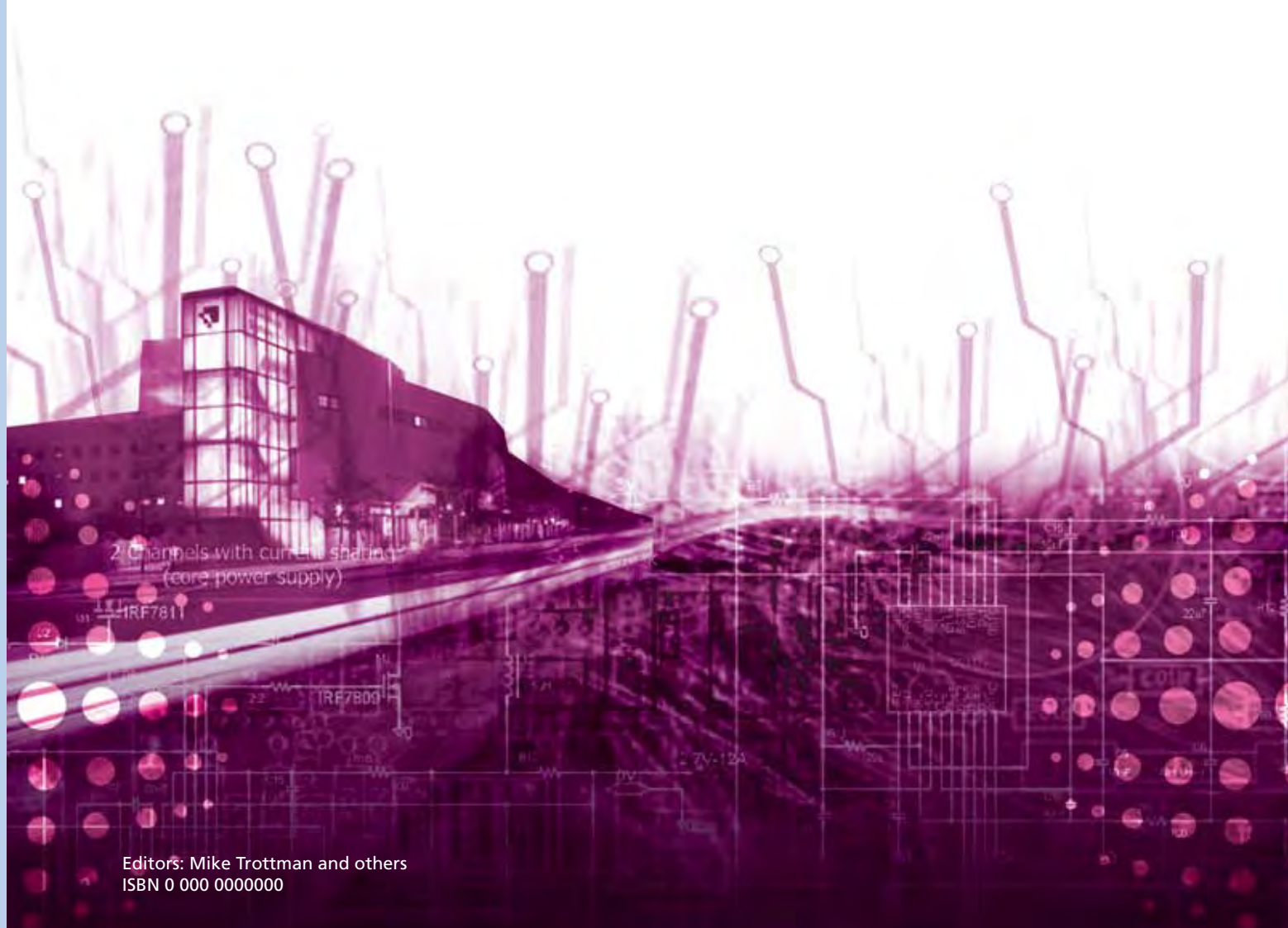
## The Challenges of a Changing World

---

### Proceedings of the Fifth International Conference of the Association for Survey Computing

---

University of Southampton, 12-14 September 2007



2 Channels with current sharing  
(core power supply)

U1 IRF7811

IRF7809

Editors: Mike Trottman and others  
ISBN 0 000 000000



**asc**  
association for survey computing  
southampton 2007

# **The Challenges of a Changing World**

## Proceedings of the Fifth ASC International Conference

University of Southampton, UK  
September 12-14, 2007

Edited by

Mike Trotman	<i>(Datalucid Limited)</i>
Tim Burrell	<i>(Office for National Statistics)</i>
Laurance Gerrard	<i>(LG Consulting)</i>
Ken Anderton	<i>(Opinion Research)</i>
Garj Basi	<i>(Kantar Operations)</i>
Mick Couper	<i>(University of Michigan)</i>
Keith Morris	<i>(Kantar Operations)</i>
David Birks	<i>(University of Southampton)</i>
AJ Johnson	<i>(Ipsos-Mori)</i>
Reg Baker	<i>(Market Strategies)</i>
Malcolm Rigg	<i>(PSI)</i>
Steve Taylor	<i>(Inputech)</i>
Andrew Westlake	<i>(Survey &amp; Statistical Computing)</i>

**Association for Survey Computing**

ASSOCIATION FOR SURVEY COMPUTING

PO Box 76  
Berkeley  
GL13 9WU  
UK

email: [admin@asc.org.uk](mailto:admin@asc.org.uk)  
UK tel +44 (0) 1453 511 511  
UK fax +44 (0) 1453 511 512  
<http://www.asc.org.uk>

ISBN: 0 9546748 2 0

Compilation © 2007 Association for Survey Computing

British Library Cataloguing in Publication Data.

A catalogue record for this book is available from the British Library.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form by means of electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of the publisher, the Association for Survey Computing, or the author of the relevant article.

## Preface

Welcome to ASC 2007, Developments in the Survey Process, the Association for Survey Computing's fifth major international conference on survey and statistical computing. As with others in the series, the aim of this one was to provide as broad a view as possible of the salient issues and challenges confronting the field.

Papers by two of the four speakers invited to address plenary sessions of the conference comprise the first section of this volume.

The remainder and majority of this volume consist of contributed papers addressing one of the four major themes that emerged when the Scientific Programme Committee (SPC) grouped together the proposals that were accepted from amongst the many excellent offers we originally received. In the order in which they appear in this volume, they are:

- The Interview
  - Excludes the 4 papers on 'New Interview Modes' for copyright reasons
- The Respondent
- The Data
- The Software

In some respects, these themes serve as headings for convenience only, as many papers could easily have fit under more than one heading. We expect the reader to find much pleasure in noting the many points of contact and convergence between the papers regardless of the immediate company they are seen to be keeping.

The contributed papers include both case studies and more theoretical reflections; some represent the outcome of purely methodological investigations, while others have a substantive focus. Regardless of content or technique, however, we are sure that the reader will find the papers in this volume instructive, thought-provoking and, in many cases, directly and practically applicable to his or her own circumstances.

On behalf of the ASC we would like to thank all the authors for preparing what we feel is an excellent volume of papers and for putting up with our many urgent reminders about the deadlines we did not want them to overrun. As co-chairs of the SPC, we would also like to thank all the members, in particular those who convened sessions and helped edit this volume.

Mike Trotman and Tim Burrell, August 2007.

Members of the ASC 2007 Scientific Programme Committee

**Mike Trotman (Co-Chair)**

*Datalucid Limited*

**Tim Burrell (Co-Chair)**

*Office for National Statistics*

**Laurance Gerrard**

*LG Consulting*

**Ken Anderton**

*Opinion Research*

**Garj Basi**

*Kantar Operations*

**Mick Couper**

*University of Michigan*

**Keith Morris**

*Kantar Operations*

**David Birks**

*University of Southampton*

**AJ Johnson**

*Ipsos-Mori*

**Reg Baker**

*Market Strategies*

**Malcolm Rigg**

*PSI*

**Steve Taylor**

*Inputech*

**Andrew Westlake**

*Survey & Statistical Computing*

## About the ASC

The Association for Survey Computing (ASC) is a not-for-profit organisation, based in the UK but drawing significant international membership and support. It provides a valuable forum for all those concerned with survey research and statistical computing and has a wide-ranging membership at both individual and corporate level. Originally formed in 1971 as the Study Group on Computers in Survey Analysis, it became the Association for Survey Computing in 1994. The ASC has links with the British Computer Society, the International Association for Statistical Computing, the Market Research Society, The Royal Statistical Society and the Social Research Association.

The ASC's many activities and links to other organisations are publicised through its website (<http://www.asc.org.uk>) and include the many conferences and workshops it runs, the register of available software it maintains and its support of other initiatives in the survey and statistical field.

For many years the ASC has organised one-day conferences in London in January and June together with occasional one-day specialised workshops. In 1992, to celebrate its 'coming of age', the ASC organised its first International 3-Day Conference, held in Bristol, England. Since then it has held similar international events at Imperial College in London, England (1996), at Edinburgh University, in Scotland (1999), at Warwick University (2003), and now in 2007 holds its fourth International Conference at the University of Southampton.

The ASC was formed in order to improve knowledge of good practice in survey computing and to disseminate information on techniques and survey software. From the outset, a particular emphasis has been to help avoid wasteful duplication of computer programs where good packages already exist and to encourage the development of common standards. In its early days the focus of involvement was on collection and analysis of survey data. Now, after more than 30 years, the Association's role has extended much more widely from the involvement of computers in survey design and formulation at one end of the process to presentation and dissemination of survey data and results at the other. Over this period it has built up a unique role as an interchange point between a variety of survey and statistical specialisms and between client and contractor organisations.

The ASC maintains a Register of Software for Statistical and Survey Analysis, listing details of all relevant packages known to the Association. This is available on the ASC website. It also publishes, for its membership, a periodic newsletter.

In co-operation with the Market Research Society is the annual 'MRS/ASC Award for Technology Effectiveness', intended to foster innovation in the involvement of computers in the survey process, whether on a large or small scale, by both companies and individual professionals. This award was presented for the first time in the autumn of 2003 and the short-list of finalists this year will be announced at the ASC2007 Conference Dinner.

While principally dedicated to serving its individual and corporate membership, non-members are welcome at all events. Membership is inexpensive and brings a number of benefits. More details are available from the ASC's Administrator, ASC, PO Box 76, Berkeley, GL13 9WU UK tel +44 (0) 1453 511 511, UK fax +44 (0) 1453 511 512, email: [admin@asc.org.uk](mailto:admin@asc.org.uk).

The activities of the ASC are organised by a volunteer committee - elected annually - supported by an Administrator. The 2006/2007 ASC Committee was:

*President:* Beverley Charles Rowe

*Chair:* Laurance Gerrard

*Vice-Chair:* Steve Taylor

*Treasurer:* Raz Kahn

*Secretary:* Andrew Westlake

*Administrator:* Christine Jenkins

*Committee:* Randy Banks,  
Garj Basi,  
Tim Burrell,  
AJ Johnson,  
Tim Macer,  
Ed Ross,  
Mike Trotman,  
Peter Wills

*Affiliates:* Suzanne Evans (*RSS*), Wendy Sykes (*SRA*), Sue Trenhaile (*MRS*)

## Table of Contents

### Keynotes

Malcolm Rigg	Affluence, Sustainability and Computing	1
Mick Couper	Whither the Web: Web 2.0 and the Changing World of Web Surveys	7

### The Interview

#### Interview Context

Ineke Stoop	Survey data, context and event data	17
Cornelia Zuell	Reporting Societal Events to Facilitate the Interpretation of Survey Results	29

#### Interview Modes

Mark Cameron	Survey Research in a Wireless World	39
AJ Johnson	Exploring the concept of Web 2.0 in the future of online research	45
Hugh Davis	Real-Time Sampling	55

#### Web Interview Enhancements

Marek Fuchs	Multimedia Web Surveys: results from a field experiment on the use of audio and video clips in Web surveys	63
Nicola Stanley	Watch what I do: Using graphical input controls in web surveys	81

### The Respondent

#### Response Rates

John Cooper	Factors Affecting Return Rates to On-Line Surveys	93
John S. Lemon	Comparing the Effect of Reminder Intervals on Response Rates for Web Surveys	103
Jelke Bethlehem	Online Panels - A Theft of Paradigm?	113



## Respondent Privacy

Adam N. Joinson	Privacy, trust and self-disclosure to web-based surveys	133
Eve-Marie Larsen	Using chat tools to perform evaluation interviews	145

## Respondent Types

Reg Baker	Separating the Wheat from the Chaff: Ensuring Data Quality in Internet Panel. Samples	157
Ineke Stoop	Increased fieldwork efforts, enhanced response rates, inefficient samples	167
Joost Kappelhof	Problems with research among ethnic minorities in the Netherlands	179

## Response rates and Interviewing Mode

Heather Wardle	Choosing web surveys: mode choices among Youth Cohort Study respondents	193
Paul Smith	What's the time? Relations between interview periods and output periods in surveys	203

## The Data

### Meta Data

Graham Hughes	The difficulty of understanding social survey questionnaires from the published documentation.	213
Andrew Westlake	Whither Statistical Meta-data?	223
Geoff Wright	Triple-S - The Broader Horizon	241

### Other Techniques

Joe Whittaker	Weighted independence graphs for finite population surveys.	247
Derek Bond	Going beyond the fence: Using projective techniques as survey tools to meet the challenges of bounded rationality	259

### Qualitative Data and Analysis

Louise Corti	Qualitative Data Exchange: Methods and Tools	271
William O'Connor	Framework and its role in increasing quality and transparency in the analysis of qualitative data, a fresh solution	275
David Birks	XSight and the shaping of Marketing Analytics	285

## Other Perspectives

Nick Read	Is Quanvert here to stay?	297
Hugh Neffendorf	Surveys and Technology - Polishing the Crystal Ball	309

## The Software

### Management Systems

Adi Cohen	Case study of a large organization's transformation of its web interview capabilities.	317
Tim Brandwood	Utopia: A Complete Research Management System	333

### Software Tools

Andrew Smith	The Challenge of Geocoding Large Scale Travel Surveys	345
Fida Hussain	Use of Seasonal Adjustment Software within the Office for National Statistics	353
Lauren Courtney	New Technologies and Tools for Study Management: Designing, Implementing and Maintaining a Web-Based Data Management System for a Multi-Site Longitudinal Intervention Study	365

### Other Software Issues

Steve Jenkins	Knitting Patterns: for interview and analysis	381
Paul Sampson & Peter Wills	The Software Licence Agreement: Friend or Foe?.	391



# Affluence, Sustainability and Computing

Malcolm Rigg

## Abstract

Western societies have grown three-fold in real income terms since WWII creating hitherto unknown levels of affluence. The consequences for individuals, society and the planet are profound. This paper explores the relationships between affluence and sustainability and raises important challenges for the IT and computing sector.

## Keywords

affluence, sustainability, computing, sustainable development, IT professionals

## 1. Introduction

In the last few months I have been heavily engaged in researching sustainable development particularly in relationship to the higher education sector. Before I joined PSI three years ago, I thought very little about the environment, and so I would like to share with you some thoughts about our consumer society, about sustainable development, then conclude with what I believe are the responsibilities of computer professionals in their different guises of developing and using computers, including researchers of course.

## 2. Affluence

I want to talk first about affluence. Most of us at this conference are affluent. Our incomes and earning power are far, far greater than the vast majority of the world's population. Indeed, incomes in Western Europe and the USA have grown threefold in real terms since 1950.

Way back in the 1960s Vance Packard (1961) was arguing that we are a society of waste makers. Since then our capacity to waste resources has risen exponentially. We have more material possessions than ever before and we replace or upgrade them more often than ever. How long did a telephone last in 1960? How many phones do we have now and how often do we replace them? How many music players, televisions, computers, have we got through in the last ten years? If you're anything like me, the answer is lots.

But does it feel as if we're so rich? Have we become more content? Have we stopped yearning for more? What has plenty done for us?

Avner Offer, in a recent book called *The Challenge of Affluence* (2006), argues that affluence is driven by novelty, and that novelty, although beguiling, unsettles people. Markets have delivered an abundance of novel and compelling opportunities. But affluence and novelty cause harm as well as good. Novelty displaces and devalues our existing stock of possessions. We want the latest model which appears to offer better performance. But it displaces and devalues not only possessions. It also, he argues, devalues virtues, relationships and values.

The 1950s and 1960s were a golden age of opportunities, and upward mobility. But the 1980s and 1990s, experienced a time when the benefits did not reach far down the social system. In this period, communities began to unravel. Social institutions began to be held in low esteem. We began to lose confidence in our politicians who promised us the benefits of the market by de-regulating industries thereby undermining social conventions such as restricted licensing hours, allowing out-of-town shopping at the expense of town centres, and offering us choice in energy, travel and public services.

At the same time, the flow of new rewards began to undermine our capacity to enjoy them. New rewards demand time and attention. Attention can be thought of as a universal currency of well-being. We can consume it by focusing on activities, or we can invest it for activity which offers greater satisfaction in the long run. We can invest in our futures by studying to enrich our lives or we can go out and enjoy ourselves. We can save for our pensions or buy the latest electronic gadgets. Our enormous and increasing levels of personal debt testify to this problem.

Personal well-being requires a sustainable balance between the present and the future. It requires a personal capacity for commitment which Offer calls prudence. The opposite of prudence is imprudence - making self-defeating choices. Affluence is a challenge, because choice is fallible. Every choice we make affects the future. It is difficult for us to bring objectives and desires at different ranges of time into a sustainable balance. Given abundance, we make many wrong choices that have implications for individuals, for society and for the planet.

### 3. Impact on the individual

The increasing body of research on well-being suggests that, above a certain point, it does not increase with wealth. There are rapidly diminishing returns. Our expectations rise faster than we can satisfy them. We seem to be on a treadmill. We spend more and more time maintaining and expanding our current inventory of possessions, always promising us more than they deliver – the new camera creates bigger picture files requiring greater storage and our effort is spent managing this rather than taking better pictures. Only too often, novelty diverts our attention to untested new rewards and often leads to unfulfilled expectations.

### 4. Impact on society

If this was the only effect then it might not matter too much. After all, personal choice is important. But we're not simply individuals. We live in societies in which inter-dependence and interaction are essential. We require systems of governance that enable us to live together but unfortunately novelty and poor individual choice don't just undermine well-being. They also undermine our social institutions.

Novelty replaces working arrangements that we know and trust with untried ones. This flow of novelty undermines existing commitment conventions. These are the values, rules and customs that hold us in

check. According to Offer, the challenge of affluence is coping with novelty. He argues that we need commitment mechanisms or ‘technologies’ in order to balance current and future needs. By that he means institutionalised arrangements that restrict our lack of ability to make good choices. The dangers Offer speaks of go well beyond consumption. They concern social and civic relationships and commitments. In a throw-away society it is all too easy to reduce issues to questions of personal choice rather than concerns about values and how we live our lives as members of interdependent communities. We increasingly need to understand how our institutional and social frameworks can support us in restraining actions which are not in our long term interests. The market and unfettered choice have created unheard of affluence but at the same time have undermined our ability to live together without damaging our institutions and social fabric.

## 5. Impact on the planet

It’s not just ourselves that get damaged. It’s the planet as well. Up to very recent times it has been taken for granted that we have the right to exploit natural resources as long as we more or less clear up the environmental and social mess we make in the process. Not so any more. The rate of exploitation of natural resources, emissions of greenhouse gases and the problems of waste threaten to overwhelm the planet. Energy reserves are rapidly depleting, climate change is happening already and we are wondering where to put all our rubbish. Just to take one example: where do all the old computers go? There’s at least 3 million in landfill in the UK alone!

## 6. So what can be done?

The first question we might ask is whether we can achieve a sustainable society? The jury is out on this. But we can certainly take a variety of steps to promote sustainable development. So what is sustainable development?

The UK government has adopted the Brundtland Commission definition (1987) in which sustainable development is defined as "meeting the needs of the present without compromising the ability of future generations to meet their own needs."

I prefer my colleague, Paul Ekins’ definition as “a process that improves the quality of life of people, and especially the least-advantaged people, in the present, while maintaining and restoring the environmental resources and services on which the quality of life of future generations will depend.” It’s all too easy to forget the least advantaged and they are the ones who are always hit hardest.

The UK’s independent Sustainable Development Commission states: “we seek to simultaneously progress economic and social and environmental goals and policies in ways that develop and maintain a good quality of life for us all and enable future generations to do the same.”

The UK Government’s guiding principles (2005) are as follows:

### *Living Within Environmental Limits*

Respecting the limits of the planet’s environment, resources and biodiversity – to improve our environment and ensure that the natural resources needed for life are unimpaired and remain so for future generations.

### *Ensuring a Strong, Healthy and Just Society*

Meeting the diverse needs of all people in existing and future communities, promoting personal wellbeing, social cohesion and inclusion, and creating equal opportunity for all.

*Achieving a Sustainable Economy*

Building a strong, stable and sustainable economy which provides prosperity and opportunities for all, and in which environmental and social costs fall on those who impose them (polluter pays), and efficient resource use is incentivised.

*Using Sound Science Responsibly*

Ensuring policy is developed and implemented on the basis of strong scientific evidence, whilst taking into account scientific uncertainty (through the precautionary principle) as well as public attitudes and values.

*Promoting Good Governance*

Actively promoting effective, participative systems of governance in all levels of society – engaging people’s creativity, energy, and diversity.

## 7. Role of computing professionals

So what can the IT sector and professionals involved in it, do or promote to address the challenges?

### **1     *Embedding it in our professional lives***

For a start, we can seek to embed the notion of sustainable development in our professional lives.

For instance, the Professional Practice for Sustainable Development (PP4SD) argues as follows:

*Professional bodies are beginning to recognise that sustainable development is a key issue for professional practice and the wider role of professionals in society. Since many professional bodies also define the curricula of degree programmes which provide the educational route to membership of the professions, this significant change in emphasis has far reaching implications for degree programmes in HE institutions. As part of this change process, 14 professional bodies have developed a common framework for sustainability to enable them to develop their thinking and practice. They have also developed a generic course on sustainable development for the professions based on systems thinking.*

***<http://www.pp4sd.org.uk/>***

Education in sustainable development is crucial both in terms of raising the general literacy about sustainable development and in enabling professionals to respond effectively in their work.

The Institute of Electrical and Electronics Engineers (IEEE) Code of Ethics starts as follows:

*“We, the members of the IEEE, in recognition of the importance of our technologies in affecting the quality of life throughout the world, and in accepting a personal obligation to our profession, its members and the communities we serve, do hereby commit ourselves to the highest ethical and professional conduct and agree:*

*. to accept responsibility in making decisions consistent with the safety, health and welfare of the public, and to disclose promptly factors that might endanger the public or the environment.”*

IEEE insists on a proactive engagement with sustainable development as a core ethical responsibility of professionals. I believe that applies to all professionals. It is not someone else’s problem: it’s ours.

## 2 *Placing sustainable design at the core of our work*

Better design is absolutely fundamental in addressing issues of energy consumption and the use of natural resources. Eighty percent of all product-related environmental impacts are determined at the design stage (Lorenz, RSA Journal, July 2007).

New ways of designing must ensure that products:

- last as long as possible
- are made of non-toxic materials
- are energy efficient
- are designed for disassembly or re-use
- and where packaging is used it should be fully bio-degradeable.

But that is just the beginning in design terms. We need to focus design on user behaviour and re-conceptualise the entire system that surrounds behaviour in order to promote more sustainable ways of living and working. For example, the winner of the recent RSA sustainable development awards was the Beeline Food Distribution Service in Canada. This was designed to reduce food miles (and thus carbon emissions) and promote the local economy by using local food suppliers. It involved optimising transport routes between local supermarkets, food shops and local producers. It is an on-line computerised system that retailers use to check availability of stock and order produce. It benefits local consumer by ensuring fresher produce, reduces transport and storage costs for suppliers, reduces storage for retailers and makes checking availability and ordering much easier.

Sustainable development principles should be applied to the design of everything we do including of course designing research studies.

## 3 *Arguing for better practices*

All organisations need a sustainable development strategy and targets are needed to address key energy, waste and emissions issues.

Wherever we work we can argue for better practices. For example, in computing by promoting more sustainable strategic practices. More sustainable practice in procurement and supply chain are also essential. There are better ways of doing things than sticking a pc on every desk and replacing it every three years. For instance, SUN claims that their thin systems desk-top service outperforms pcs as follows:

	Typical pc	SUN Desktop service
Average life	3 years	15-20 years
Power use per user	150 watts	4-15 watts
Noise	5dbs	0dbs
Air quality	Negative impact	No impact

We can also review software updating strategies and assess their environmental costs and benefits

We can demand better practices and practise what we preach, always starting with using less of all resources.

At the very least we should support

- regular auditing and monitoring of energy consumption



- good housekeeping practices such as:
  - no screensavers
  - automated switching off processes.

We should tackle materials consumption:

- Paper and printing policies and strategies and defaults
  - Minimal printing
  - Recycled paper
  - Double sided
  - Black & white printing
  - Recycling of toner cartridges.

We should also support recycling and, better still, re-using equipment and supplies.

## 8. Conclusions

But none of this is likely to be enough unless we also start to tackle our collective addiction and dependence on superficial novelty. It is also increasingly clear that there are biophysical limits to growth that pose fundamental challenges to society. This requires more profound reflections on the way we live and a transformation of our relations with the planet. Jonathan Porritt concludes his book 'Capitalism: As If The World Matters' (2005) by arguing that 'it is only sustainable development that can provide both the intellectual foundations and the operational pragmatism upon which to base such a transformation'. It is the best, and perhaps the only chance that we've got.

## References

- Brundtland Commission - WCED (The World Commission on Environment and Development; 1987): Our common future. Oxford.
- Offer, A. (2006), The Challenge of Affluence: Self-Control and Well-Being in the United States and Britain Since 1950, Oxford.
- Porritt J. (2005), Capitalism: As If The World Matters, Earthscan.
- Packard V. (1961), The Waste Makers, Longmans.
- Securing the Future - UK Government sustainable development strategy (2005)  
<http://www.sustainable-development.gov.uk/publications/uk-strategy/index.htm>

## About the Author

Malcolm Rigg is Director of the Policy Studies Institute. He was formerly Managing Director of BMRB International and before that Director of Research at COI Communications. He has published widely on training, education, employment, sport and electronic government. He is currently working on a review of sustainable development in English Higher Education Institutions. He is a Fellow of the RSA, a Fellow of the Market Research Society and an executive member of the Association of Research Institutes in the Social Sciences (ARCISS). The weekend before the conference he is taking part in a sponsored bike ride from Ghent to Aix on behalf of the Market Research Benevolent Association. <http://www.justgiving.com/jgutmann>

# Whither the Web: Web 2.0 and the Changing World of Web Surveys

Mick P. Couper

## Abstract

The World Wide Web is undergoing profound changes. The term “Web 2.0” is being used to refer to the second generation of Internet-based services focusing on collaboration and interaction, as opposed to information. This paper reviews the various flavours of Web 2.0 and what they may mean for the survey research profession. Specifically, the first generation of Web surveys often emulated paper or computer-assisted surveys, without exploiting the unique medium of the Web. As Web surveys mature, we are seeing ways that the changing medium can be used to redefine the way we think of traditional surveys and questionnaires delivered online. The paper reviews some of the design challenges and opportunities for Web surveys.

## Keywords

Internet surveys; Web 2.0; survey design

## 1. Background

I feel like a fraud talking about Web 2.0 and the future of the Internet. My first ever use of a Wiki was as part of the Scientific Program Committee. I don't use Facebook or MySpace. I don't blog. I don't even maintain my own Web page. I've been called a “whining obstructionist” and a Luddite in the past. There are several papers at this conference that discuss Web 2.0 and the implications for survey research (I urge you to read these). So, why me?

I've spent a large fraction of my career thus far trying to understand the impact of technology change on survey research, and have lived through the CATI, CAPI and Web “revolutions”. Web 2.0 may be the next big transition, or it may simply be an evolutionary change in one method of survey data collection.

Along with Roger Tourangeau, Fred Conrad, Reg Baker and others, I have devoted a great deal of research effort to understanding how the design of Web surveys may affect the measurement process and the resultant data, whether positively or negatively. During this time, I have been struck by the degree to which many Web surveys still look and feel like paper surveys. Admittedly, progress is very uneven, with many of the commercial sector way ahead of others, especially in the government and academic sectors. There are many possible

reasons for this, including concerns about perturbing data series, issues around mixed-mode data collection, and maybe a lack of evidence of the benefits of some of these enhancements. Notwithstanding these concerns, I believe we are not exploiting the features already available to us on the Web to enhance or extend survey measurement, to find new ways of asking questions or generating data, or even to find new questions to ask. Web 2.0 may help us do this.

I'm fascinated by the potential of the new Web tools and processes to go beyond the narrow constraints of questions and answers that have characterized surveys for the first century of their use (see Couper, 2005). I'm also fascinated by the hype surrounding Web 2.0, and get that feeling of "here we go again." The introduction of each new survey technology—from CATI to CAPI to Web—has been heralded with overblown promises of how the new method will fundamentally change the way we do surveys. Yet the survey industry has changed remarkably little in its core processes. The fundamentals have remained largely unchanged since the days of the early pioneers. I'm not implying that we have not changed. I'm simply suggesting the change has been more evolutionary than revolutionary.

However, the declining response rates and increasing costs associated with traditional survey methods, suggests that we may be approaching a point where a fundamental shift in thinking may be required. It may be time to think seriously about *what* we do, not just *how* we do it. In terms of inference, we need to engage in serious discussion about the role of probability sampling and the importance of maximizing response rates. In terms of measurement, we may need to think about approaching the issue from a new perspective. Can the Web (whether 2.0 or any other incarnation) do this? No, but the innovations on the Web can certainly serve as a catalyst to get us to address some of these issues. Given this context, let me move to the specific topic of this paper—Web 2.0.

## 2. What is Web 2.0?

So, what is Web 2.0, and what does it mean for us as survey researchers? The term "Web 2.0" was coined by O'Reilly Media in 2004 to refer to the second-generation of Internet-based services, primarily focused on online collaboration. But there are other aspects of so-called second-generation Internet that fall under the broad rubric of 2.0. There are many disagreements about what Web 2.0 means, and who "owns" the term. The term has already become so overused as to lose meaning. When I hear about "Al Qaeda 3.0" on MNSBC, read about "Flaw-o-matic 2.0" in the New York Times, and see an article on "Life 2.0" in the Delta In-flight magazine, all within the same month, it's clear that the 2.0 tag has already become a cliché. A visit to [www.technorati.com](http://www.technorati.com) (itself a Web 2.0 site) in late May 2007 revealed 51,027 blog posts on the topic of "Web 2.0." I'm not going to get bogged down on the semantics of the Web 2.0 label. Instead, I'll offer my view of the evolving Internet technologies and what they mean for survey research.

I see two main streams of development characterizing the next generation Internet tools. The first is the world of MySpace, YouTube, Flickr, blogs, wikis, and the like. The social, collaborative or participatory Web, or Web 2.0.s, if you like, is characterized by broad participation in the process of creating and sharing Web content. No longer is Web content under the control of a few "Webmasters," but now everyone can contribute, in a virtual world of massive co-creation (e.g., Bricklin, 2006). In declaring "You" its Person of the Year for 2006, Time Magazine noted that "...the New Web ... is a tool for bringing together the small

contributions of millions of people and making them matter.” The Time article continues, “Web 2.0 is a massive social experiment and, like any experiment worth trying, it could fail” (Grossman, 2006).

The second major development relates to software tools to permit the social web to exist and thrive, and to transform traditional Web browsing experiences into more fluid interactive exchanges. Probably the best known of these tools is AJAX (or Ajax), or asynchronous JavaScript and XML. The term was coined by Jesse James Garrett of Adaptive Path in 2005, noting that “Ajax isn’t a technology. It’s really several technologies, each flourishing in its own right.” Examples of AJAX can be seen in action at Google Maps, Paguna.com, Pandora.com, and a host of other sites. Internet communication in earlier times (Web 1.0) typically involved a user (client) composing a request or completing a form on a static HTML page, which was then transmitted to the Web server. The server processed the request and then delivered another static web page to the user. While some interactivity was achieved using client-side scripting (such as DHTML or JavaScript), this did not involve communication with the server. Most of the control resided on the server, and communication was a turn-taking process, with the server receiving the entire client request before acting upon it. Even the term “browsing” implied passive behaviour on the part of the users.

In asynchronous applications such as AJAX, the web server can be processing part of the request while the user is in the process of entering information or making choices on the page. As the user narrows the selections, for example, the Web server is delivering appropriate content in the background, ready for when it is needed. This means that the entire Web page does not have to be reloaded every time the user requests a change. The use of the term “asynchronous” in AJAX is a bit of misnomer in that all Internet communication is asynchronous<sup>1</sup>. The idea is that the server does not need to wait for the client to finish in order to begin processing requests, allowing for a much smoother and more “real-time” interaction. AJAX is but one of the tools that permit such interactivity; similar effects can be achieved with Flash, XForms, Lazlo, Adobe Flex, and other second-generation tools. We can call these developments Web 2.0.i, or the “interactions Web” (USA Today, October 26<sup>th</sup>, 2005).

Viewing these developments in parallel, the Web is evolving from a mostly passive text-based information medium with one-to-many distribution to a multimedia, collaborative, and interactive network characterized by many-to-many distribution. What does this all mean for survey research? In addressing this question, I’m going to focus primarily on the second set of recent developments, namely the “interactions Web” or Web 2.0.i. But first, I’ll look briefly at possible implications of the “social Web”, or Web 2.0.s.

### 3. The Social Web and Survey Research

What does the “social Web”, or Web 2.0.s, bring to survey research? Market and social researchers are eying the new developments with a view to gaining access to the thoughts, preferences, suggestions, etc., of a large and growing portion of the population (especially the kinds of consumers that advertisers are most interested in). The amount of personal

<sup>1</sup> In the world of telecommunication, asynchronous operations are those without the use of fixed time intervals. In the computing world, asynchronous means having each operation start only after the preceding operation is completed. In the communication world, asynchronous communication involves a time delay (e.g., e-mail and voice-mail are asynchronous, whereas instant messaging and telephone conversations are viewed as synchronous).

information people share online has grown exponentially over the last several years, with the growth of blogs and other tools for sharing of content. In addition to content, users can organize existing materials online. For example, social bookmarking sites such as Flickr, del.icio.us, or 43things.com allow user-driven tags and classifications, or so-called folksonomies. The data and organizational structures already on the Internet are viewed as a potential goldmine for researchers. In addition, the analytic tools to permit searching, sorting, and analyzing vast amounts of digital data have increasingly made these data readily accessible to researchers. This is not new. Since the early days of the Web researchers have done content analysis of listserv exchanges and studied communication and interaction online through these media. What is changing is the number of people engaged in such activity and the richness of information posted, going beyond text-based exchanges to sharing of visual and other media, such as on Flickr or YouTube. Much of this research was presented at substantive conferences (e.g., sociology, psychology, communication) and the Association of Internet Researchers (a.o.i.r, [www.aoir.org](http://www.aoir.org))<sup>2</sup>. As evidence of this interest, Microsoft Research recently hired its first sociologist, Marc Smith, who studies social networks and social cyberspaces (see Smith, 2006).

As bandwidth increases, these social works are increasing in the richness of the interaction. A number of virtual worlds, inhabited by avatars, already exist. One of the best-known of these is Second Life, with over 6.6 million “residents.” Gartner, Inc. estimates that 80 percent of active Internet users will have a virtual presence by 2011, just four years away (*BusinessWire*, 5/22/07). Several technology companies have a presence on Second Life (see Wagner, 2007), and virtual religion, virtual job fairs and business-to-business communication is already occurring. Virtual market research or social research may be next. Even if research is not conducted in these virtual worlds, the tools are already being evaluated for their potential application to survey research. Several researchers are exploring avatars for survey interviewing, as presented at this conference (e.g., Schober and Conrad, 2007; Johnston, 2007).

For market researchers in particular, these social networks offer insight into how people think, and what’s hot. The wide range of recommendation-based systems, such as the customer reviews on Amazon.com or TripAdvisor.com, can provide much information on the spread of new ideas or products. Research use of these tools can range from observation and analysis, to unobtrusive manipulation (e.g., seeding a site with specific information and seeing what develops), to explicit recruitment of participants in large-scale social networks for market research purposes (e.g., Gadeib and Lichter, 2007). For example, Nielsen BuzzMetrics (<http://www.nielsenbuzzmetrics.com/>) focuses on “consumer-generated media,” tracking and analyzing the social web using sophisticated real-time data mining tools to identify trends, and word of mouth (WOM) marketing (see <http://www.womma.org/>) uses similar approaches to spread new ideas. Social researchers are similarly studying the formation and evolution of online communities, the rules and norms of online interaction, the spread of information in cyberspace, and the like. Some market research companies are developing new research tools to exploit these features; examples include Virtual Surveys, [www.Brainjuicer.com](http://www.Brainjuicer.com), the MindVoyager qualitative idea generation tool from [www.Dialego.de](http://www.Dialego.de), and the Opinionator from Ipsos MORI (Johnson, Mills, and Davies, 2007).

---

<sup>2</sup> This organization may now be moribund, as the website has not been updated in over a year, although there is a posting for the October 2007 conference.

While the social Web opens up a wide range of new research opportunities, often using large amounts of data and complex modelling and analytic methods, this is not really survey research as we know it. For this we need to turn to the new developments in interactivity.

#### 4. The Interactions Web and Survey Research

The second flavour of the new Web, Web 2.0.i, has much more direct relevance for survey research as we know it. The old style of interaction—in which a user completes a form and submits the request to the server, which in turn processes the request and delivers the next page to the user—is being replaced with a much more dynamic exchange of information in a continuous interaction rather than a series of discrete turns. AJAX is the tool most closely associated with these developments, but such interaction is also made possible by improvements in bandwidth and compression and transmission technology. Thus, because of developments in both hardware and software, the old stop-and-go Web is being replaced by a fluid interaction and increased richness of the medium.

The key question for us is whether increasing interactivity and richness of the media improve the quality of survey data, or the survey experience in general. Thus far the answer may be sometimes yes, sometimes no.

While there have already been several demonstrations of the new interactive tools in Web surveys, such as dynamic maps (e.g., Sinibaldi et al., 2006) and virtual product displays (Gadie and Kunath, 2006), research on their effectiveness relative to existing tools still has to be done. To get a partial answer to the question, we must look to research on earlier incarnations of interactive features.

For example, several studies have explored the use of progress indicators in Web surveys. The evidence for use of this feedback feature is quite mixed, with only modest reductions in breakoffs in some cases, and increases in others (see Conrad et al., 2005; Heerwegh and Loosveldt, 2006; Yan et al., 2007). Dynamic progress indicators are ubiquitous on the Web, yet seem to have little if any positive effect in online surveys. This may mean that respondents are not interested in such dynamic feedback, or it may mean that we haven't figured how best to exploit the dynamic nature of the Web to convey a sense of progress and movement through the survey.

Similarly, a number of studies have explored whether (and how) visual analogue scales improve measurement. Decades before the Web was invented, researchers were extolling the value of dynamic rating scales. They seemed ideal for implementation on the Web. Thus far, however, the evidence for their use is mixed at best. Despite a wide range of such tools being made available in Web survey software, and exploration of alternative designs (e.g., Couper et al., 2006; Funke, 2005; Lütters, Westphal, and Heublein, 2007; Stanley and Jenkins, 2007; Thomas and Couper, 2007), they do not appear to perform any better than a series of radio buttons, whether in terms of measurement error, speed of completion, or subjective enjoyment<sup>3</sup>.

A third set of examples relates to exploiting the rich visual nature of the Web. Again, across several studies, the inclusion of images has been shown to affect the answers obtained in

<sup>3</sup> Saris (1989; personal communication) argues that visual magnitude scales work best when respondents can make fine-grained distinctions (e.g., between political candidates). An online experiment currently underway is testing this idea.

surveys but change the context of measurement (e.g., Couper, Conrad, and Tourangeau, 2006; Couper, Kenyon, and Tourangeau, 2004; Hemsing and Hellwig 2006; Witte et al., 2004). This suggests that great care needs to be taken when enhancing Web surveys in this way, as unintended consequences are possible. Much more research is needed to understand how images can be used to improve measurement in surveys, but the preliminary results suggest that doing so may have risks as well as benefits. As we think of the evolving Web, our research needs to expand from static images to video, and to images that can be manipulated directly by respondents (e.g., Athale et al., 2004).

While the research on progress indicators, visual analogue scale, images, and other dynamic tools may suggest that the addition of rich media and interactive features to Web surveys may not enhance measurement or improve the respondent experience, these are early days and much work remains to explore how best to use these tools.

There are some examples of dynamic features where the evidence suggests improved data quality with increased interactivity. One example is the running total or constant sum question (see Conrad et al., 2005). Using client-side scripting, respondents are provided with running tallies as they enter a series of numbers. Doing so leads to better-formed answers than providing no feedback, or providing delayed (client-side) feedback. Similarly the use of dynamic lookup lists which permit respondents to search for a select their answer from a large database appear to improve the quality of answers and reduce the need for coding (see Funke and Reips, 2007; Tourangeau et al., 2004). Finally, work on dynamically changing the shading of items on grids as they are completed provides some evidence for improvements in data quality and the subjective respondent experience (Galesic et al., 2007). While the empirical evidence for the benefits of these tools remains sparse, there is enough promise to suggest more work in this area.

A variety of other visual and interactive tools have been explored in Web surveys, and these may serve as precursors to the potential riches of Web 2.0.i tools. In addition to the use of dynamic maps (Sinibaldi et al., 2006) and virtual store shelves (Gadie and Kunath, 2006) mentioned earlier, card sorting tasks (e.g., Thomas et al., 2005) and ranking tasks (Neubarth, 2006) using graphics and client-side interactivity online have been tested. While evidence for their value has yet to be presented, the level of research activity suggests much promise for these new tools, and the move toward more fluid interaction, rather than relying on client-side scripts, may provide further advantages for these tools.

## 5. Implications of Web 2.0 for Online Survey Research

What are the implications of these developments for survey research conducted via the Internet? Or, why should we care about Web 2.0 and the rapidly changing online world? Will these changes affect what we do as survey researchers, or will it be business as usual? Prognostications vary from the slow, evolutionary perspective to the radical, revolutionary view.

Is it time for Surveys 2.0? Or have we outlived the survey paradigm? Do we need to find new ways to measure highly mobile, highly diverse, hard-to-reach and largely uninterested populations? There are already some proponents of Web 2.0 for whom the answer is a resounding yes, much like earlier proponents of Internet research argued that the old way of

doing surveys was dead (e.g., see Couper, 2000). But these are concerns that go beyond technology shifts.

The biggest challenge facing survey research is not one of technology—although technology can certainly help—but one of the fundamental statistical and social scientific premises underlying surveys. As a method, surveys have served us well for nigh on 100 years. Has their time come to an end? I do not think so; but, nor do I think that doing things the same old way, with greater effort and expense, will do it. The coverage and nonresponse problems facing RDD telephone surveys, along with the challenges of mobile telephone and VOIP, mean that the method is increasingly threatened. The cost of face-to-face surveys has limited their application to a relatively few instances that demand high quality estimates. Representative sample surveys, especially of national populations, are an increasingly beleaguered commodity. I believe it is inevitable that we must change. The question whether this change will be evolutionary or revolutionary is orthogonal to the issue of Web 2.0, but as the world around us changes—and the social Web is a prime example of such change—we need to seriously consider new ways of understanding and gathering data about social behaviour and its correlates. Whether this means adapting our methods to meet the new challenges, or completely rethinking how we conduct research, or both, is an issue we will likely confront over the next several years.

I believe these new tools offer enormous opportunities in terms of improved measurement. Unfortunately the biggest problems facing survey research today are not those of measurement, but rather issues of representation—coverage, sampling and nonresponse. However, we may still well need to change our approaches to measurement, in order to reflect changing user/respondent expectations. The more the Web becomes dynamic, interactive, and adaptive, the more respondents may come to expect from our surveys. Even if we don't find much to gain from Web 2.0 applications, we may need to adapt our methods.

What do we have to do to generate and maintain respondent interest in our surveys? In face-to-face surveys, we believed or argued that the act of being interviewed by a human interviewer had intrinsic social value. This argument is less strong for telephone surveys, although in the early days of telephone interviewing it was believed that the act of receiving a long-distance call from a survey organization was both a rare and important event. This is most certainly no longer the case.

For mail surveys, the value for the respondent must be derived primarily outside of the medium, whether through extrinsic rewards (e.g., incentives) or intrinsic motivations (e.g., normative behaviour, societal good, etc.). We face an even bigger challenge with Web surveys—the very cost factors that make it such an attractive survey medium have led to the proliferation and widespread abuse of Web surveys. Anyone can create and distribute a Web survey, and some days it seems that almost everyone does.

As the Web matured, the relative rarity and novelty of Web surveys quickly wore off. Evidence of this can be seen in the growing number of online panels, and the rapidly declining response rates among members of those panels. Panel members are becoming increasingly selective in which surveys they choose to do, and quickly break off if a survey is not to their liking.

In terms of the Internet in general, the number of alternative sites and the attractiveness of those sites potentially overwhelm the user or respondent. Web sites are increasingly improving the user experience to attract and keep consumers. Given this backdrop, there is



nothing rare, unique, or even attractive about Web surveys. Would changing the intrinsic interest of the survey design improve response rates or data quality? I think not. But would not doing some of these things further erode response rates? I think so.

Given all of this, I would argue that our primary focus should not be on the surface features or aesthetics—entertainment value, if you will—of our Web surveys. We cannot hope to compete against the top commercial and media sites, either in terms of the products and services they offer, or in the resources available to enhance the Web experience for users. Furthermore, going too far in this direction may even backfire, conveying less legitimacy and seriousness of purpose. Rather, our primary focus should be on using the new features of the Web 2.0 world to improve survey measurement. This is not to suggest that aesthetics or the user experience are unimportant—I have spent a good part of my career advocating for improved usability of survey instruments. It is instead to suggest that the focus of our attention—and our research—should be on improving and enhancing measurement. This suggests, as implied earlier, that I think Web 2.0 is not going to solve the most pressing survey problems of the day.

In a way, the two flavours of the new Web mirror the arguments outlined above. Web 2.0.s offers a radical new perspective on measuring a wide range of social attitudes, opinions, behaviours, etc., that have traditionally been explored using surveys. This represents a departure from the paradigm of representative samples with standardized measurement that is characteristic of the survey method.

Web 2.0.i offers a variety of tools to enhance the survey experience and flow of the interview, and improve survey measurement. It is an extension of the methods we currently employ, and does not require a completely new way of thinking. My guess is that we will see—and are already seeing—developments on both fronts in the survey world. Whether one or the other will come to dominate is not clear. At least for the time being, Web 2.0 offers both an expansion of existing research methods and the opportunity to try new things while we attempt to grapple with the key challenges facing representative sample surveys.

## References

- Athale N, Sturley A, Skoczen S, Kavanaugh A, and Lenert L (2004) A Web-Compatible Instrument for Measuring Self-Reported Disease Activity in Arthritis. *Journal of Rheumatology*, 31 (2): 223-228.
- Bricklin D (2006) The Cornucopia of the Commons: How to Get Volunteer Labor. <<http://www.bricklin.com/cornucopia.htm>>
- Conrad FG, Couper MP, Tourangeau R, and Galesic M (2005) Interactive Feedback Can Improve Quality of Responses in Web Surveys. Paper presented at the annual meeting of the American Association for Public Opinion Research, Miami Beach, May.
- Couper MP (2000) Web Surveys: A Review of Issues and Approaches. *Public Opinion Quarterly*, 64 (4), 464-494.
- Couper MP (2005) Technology Trends in Survey Data Collection. *Social Science Computer Review*, 23 (4): 486-501.

- Couper MP, Conrad FG, and Tourangeau R (In press) Visual Context Effects in Web Surveys. *Public Opinion Quarterly*, in press.
- Couper MP, Kenyon K, and Tourangeau R (2004) Picture This! An Analysis of Visual Effects in Web Surveys. *Public Opinion Quarterly*, 68 (2): 255-266.
- Couper MP, Singer E, Tourangeau R, and Conrad FG (2006) Evaluating the Effectiveness of Visual Analog Scales: A Web Experiment. *Social Science Computer Review*, 24 (2): 227-245.
- Fuchs M and Funke F (2007) Video Web Survey: Results of an Experimental Comparison with a Text-Based Web Survey. Paper presented at the Association for Survey Computing's Fifth International Conference on the Impact of Technology on the Survey Process, Southampton, England, September.
- Funke F (2005) Visual Analogue Scales in Online Surveys. Paper presented at the General Online Research Conference (GOR'05), Zurich, March.
- Funke F and Reips U-D (2007) Dynamic Forms: Online Surveys 2.0. Paper presented at the General Online Research Conference (GOR'07), Leipzig, March.
- Gadieb A and Kunath J (2006) Virtual Research Worlds – Simulate the Difference! Efficient Concept Testing with in Virtual Market Simulations Online. Paper presented at the General Online Research (GOR '06) conference, Bielefeld, Germany, March.
- Gadieb A and Lichter C (2007) Online Research 2.0. Join the MindVoyage. Paper presented at the General Online Research Conference (GOR '07), Leipzig, Germany, March.
- Galesic M, Tourangeau R, Couper MP, and Conrad FG (2007) Using Change to Improve Navigation in Grid Questions. Paper presented at the General Online Research Conference (GOR'07), Leipzig, March.
- Garrett JJ (2005) Ajax: A New Approach to Web Applications.  
<<http://www.adaptivepath.com/publications/essays/archives/000385.php>>
- Grossman L (2006) Time's Person of the Year: You. *Time*, Wednesday December 13th, 2006  
<<http://www.time.com>>.
- Hemsing W and Hellwig JO (2006) The Impact of Visualization of Question Types and Screen Pages on the Answering Behavior in Online Surveys. Paper presented at the General Online Research (GOR '06) conference, Bielefeld, Germany, March.
- Johnson AJ, Mills J, and Davies G (2007) Exploring the Practicalities of Developing Web 2.0 Applications in Online Research. Paper presented at the Association for Survey Computing's Fifth International Conference on the Impact of Technology on the Survey Process, Southampton, England, September.
- Johnston M (2007) Automating the Survey Interview with Dynamic Multimodal Interfaces. Paper presented at the Association for Survey Computing's Fifth International Conference on the Impact of Technology on the Survey Process, Southampton, England, September.
- Lütters H, Westphal D, and Heublein F (2007) SniperScale: Graphical Scaling in Data Collection and its Effect on the Response Behaviour of Participants in Online Studies. Paper presented at the General Online Research Conference (GOR'07), Leipzig, March.

- Neubarth W (2006) Ranking vs. Rating in an Online Environment. Paper presented at the General Online Research Conference (GOR'06), Bielefeld, Germany, March.
- Saris WE (1989) Continuous Scales in the Social Sciences: An Attractive Possibility. University of Amsterdam. Methodology Department, Political Science Faculty.
- Schober MF and Conrad FG (2007) Dialogue Capability and Perceptual Realism in Survey Interviewing Agents. Paper presented at the Association for Survey Computing's Fifth International Conference on the Impact of Technology on the Survey Process, Southampton, England, September.
- Sinibaldi J, Crawford SD, Saltz R, and Showen S (2006) Using Interactive Web-Based Maps to Collect Geographical Data in a Student Survey. Paper presented at the annual meeting of the American Association for Public Opinion Research, Montreal, Canada, May.
- Smith M (2006) Visualization of Online Social Networks: Working with MSR Community Tools and Related Social Software. Keynote address at the General Online Research Conference (GOR '06), Bielefeld, Germany, March.
- Stanley N and Jenkins S (2007) Watch What I Do! Using Graphic Input Controls in Web Surveys. Paper presented at the Association for Survey Computing's Fifth International Conference on the Impact of Technology on the Survey Process, Southampton, England, September.
- Thomas RK, Bayer LR, Johnson A, and Behnke S (2005) A Comparison of an Online Card Sorting Task to a Rating Task. Paper presented at the annual meeting of the American Association for Public Opinion Research, Miami Beach, FL, May.
- Thomas RK and Couper MP (2007) A Comparison of Visual Analog and Graphic Rating Scales. Paper presented at the General Online Research Conference (GOR'07), Leipzig, March.
- Witte JC, Pargas RP, Mobley C, and Hawdon J (2004) Instrument Effects of Images in Web Surveys: A Research Note. *Social Science Computer Review*, 22 (3): 363-369.

## About the Author

**Mick Couper**, PhD, is a Research Professor in the Survey Research Center, Institute for Social Research, at the University of Michigan, and a Research Professor in the Joint Program in Survey Methodology. He conducts methodological research on a wide range of issues, including nonresponse, computer-assisted surveys and Web data collection. He can be contacted at ISR, P.O. Box 1248, Ann Arbor, MI 48106, USA; tel. +1-734-647-3577; e-mail [mcouper@umich.edu](mailto:mcouper@umich.edu).

# Survey data, context and event data<sup>1</sup>

Ineke Stoop

## Abstract

In line with the development of the social sciences into a more empirically oriented and methodologically sophisticated science, the last half century has seen a large increase in the use of cross-national survey data. One drawback of surveys, however, is that the data they generate from respondents are often isolated from their context and thus potentially misleading (or at least uninformative) about material facts. As a result, individual behaviour and attitudes are described and analysed as if the individuals concerned were living in a society without history, geography, social relationships, social and institutional structures and social and political events. From the start of the European Social Survey (ESS), the need was recognised for a context and event data inventory to provide the necessary background information when analysing survey data. This paper presents an overview of what has been done so far – with examples from different rounds of the ESS – and of the plans for the future.

## Keywords

Contextual information, integrating survey data with data from other sources, events

## 1. Introduction

Cross-national projects require an amount of documentation that is usually neither provided nor necessary in national studies. Mohler and Uher (2003, p. 312) mention the need for:

*A 'thick' description of the socio-cultural contexts of a cross-national study and the different 'fields', including all the relevant contextual statistics. Some of these are standard and basic (e.g., population statistics), others are determined by the scope and topic of a study (e.g., divorce rates or GNP). This information should also cover regulations and infrastructure information relevant for understanding respondent answers in a given context (e.g., divorce laws, abortion regulation).*

The European Social Survey ([www.europeansocialsurvey.org](http://www.europeansocialsurvey.org), ESS) is a bi-annual survey measuring social, cultural and attitudinal changes across Europe. By the third round of data collection in 2006/2007 more than 30 European countries were involved. The need to collect contextual information was recognised early on in the early preparations of the ESS; in addition, in order to aggregate data and institutional information, it was decided to construct

<sup>1</sup> A number of the ideas in this chapter on the use of news sources and ways to improve event reporting stem from discussions with and papers from Howard Tumber (City University London) and Paul Statham (Bristol University). Thomas van Putten (SCP) designed and built ESS-Eventnet and Josine Verhagen (SCP) analysed the relationship between events and response behaviour. Ideas on context information in relation to survey data come from discussions with and papers from Kristine Kolsrud and Knut Skjåk (Norwegian Social Science Data Services), both members of the Central Coordinating Team of the European Social Survey.

an event database both to provide context for the survey data and to act as a source for historical micro-analysis (ESF, 1998, p. 33). The aim was to control the impact of contextual factors or events, as they can influence whole societies in a way that is not uniform across the countries in the ESS. The intention is that the contextual and event data inventory should offer the researcher a concise, pertinent synopsis of major political, social and other potentially relevant events in the ESS countries. This is especially important as the ESS covers a wide range of topics that partly changes over time.

Considering that the ESS survey data will be analysed by future analysts from a wide range of countries (as at mid-March 2007 the ESS had 13,158 registered users from 162 countries, including the Cocos Islands, the Holy See, Mongolia, Turkmenistan and Zimbabwe, and 7,496 data downloaders), one can imagine the different kinds of national context factors that need to be taken into account when analysing cross-national surveys. To start with, one needs to have some general knowledge of living conditions and the social and political situation in a given country. Can a farmer in Russia be compared with a farmer in Denmark? Is living in a flat in central Paris to be compared with living in a flat in rural Latvia? Do opinions on European government vary (positively or negatively?) with the length of EU membership, or do they depend on whether or not a country holds the EU presidency at the time of the survey (EOS Gallup Europe, 2003)?

National laws and regulations may also have an impact on survey outcomes. Consider the well-known fact that divorce has a negative impact on satisfaction and happiness (Lucas, 2005). Can it be inferred from this that a low divorce rate indicates a high level of happiness? Veenhoven (1983) found that married people are happier than singles, and that among singles divorcees are the most unhappy, followed by widows and widowers – except in Italy and Ireland, where the differences were minimal and people were much unhappier in general. Braun and Mohler (2003, p.111) observe that divorce in Ireland is not possible. This might explain why divorce and happiness are not related in Ireland, and why Irish married people are relatively unhappy. Without knowledge about institutional arrangements, one might draw faulty conclusions. This would also be true when studying the relationship between response behaviour and voting behaviour. Those who are interested in politics are more likely to vote and more likely to participate in surveys (Stoop, 2005) irrespective of the survey topic. In those countries where voting is mandatory, this relationship will be absent.

In addition to general living conditions and institutional arrangements, events occurring during fieldwork may have an impact on survey outcomes. Bradburn (1969) gave an example of the (unexpected) consequences of major events in his study of the trauma caused by the assassination of President John F. Kennedy. This event not only caused feelings of shock, grief and personal loss, but also occasioned an increase in interpersonal communication and social cohesion. Other possible effect of major events are growing insecurity and dissatisfaction during redundancies, bankruptcies and shutdowns and following terrorist attacks, and less trust in politicians during political scandals. Although personal and family events may of course usually have a more profound impact on people's lives and thoughts than more distant political events, these personal sources of turbulence do not tend to have a systematic impact on survey outcomes. They are, in effect, randomly distributed across the population. In contrast, what is important in comparative surveys are any systematic effects of events on attitudes at a particular time or in a particular place. It is this sort of turbulence that can cause differences between countries, changes over time and variations between subgroups of the population.

This paper will describe how information on context and events has been collected in the European Social Survey, give an impression of important events during fieldwork and outline problems and plans for the future. A theoretical background and more information on the history of event reporting in the ESS are given in Stoop (2007).

## 2. Context

The most common type of data used to measure context are usually called ‘regional data’ (NSD, 2006). These data describe the characteristics of territorial and spatial units, such as municipalities, communes, census tracts, voting districts, counties, countries, etc. The units are usually (but not necessarily) clearly defined in space, and differ widely in their purpose. They may be institutional, such as a constitution, electoral system, or school system for a country. They may be distributions or aggregates of micro-units such as population data, voter counts, etc. Or, they may be more structural, such as a party system. Viewed from the bottom up, these data are termed ecological data, and any sophisticated collection of survey data should ideally be supplemented by such data. These concepts of regional, aggregate and ecological data focus on different aspects of the information and the differences may have some practical and some substantive relevance.

From the beginning of the ESS it was envisaged that contextual information would be provided on each of the participating countries. Ideally, a Europe-wide database of background contextual data should become available that will routinely serve as a source for European surveys. Academically driven European regional and contextual databases are hard to find. Electoral political scientists have built an elections and electoral systems database, but almost all the major database developments at regional level (such as NUTS) have been produced by official statistical agencies, notably Eurostat. Valuable as they are, they cannot quite cater to the needs of the academic community, partly because of their content, partly because of their scope (not all European countries are included) and partly because of their limited user-friendliness. Other Europe-wide regional data are prohibitively expensive and are incomplete in terms of both content and documentation. For comparative purposes – as in the ESS – users are regularly confronted with long lists of footnotes and caveats indicating that comparability may be limited.

Due to lack of funds and lack of time, the original aspirations of the ESS with regard to contextual information were seriously downscaled. In the first two rounds of the ESS (2002-2005), only the following context resources were available:

- an overview of freely accessible sources of country information available on the Internet: [www.scp.nl/users/stoop/ess\\_events/links\\_contextual\\_data2003.htm](http://www.scp.nl/users/stoop/ess_events/links_contextual_data2003.htm)
- contextual information on countries and regions for Round 1, provided by the ESS data archive at NSD: <http://ess.nsd.uib.no/files/2003/ESSMacrodata.xls>
- contextual information provided in each round by National Coordinators: <http://ess.nsd.uib.no> → ESS Round 1 and ESS Round 2 → Survey Documentation → Appendix 1, Population statistics and other documentation.

In 2005 the ESS was awarded Europe’s top annual science award, the Descartes Prize. A large slice of the prize money will be spent on developing a contextual dataset framework (NSD, 2006). The aim is to build a Europe-wide information framework in which background contextual data are linked with data from the ESS. The framework will facilitate integration of relevant information from scattered sources throughout Europe, generating added value for

the ESS. The individual ESS data and the contextual data will be available online using the same data dissemination tool, and will be linked to data at European regional level, enabling mouse-over data browsing and cartographical and graphical presentations. The aim is to include documentation on structural differences at the national and regional level, such as electoral and educational systems. The use of NUTS as a common unit dimension opens the

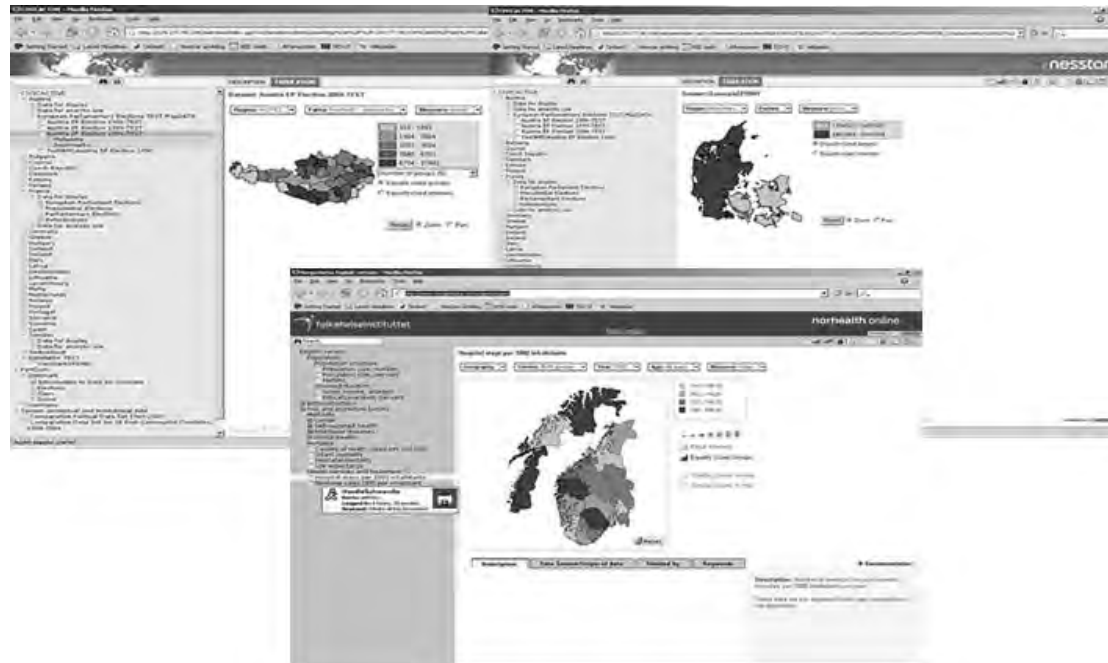


Figure 1. An example of linking survey data, context data and NUTS levels

way for cartographic presentation of the data, since coordinates for maps are developed according to the NUTS standard. The Nesstar technology facilitates the use of maps based on these kinds of coordinates. An example of data presentation at NUTS level is given in figure 1 (NSD, 2006).

### 3. Media-reported events

The death of a loved one, the loss of a job or being a victim of a serious crime may have a major impact on individual well-being and attitudes. The contextual impact on individual response behaviour will not create major effects, however, as long as the contexts and events vary individually in an idiosyncratic fashion. The impact of events must be considered and, whenever possible, controlled for as soon as whole societies are influenced in a way which is not uniform across countries. This is especially true in general, longitudinal studies. For this reason, the ESS focuses on events that draw the attention of the general public.

From the range of events that occur in the world every day, only a small selection are salient to the public. It is these salient, well-popularised events, with high exposure either to large subgroups in a country or to an entire country, or even to a large group of countries, that have the potential for focusing and shaping the attention of members of the public. Information on these events is conveyed through the mass media, hence the focus on media-reported events. Communication researchers have developed several theories on how the mass media may influence public opinion (Cohen, 1963; Nas, 2000), from the hypodermic needle theory via the existence of selective filters between medium and recipient, to the agenda-setting

approach. Increasingly, the role of the 'recipient' of information and cultural difference is being acknowledged (Pfetsch, 2004).

Stoop (2007) presents the rationale of event reporting in the ESS. Here, suffice it to say that if, to give a hypothetical example, questions on values and attitudes to the environment are part of a survey, one would very much like to know if during the fieldwork newspapers were full of reports on Al Gore's 'An Inconvenient Truth'. Relevant considerations in a cross-national study are that this media attention will be specific for one round of the survey, may vary across countries (especially as fieldwork periods are not always identical) and may have a different impact depending on the timing of each individual interview.

Within the context of the ESS, financially driven choices have been made as to which media to use as source material for event reporting. Although there were arguments for monitoring television news, the difficulties and costs of systematically recording and coding television news bulletins across over 30 nations were too daunting. Instead, only newspapers have been monitored, in the hope that television news agendas and press agendas tend to coincide in many cases.

In the ESS major events are reported that draw headlines in the newspapers, whether national or international, as events that take place far away may have an equivalent impact to events closer to home. The war in Iraq, for instance, had seemingly major (and different) national ramifications which had a major impact on national politics in many countries. The ultimate aim is to be able to link events to survey responses, so in the ideal situation events could be identified that have a clear start and end. However, many events do not behave like this in reality, but linger on and sometimes re-emerge. Consider the referenda on the European Constitution in France and the Netherlands in 2005, where the Constitution was rejected. These 'events' continue to have an impact on EU governance in 2007. In addition, the referenda took place during 'difficult' political times in their respective countries, and may well have tapped into hostility towards their own national governments as much as opinions on the proposed Constitution. The effect of the Dutch and French 'No' votes may have influenced attitudes towards the EU in many other countries simultaneously. This shows that not all 'events' are related to their ostensible subject matter, nor to the country where they take place. On the other hand, 'identical' events may have different consequences in different countries. Similar political scandals, for instance, can lead to a decrease in trust in one country and to mild amusement in another where scandals are a fact of life.

### Event reporting guidelines and tools

Event reporting in the ESS started in Round 1 (2002/2003) with rather loosely formulated guidelines asking National Coordinators and their teams in each participating country to send in monthly reports on events that received 'prominent attention' in national newspapers. These events were defined as meaning 'front page news' or 'appearing regularly in larger articles on the inside pages' for several days. The selection of newspapers was more or less left to the national event reporters. Reporters were asked to place events in fixed categories, partly inspired by the World Handbook of Political and Social Indicators Series (Taylor and Jodice, 1986), to provide keywords and a description, to give a start and end date (if possible), to mention the source, and to assess the likely size and direction of the event's impact on the survey responses.

In Round 2 (2004/2005) the guidelines were revised, resulting in a more standardised format. Events now had to be reported weekly, rather than monthly. As in Round 1, incoming event



reports were posted on a website (accessible via <http://ess.nsd.uib.no>) to give an up-to-date overview of incoming event reports, and also to show guidelines, information notes and background information about the process. This transparency was helpful not only to users who wanted an overview of weekly or monthly events in each country, but also to event reporters themselves as a way of checking how their colleagues in other countries were using the system. The webpage also contains FAQs and ultimately provides the final ESS media-reported event inventory for each round.

In Round 3 (2006/2007), ESS-eventnet was implemented ([www.scp.nl/ess/events](http://www.scp.nl/ess/events)), a web tool that allows national event reporters to upload events themselves. The reported events are now part of a database. Figure 2 shows what ESS-eventnet looks like. The webpage provides access to reports on each individual event and can also generate a number of simple reports. Practical improvements are foreseen for Round 4 (2008/2009).

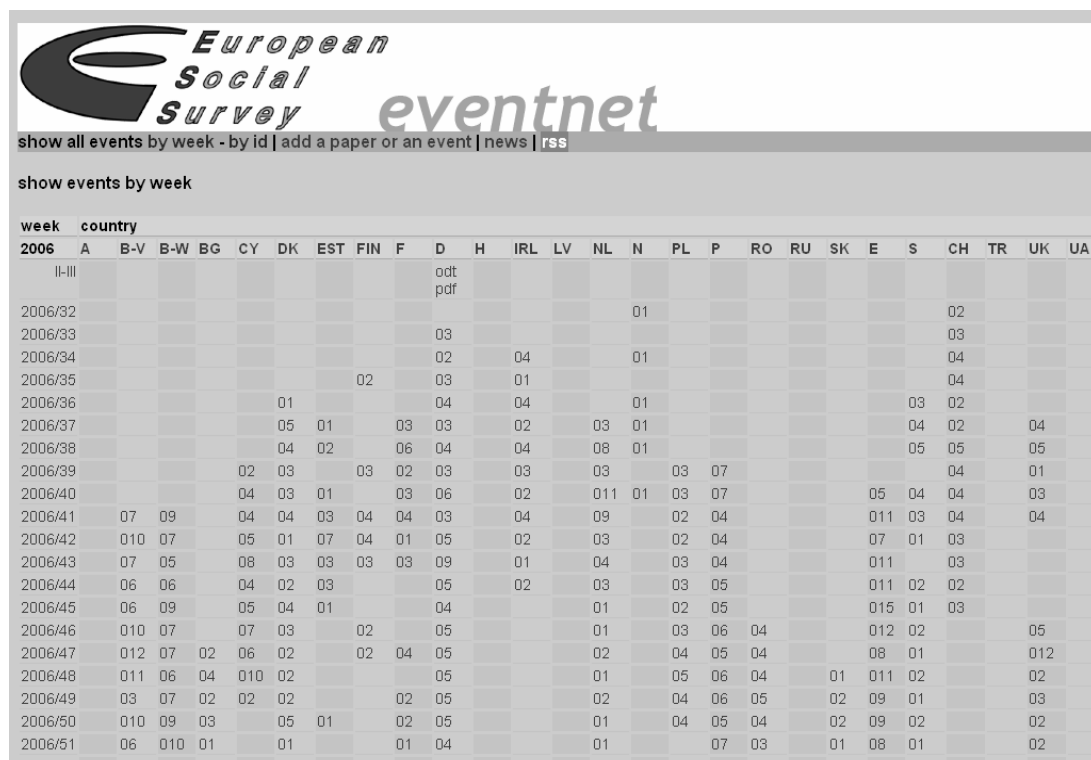


Figure 2. Screenshot of ESS-eventnet: number of events per country per week

### Media-reported events reported in the ESS

Many things happened in Europe during the ESS fieldwork, and national event reporters worked hard to cover these events. Early in April 2007, when a number of countries had not yet finished the fieldwork for Round 3, almost 1,200 events were described in the ESS-eventnet database. Table 1 gives an overview of the number of events per country.

Table 1. Number of media-reported events reported in ESS Round 3 by April 2007

	2006	2007	Total		2006	2007	Total		2006	2007	Total
Belgium-	90	74	164	Netherlands	57	3	60	Ireland	29		29
Belgium-	88	30	118	Cyprus	57		57	Romania	27	1	28
Spain	117	1	118	UK	44	4	48	Estonia	22	3	25
Germany	88	11	99	Denmark	45		45	Finland	20	3	23
Portugal	77		77	Poland	38		38	Bulgaria	13	3	16
Switzerland	74		74	Sweden	33	5	38	Slovak Republic	7	9	16

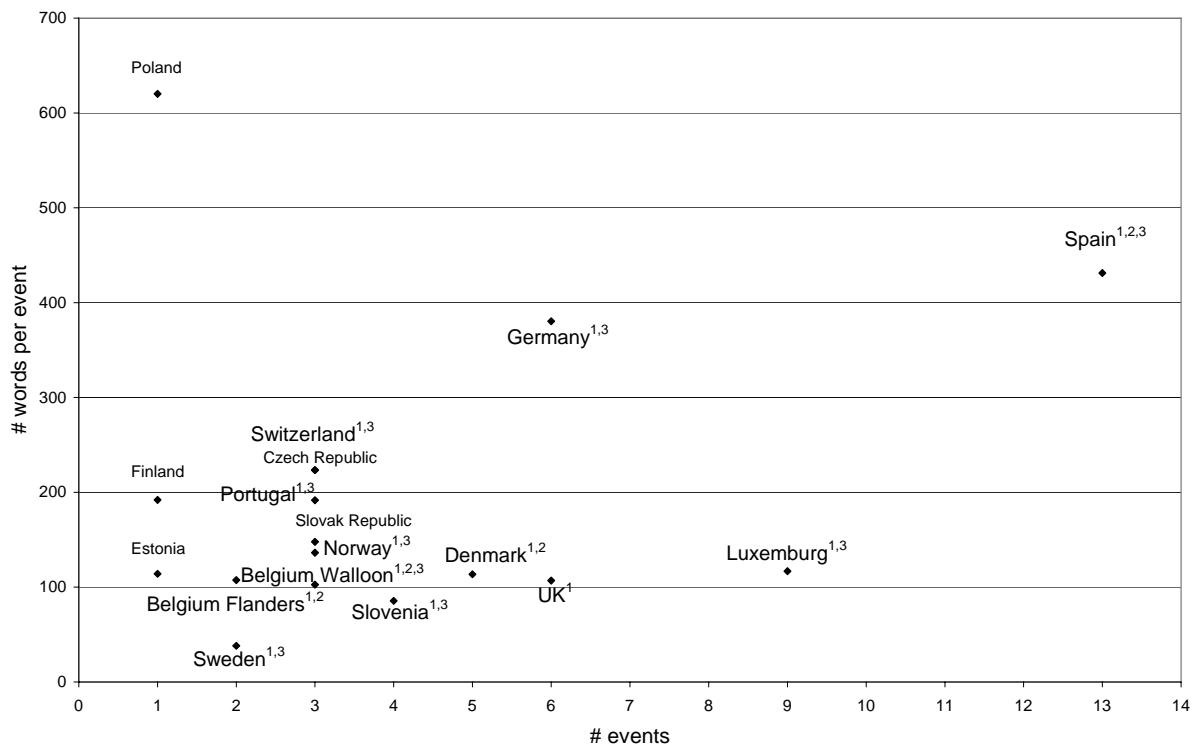


Figure 3. Number of events and average number of words per event in ESS Round 2, 1-7 November 2004

1) Mentions US elections; George Bush; 2) Mentions murder of Theo van Gogh in the Netherlands; 3) Mentions Arafat

France	36	25	61	Slovenia	30	30	Norway	6	6
Total								773	148 921

Table 1 illustrates the vast differences in the number of events reported in Round 3 of the ESS. Figure 3 gives more detail for one week in Round 2, in November 2004. The number of reported events in this week varies from 1 to 13, and the average number of words in reported events (including category labels, etc.) ranges from 38 in Norway to 620 in Poland. Major events that drew front page headlines for many days (US elections, Arafat dying, assassination of Theo van Gogh) were mentioned in only some of the countries. These results indicate that there are substantial differences in event reporting between countries that are unlikely to be due to the absence or presence of actual events, and more to national or even individual reporting habits.

### Countries, events and opinions?

Figure 4 presents the results of a correspondence analysis based on word counts of event reports in Round 1 (2002/2003). Events that were infrequently mentioned or related only to a single country have been removed (such as Haider's success in Austria or the Prestige oil spills in Spain). The data used for this correspondence analysis are the number of times a particular word is mentioned in a particular country. In the graph, countries that report similar events are plotted close together, along with the events that distinguish them from other countries. Events that are reported by every country are in the centre of the graph.

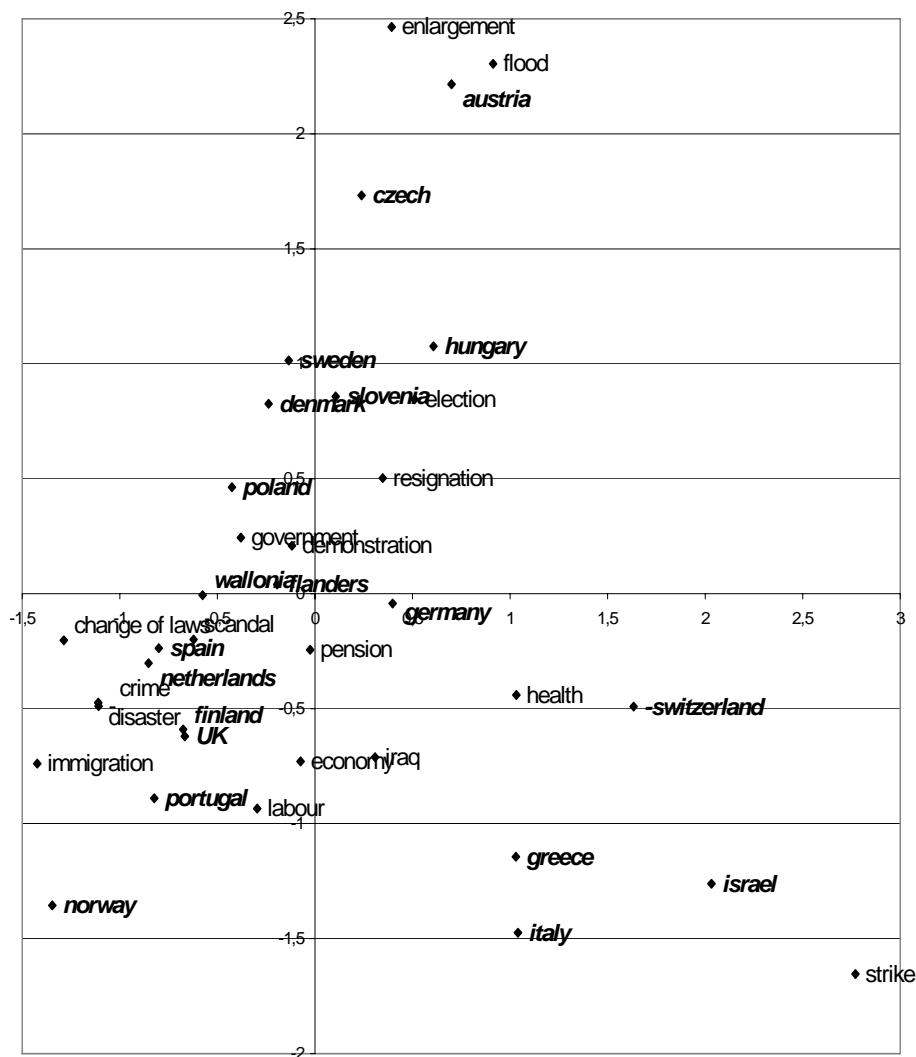


Figure 4. Countries and events, ESS-Round 1, weighted by word count

Austria, Hungary and the Czech Republic are all located in the upper right quadrant of figure 3, because they all suffered severely from flooding in the autumn of 2002. EU enlargement was also an important topic in these countries, as it was in Denmark, which at the time held the EU presidency. Greece, Israel and Italy were hit by strikes. Despite the less than standardised event reporting from different countries, figure 3 shows that the coding and analysis of the present data can provide an overview of major events during Round 1 that might have had an impact attitudes and opinions. On the other hand, it is difficult to obtain a clear picture of the short-term effect of particular events on attitudes. There was, for instance, no clear impact on trust in politicians or institutions (not even the UN), which might have been expected because of the considerable political turmoil at the start of the Iraq war.

The reality is that measuring the impact of events on attitudes is a highly uncertain and complex affair. For instance, despite the political turmoil in the Netherlands following the murder of Theo van Gogh by a Muslim fundamentalist in November 2004, responses to a question on freedom of speech in a Dutch survey (Verhagen, 2006) showed an abrupt change in the immediate aftermath, but rapidly returned to the original level. Figure 5 presents the average score per week on the question of whether immigrants have a negative or positive impact on Dutch society in Round 2 of the ESS. The figure shows a clear dip after the murder of Van Gogh, but weekly averages quickly return to their original level in December. Surprisingly, there is a very sharp drop in January 2005. One possible cause is that converted refusers – i.e. people who originally refused to participate in the survey but later did so – may

have a more negative attitude towards immigrants (Billiet and Philippens, 2004). Surprisingly, the responses of the small number of respondents who participated in 2005 without having ever refused were just as negative as those of the initial refusers. There is still no explanation for this. As Bradburn (1969) has recommended, the event reports may need to be supplemented by more systematic research on psychological reactions to significant events.

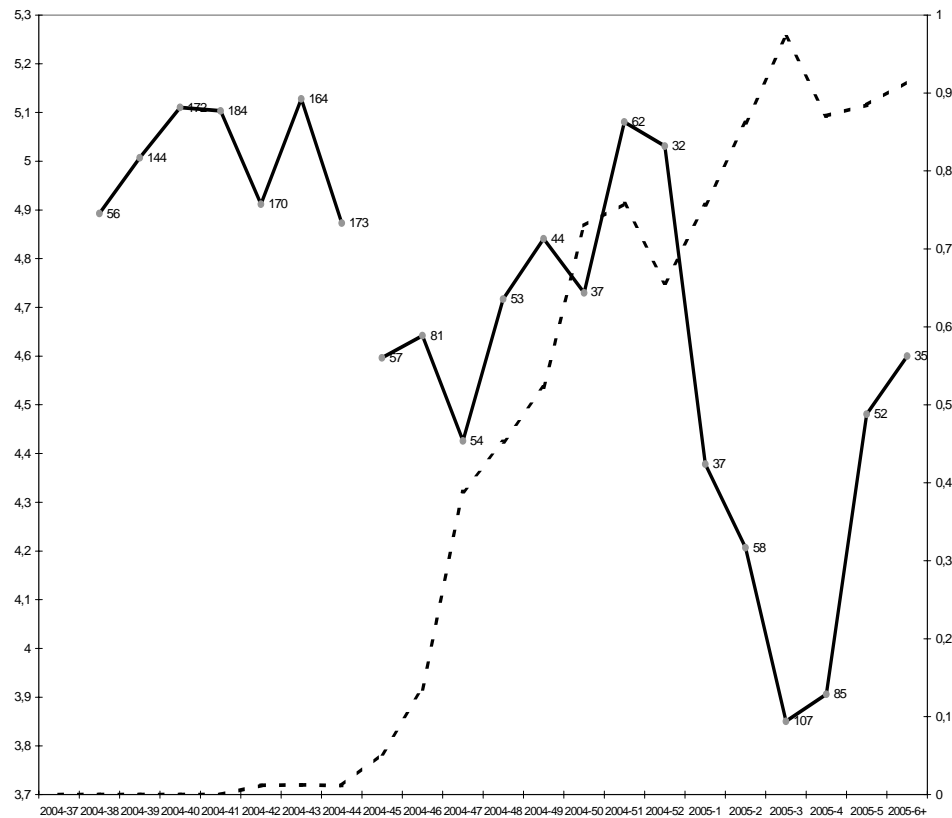


Figure 5. Average score per week on question of whether immigrants are an asset to society (0-9) (left-hand scale —) and percentage of respondents who were initially reluctant (right-hand scale ----). Numbers indicate number of respondents per week

#### 4. Problems and plans

This overview of contextual information and event reporting in the ESS shows that – useful as it undoubtedly is – it could certainly be improved upon. Several issues remain to be resolved. With regard to contextual information, as outlined in section 2, there are specific plans and projects for improvement. The situation with regard to event reporting is more complicated. Firstly, event reporting in its current form, being carried out by ESS National Coordinators and their team, is very labour-intensive. Secondly, and probably related to this, there are large differences between countries in the number, type and detail of events reported. Thirdly, event reports are based on what is reported on the front pages of newspapers. Front pages within one and the same country differ from newspaper to newspaper; newspapers also differ across countries, and the selection of newspapers is left to the national event reporters. Events might in future have to be taken from articles on other pages, and possibly from editorials as well as Op-Ed pages. Another aspect that cannot be ignored in this perspective is the problem of language: the ESS event reports are in English, whereas the newspapers used are written in more than 30 different languages.

Underlying problems related to event reporting are reflected in basic questions such as ‘What is an event? (the war in Iraq? the release of British hostages in Iraq?)’, ‘When is an event? (on 9/11? in the months after 9/11? at the first anniversary of 9/11?’ and ‘Where is an event? (in Iraq? at the homes of the families of Italian soldiers serving in Iraq?)’. These questions have certainly not been resolved in the present ESS event reports.

A different problem is that events that have the largest impact on opinions and attitudes may well occur between fieldwork periods. In the present set-up these events will be ignored. A last and very fundamental problem is how to relate individual events directly to individual opinions. Can they be modelled, and can specific events be part of the explanatory variables in multilevel analyses? Figure 5 illustrates how complicated this might be in the light of the relationship between response behaviour and the timing of events.

Modelling events will as a minimum require more standardised and cross-nationally comparable event data. At present, two ways are foreseen to achieve this. Firstly, an experiment has been conducted on computer coding of events (Schrodt, 2001). As part of a project intended to provide an infrastructure for the ESS, funded by the European Commission, Cornelia Züll and Juliane Landmann (2003) have already begun carrying out experiments on ESS data using automatic coding. The results will be available in 2007. As part of the same project, researchers at Bristol University and City University (Statham and Tumber, 2007) plan to develop a more standardised, impartial, comprehensive and accessible tool for event reporting. They will still use newspapers as the primary source of event reports, but will use the electronic versions from the Lexis-Nexis database rather than the paper versions. Lexis-Nexis provides an on-line tool for searching newspapers in a wide variety of countries, but – where no such coverage exists – relevant newspapers (intact or on microfiche) can be bought for quarrying at a later date. Events will be coded *post hoc*. This enables the coding of different countries to proceed at different paces.

In building this tool a wide range of newsprint media will be studied (left-wing/right-wing; elite/mass readership). A systematic coding frame of event variables will be developed comprising elements such as time of event, place, actor, geographical scope, etc. The proposed methodological starting point for coding news is the ‘political claims analysis’ (Statham and Tumber, 2007). ‘A claims-making act (shorthand: a claim) is a strategic action in the public sphere. It consists of intentional public speech acts which articulate political demands, calls to action, proposals, and criticisms, which, actually or potentially, affect the interests or integrity of claimants and/or other collective actors in a specific issue-field.’ Of course, in the ESS framework, the scope will have to be broadened from the political arena to include more encompassing economic, social and cultural events.

The resultant coded event database is intended to help researchers determine what has happened during fieldwork (or possibly between fieldwork) periods in the ESS countries, but also the relative salience of events in different countries and the national ‘meaning of events’. The database would also provide information on specific time periods and/or issue fields, making it possible to find clues as to whether any special national or cross-national factors have influenced the responses in particular rounds of the ESS, and of course in any other comparative survey.

## References

- Billiet, J. and M. Philippens, M. (2004) Data quality assessment in ESS Round 1. Between wishes and reality. Leuven: K.U.Leuven, Departement Sociologie. Onderzoeksverslag Afdeling Data-verzameling en Analyse. DA/2004-38. 24p. (Paper presented at the Sixth International Conference on Social Science Methodology. Amsterdam: August 17 – 20).
- Bradburn, Norman M. (1969) The Structure of Psychological Well-Being (chapter 11: Social Trauma: The Assassination of President Kennedy). Chicago: Aldine Pub. Co. Online edition: Published May 2001 by the NORC Library. [cloud9.norc.uchicago.edu/dlib/spwb/index.htm](http://cloud9.norc.uchicago.edu/dlib/spwb/index.htm)
- Braun, Michael, and Peter Ph. Mohler (2003), Background Variables. In: J.A. Harkness; F. J. R. van de Vijver, P. Ph. Mohler, Cross-Cultural Survey Methods. Hoboken: Wiley, pp. 101-116.
- Cohen, B. (1963) The Press and Foreign Policy. Princeton: Princeton University Press.
- EOS Gallup Europe (2003) Flash Eurobarometer 142: Convention on the Future of Europe.
- ESF (1998) Blueprint for a European Social Survey. Strasbourg: European Science Foundation.
- Lucas, Richard E. (2005) Time Does Not Heal All Wounds. A Longitudinal Study of Reaction and Adaptation to Divorce. *Psychological Science* 16 (12), 945–950.
- Mohler, Peter Ph., and Rolf Uher (2003) Documenting Comparative Surveys for Secondary Analysis. In: J.A. Harkness; F. J. R. van de Vijver, P. Ph. Mohler, Cross-Cultural Survey Methods. Hoboken: Wiley, pp. 311-327.
- Nas, M.A.J.C. (2000) Sustainable Environment, Unsustained Attention. A Study of Attitudes, the Media and the Environment. Den Haag, Sociaal en Cultureel Planbureau (in Dutch, summary in English).
- [www.scp.nl/english/publications/summaries/9057495244.html](http://www.scp.nl/english/publications/summaries/9057495244.html).
- NSD (2006) Proposal for integrating contextual data with survey data in ESS. Paper for the 1st joint meeting of the ESS Scientific Advisory Board and Funders' Forum, Strasbourg 9 November 2006.
- Pfetsch, Barbara (2004) The Voice of the Media in European Public Sphere: Comparative Analysis of Newspaper Editorials. Integrated Report WP3 of EC FP5 project: The Transformation of Political Mobilisation and Communication in European Public Spheres.
- Schrodt, Philip A. (2001) Automated Coding of International Event Data Using Sparse Parsing Techniques. Paper presented at the International Studies Association, Chicago, February 2001.
- Statham, Paul, and Howard Tumber (2007) News Events Analysis: Measuring the 'Political Climate'. Briefing Paper ESSi: NA4 Developing Event Reporting. The development of a methodological tool. Presented at the Meeting of National Co-ordinators, Mannheim, February 2007.
- Stoop, I.A.L. (2005) The Hunt for the Last Respondent. The Hague: Social and Cultural Planning Office.

- Stoop, Ineke (2007) If it bleeds, it leads: the impact of media-reported events. In: Roger Jowell, Caroline Roberts, Rory Fitzgerald & Gillian Eva. Measuring attitudes cross-nationally – lessons from the European Social Survey. London: Sage, pp. 95-112.
- Taylor, Charles Lewis, and David A. Jodice (1986) World Handbook of Political and Social Indicators III: 1948-1982 [Computer file]. Compiled by Charles Lewis Taylor, Virginia Polytechnic Institute and State University. 2nd ICPSR ed. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor].
- Veenhoven, R. (1983 ) The growing impact of marriage. Social Indicators Research, 12, 49-63.
- Verhagen, Josine (2006) Robuuste meningen. Het effect van responsverhogende strategieën bij het onderzoek Culturele Veranderingen in Nederland. The Hague, SCP (in press).
- Züll, Cornelia, and Juliane Landmann (2003) European Social Survey and Event Data. ZUMA, working paper.

## About the Author

**Ineke Stoop** is head of the department of ICT and Data Services at the Netherlands Institute for Social Research/SCP. She is responsible for the surveys sponsored by the SCP and for research on data quality. She has specialised in nonresponse issues. She is also a member of the Central Coordinating Team of the European Social Survey. She can be contacted at SCP, Postbus 16164, 2500 BD, The Hague, Netherlands, tel. +31 70 3407671, fax +31 70 3407044, e-mail [i.stoop@scp.nl](mailto:i.stoop@scp.nl).

# Reporting Societal Events to Facilitate the Interpretation of Survey Results

Cornelia Zuell & Juliane Landmann

## Abstract

Societal events are events like elections, significant changes in laws, demonstrations, but also extreme weather conditions. All such events can have an effect on a society and, consequently, influence the attitudes of its population. When conducting a survey, the impact of an event must be considered and, whenever possible, controlled especially in survey projects in which different countries participate. However, manual identification of significant events that occur during the data collection phase is a very error-prone and time-consuming task. Therefore, we have developed a procedure to identify events using a combination of two different approaches of the (quantitative) computer-assisted content analysis: the reference text technique and the statistical association approach. On the basis of distinctive features of word usage in a so-called reference text corpus and in newspaper texts of a specific time period in which events should be located, words are selected and classified by means of an exploratory factor analysis. We will demonstrate this procedure by means of an example in which we will identify events automatically in Germany as well as in Great Britain. First, we will describe the composition of a reference text corpus. Thereafter, we will provide details for the calculation of relative differences between the relative frequencies for each word in the reference corpus and in the newspaper text. We will use the words with the highest relative differences as keywords for further analysis. Based on the co-occurrence of these keywords in each newspaper article, we will conduct a factor analysis to identify the events.

## Keywords

Computer-assisted content analysis, event reporting, survey context variables

## 1. Introduction

Since the beginning of the first data collection period of the ESS, the events occurring during the data collection phase have been documented in all countries of the ESS (Stoop 2002, 2004, 2006). The assumption is that respondent behaviour or answers to some questions are influenced by significant events in various areas. Hence, the impact of an event must be considered and, whenever possible, controlled across the countries in the ESS. Each national coordinator in the ESS is asked to provide major national events that could influence the answers to questions or respondent behaviour in general. The way how events are to be identified is specified in the guidelines for national coordinators<sup>1</sup>. In 2002, survey



coordinators were asked to send an overview of events each month. In further rounds, weekly reports have been requested.

Unfortunately, manual documentation of events is a very time-consuming and error-prone task in such cross-border projects. The ESS event data collection lacks consistency, in the sense that the definition of a major event seems to be handled quite differently in each ESS country. The number of reported events in round 3 of the ESS varies from 7 (in Norway) to 197 (in Spain) for the whole data collection period. The number of reported events varies from 1 to 15 per week (<http://www.scp.nl/ess/eventnet/>). Major events that drew front-page headlines for many days (for example, the climate change report or the execution of Saddam Hussein) were mentioned in only some of the countries.

Therefore, we will propose an approach to identify major events using computer-assisted content analysis.

The most frequently used approach of computer-assisted content analysis is the dictionary-based approach. This approach requires an a priori developed dictionary defining all possible events. The dictionary is used to code texts according to the specified categories. The coding results can be used to identify the most frequently reported and a priori categorised events.

Another approach, the Statistical Association Approach, is based on consideration of co-occurrences of words. The co-occurrence of words in a text unit defines a matrix of similarities between words and this matrix can be further analysed by classification methods.

In the following, we will discuss the applicability of the two approaches and we will propose a procedure to identify the major events reported in newspapers during a specific time period combining the statistical association approach with a reference text technique.

## 2. The Dictionary-based Approach

Initially, we preferred the dictionary-based approach as the most often used approach in computer-assisted content analysis. The basis for this approach is a user-defined dictionary containing the definition of categories in form of word lists. Based on our knowledge about the dictionary-based approach, we do not recommend this approach to identify events for several reasons. The following two main reasons can be outlined as:

### **Time-consuming Development**

- The development of dictionaries is very time-consuming. Philipp Schrodtt (2001:2-7) mentions in his paper that it took nearly four years to develop and evaluate a dictionary to code international events in English texts.
- The dictionary has to be developed and validated for every language spoken in the countries participating in the ESS or all newspaper texts have to be translated to English before coding.

### **A Priori Development**

- The dictionary has to be developed a priori which means you have to know which events can possibly occur because you have to define categories in the dictionary.
- The dictionaries have to be updated every time a new coding phase starts because, for example, new events can occur and politicians change. Word lists to define new events have to be added as soon as these events happen.

### 3. Combination of a reference text technique with the Statistical Association Analysis

Recognising these problems, we decided to test a second approach of computer-assisted content analysis to identify major events. Some time ago, we discussed the statistical association approach as an alternative to the dictionary-based approach (Landmann & Zuell 2004). One result of this test was that the statistical association approach offers possibilities for an explorative analysis and the enormous time-consuming text pre-processing phase can be significantly reduced by lemmatisation and parsing routines. Regarding the aim of identifying events, the crucial advantage is that one does not need a priori defined categories, which means that such an approach could be very appropriate for finding events without too much previous knowledge about the text itself.

One major problem of this kind of analysis is how to differentiate between words which are indicators for events and words which are so-called meaningless words. Our assumptions are that a) major events are reported frequently in a specific time period and can be identified by frequently used words and b) the words used to describe the events are distinguishable from other words because they occur much more in the texts of a specific time period than in a larger text sample of general language usage. For our research question we use newspaper texts because newspapers are the medium in which societal events are reported typically.

Based on these assumptions we compare a reference text corpus composed of newspaper texts for a longer time period with a corpus of texts for a specific period in which we expect to discover events (the so-called event text corpus). Our assumption is that the reference text represents the typical vocabulary usage in newspapers and the event corpus contains specific event words for the selected time period.

### 4. The Procedure

In the following we will describe our procedure in a more detailed fashion. In general, the procedure can be described in four basic steps, starting with the composition of the reference text corpus, selecting the event text corpus, moving to the calculation of word frequencies and differences between the word frequencies, and concluding with the application of a statistical association analysis based on the word frequencies of selected words to identify the events.

In the first step, we determined the reference text corpus. For our tests we selected “The Guardian” and “The Times” as representatives of British coverage. For Germany, we selected “Süddeutsche Zeitung” and “Welt”. The decision for a specific newspaper does not seem really important for our analysis because we are looking only for outstanding events. We suppose that these events are reported in all newspapers as well as in television and broadcast independently from the political or cultural tendency of the medium. Here we emphasise that we are not interested in how something is reported but in what is reported. We collected all articles published in a two-year period (December 2004 to November 2006) in the sections of national and international news as well as all articles published on the first page of each newspaper edition of the selected newspapers. Finally, the reference text corpus for Great Britain consists of 102.949 articles composed of 38.994.210 words and the corpus for Germany consists of 56.845 articles composed of 16.942.771 words. We consider the words in these corpora as normal use of vocabulary in newspaper articles. The corpora also include the articles of the time period to be analysed. All texts were automatically lemmatised by the

programme TreeTagger (<http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>). Lemmatisation refers to the matching of all different forms of a word regardless of whether its root is the same, e.g. ‘say’ as well as ‘said’ share the same lemma. We assume that using the lemmata instead of the words will lead to clearer event groups.

In the second step, we prepared the event text corpus, containing newspaper texts over a specific time period where we expected to detect the major events. We illustrate the functionality of the procedure which we outlined in the previous section with an example. We were interested in events occurring at the beginning of the data collection phase of the ESS in September 2006. Our interest was focused on major events in Great Britain and Germany during the first week of September (week 36). As event text corpus, we selected the specific texts published in the above mentioned newspapers in the specific week. With these texts we expected to identify the major events which occurred in this week.

These two steps are the most relevant steps in our procedure because they establish the base for isolating words as indicators for events.

In the third step, we calculated word frequencies for all words of the reference text corpus and for all words of the event text corpus. Additionally, the relative frequency of each word was calculated as the proportion of the total number of words in the text.

Afterwards, we calculated the differences between the relative frequencies of the words of the reference text corpus and the frequencies of the words in the event text corpus. For our purposes we have to use relative differences because for words with higher relative frequencies higher differences are expected. To avoid this effect we calculated the relative differences as the portion of the total number of words in the reference text corpus.

Words were then sorted by relative differences and the words with the highest relative differences were used as event words in the further analysis following our assumption that these differences are indicators for specific events.

In our example we set up two more restrictions. Firstly, we removed all words that have very low frequencies, in our case smaller than 25, because of our assumption that major events are reported frequently in newspapers in a specific time period and can be identified by frequently used words. Moreover, we determined that we consider only words in our analysis which are found at least in two percent of all articles of the event text corpus. These restrictions are necessary because the importance and relevance of an event can be determined by the frequency of reporting. Additionally, we limited the number of words selected for the factor analysis to 30 words with the highest deviation from the reference text corpus and which, additionally, meet all other conditions. The specific cut-off points are somewhat arbitrary but can be based on some familiarity with the data and the decision on how many events should be handled as major events. Looser restrictions result in more events to be considered in the following analysis. The selected words can be found in table 1.

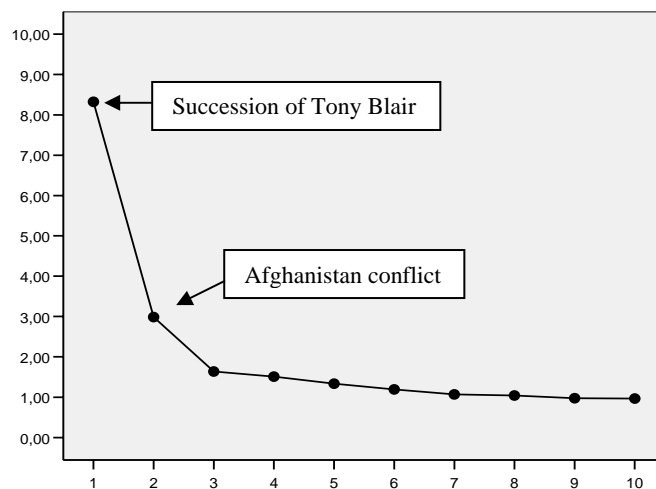
We applied an exploratory factor analysis to identify the latent semantic fields of the event words within the text under examination and to identify reported events. The goal is the representation of the latent semantic fields of the correlations of word frequencies. The factor analysis allows us to replace many more or less correlated variables by few independent factors without crucial information loss.

**Table 1: Selected Words**

Great Britain	Afghanistan, Blair, Blairite, Blairites, Brown, Campbell, Festival, Gordon, Johnson, Kabul, Laden, Milburn, NATO, Nimrod, Operation, PM, Science, Steve, War, Watson, Wright, aide, chancellor, code, departure, festival, leadership, quit, resignation, timetable
Germany	Kampusch, Seeblockade, Natascha, Lebensmittelkontrolle, Gammelfleisch, Schnappauf, Datei, Fleisch, Anforderung, Gesundheitsfonds, Bonn, Marine, Beirut, Seehofer, Schiff, Ware, Blockade, Labour, Küste, libanesisch, Libanon, Überwachung, Skandal, Terror, Taliban, kontrollieren, Anti, Gesundheitsreform, Fond, CIA  (Kampusch, sea blockade, Natascha, checking foodstuff, rotten meat, Schnappauf, file, meat, requirement, "health pool", Bonn, Navy, Beirut, Seehofer, ship, goods, blockade, Labour, coast, Lebanese, Lebanon, control, Scandal, terror, Taleban, control, anti, "health reform", pool, CIA)

Based on word frequencies, correlations were calculated and subsequently we performed an exploratory factor analysis using principle component as extracting method, and Varimax rotation. Our strict rules for selections of the event words (e.g. the variables for the analysis) and our relatively strong expectations concerning the factor patterns to be revealed in the text help us to interpret the factor results.

To limit the number of dimensions, we used the screeplots of Eigenvalues. For Great Britain the screeplot (fig. 1) indicates that the first two factors can be interpreted as indicators for events. These two factors (see table 2) explain the most amount of variance.

**Figure 1: Screeplot of the Eigenvalues (Great Britain)**

The first factor comprises the words "Blairite, Watson, Brown, Milburn, chancellor, Gordon, leadership, aide, Blair" and represents the event "succession of Tony Blair". Blair announced his resignation and ignited a discussion about his successor. Blair was not willing to give Brown the public endorsement he wanted as his successor.

The second factor is explained by the words "Nimrod, NATO, Kabul, Afghanistan, operation", which are indicators for the event "Afghanistan conflict". The continued discussion of the reinforcement of the troops of the NATO and the engagement of British

soldiers as well as the increasing number of killed soldiers are the main topics concerning this event.

**Table 2: Great Britain: Factor loadings (Varimax rotated) of the selected factors**

	1	2
Nimrod	-,026	<b>,539</b>
Blairites	,428	-,047
Blairite	<b>,812</b>	-,009
Watson	<b>,670</b>	-,010
NATO	-,025	<b>,859</b>
Kabul	-,043	<b>,801</b>
Brown	<b>,878</b>	-,031
departure	,460	-,046
timetable	,166	-,050
Milburn	<b>,542</b>	-,020
Afghanistan	-,031	<b>,852</b>
Laden	,005	-,013
chancellor	<b>,919</b>	-,007
Johnson	,349	-,040
PM	,360	-,038
Science	-,024	-,058
Wright	,073	,054
Gordon	<b>,779</b>	-,049
resignation	,434	-,041
Operation	-,033	<b>,728</b>
quit	,463	-,040
Steve	-,081	-,102
leadership	<b>,723</b>	-,061
aide	<b>,592</b>	-,031
Festival	-,030	-,060
Campbell	,097	-,062
Blair	<b>,840</b>	-,034
festival	-,060	-,069
War	-,088	-,106
code	,011	-,058

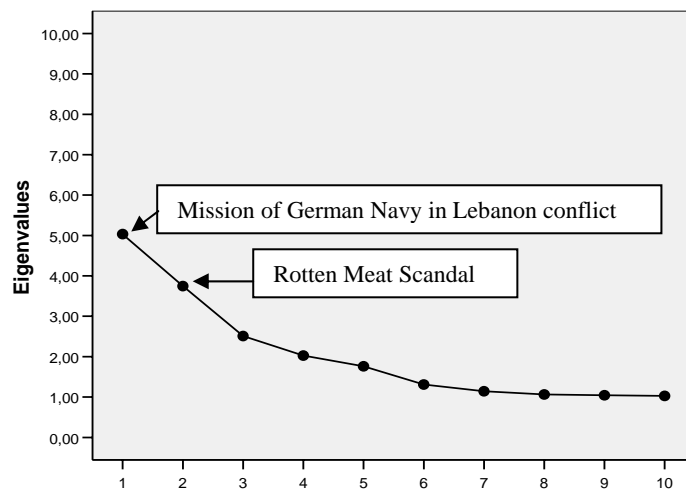
For Germany, too, the screeplot (fig. 2) indicates that two factors can be interpreted.

The first factor comprises the words "Seeblockade, Anforderung, Beirut, Blockade, libanesisch, Libanon" ("sea blockade, request, Beirut, blockade, Lebanese, and Lebanon") and represents the event "mission of German Navy in Lebanon conflict". The German Navy should secure the sea frontier in the Lebanon conflict as part of the NATO to support freedom. But the government of Lebanon formulated some conditions before allowing foreign troops in the country. Therefore, the German mission was delayed.

The second factor is explained by the words "Lebensmittelkontrolle, Gammelfleisch, Schnappauf, Fleisch, Seehofer, Ware, and Skandal" ("checking foodstuff, Rotten Meat, Schnappauf, Meat, Seehofer, Goods, Scandal), which are indicators for the event "Rotten Meat Scandal". In Germany, rotten meat was found in a wholesaler store in Bavaria. This resulted in a discussion between the German Minister of Agriculture Seehofer and the

Bavarian Minister for Environment, Health and Consumer Protection Schnappauf about the responsibility for and consequences of this scandal.

**Figure 2: Screeplot of the Eigenvalues (Germany)**



**Table 3: Germany: Factor loadings (Varimax rotated) for the selected factors**

	1	2
Kampusch	-,020	-,039
Seeblockade	<b>,802</b>	-,031
Natascha	-,033	-,045
Lebensmittelkontrolle	-,062	<b>,709</b>
Gammelfleisch	-,032	<b>,763</b>
Schnappauf	-,043	<b>,589</b>
Datei	-,017	-,004
Fleisch	-,029	<b>,678</b>
Anforderung	<b>,575</b>	-,029
Gesundheitsfonds	-,036	-,048
Bonn	-,041	-,088
Marine	,284	-,068
Beirut	<b>,904</b>	-,043
Seehofer	-,061	<b>,774</b>
Schiff	,239	-,057
Ware	-,005	<b>,635</b>
Blockade	<b>,714</b>	-,053
Labour	-,007	,031
Küste	,323	-,065
libanesisch	<b>,820</b>	-,059
Libanon	<b>,788</b>	-,084
Überwachung	,137	,145
Skandal	-,098	<b>,636</b>
Terror	-,092	-,148
Taliban	-,085	-,238
kontrollieren	,179	,238
Anti	-,049	-,067
Gesundheitsreform	-,056	-,080
Fond	-,032	-,040
CIA	-,064	-,164

## 5. Discussion

In conclusion, one can say that our procedure leads to good results when searching for major events in newspaper articles without too much pre-processing work done by humans. The advantage of such an approach is that

- it identifies events uniformly (in contrast to manual coding as described above),
- no knowledge about events is necessary in advance (no a priori categorization as necessary with a dictionary-based approach), and
- the number of events selected can be regulated by setting the analysis parameters more or less restrictively (number of words, frequency of words, etc.)

The procedure offers a systematic way to create the event data base for all countries participating in the ESS. Nevertheless, the short description of the different selected factors based on (automatically selected) newspaper articles remains an important task. Moreover, the decision about the presumed effects of an event on respondent behaviour remains: The decision of the effect of an event to be expected on respondent behaviour cannot be made with content analysis and not even by coders. In our opinion it has to be done by researchers, for example by those who developed the questionnaire and/or those who analyse the data.

Additionally, a further question remains. Texts were automatically lemmatised for our project. The lemmatisation routine works rule-based. This results in some words not combined to one root (for example, Blair, Blairite, Blairites). At the project start we assumed lemmatisation would be very important to get clearer event groups. After our tests, we propose to prove the necessity of lemmatisation. In our experiences the words with different word forms are always combined in one factor. For working without lemmatisation it could be helpful to change some of the analysis parameters (word frequencies, number of event words). Although programmes for lemmatisation for different languages are available, it is a lot of work to apply them to the amount of texts necessary for the reference text corpora. Further test are planned and necessary.

## References

- Landmann, Juliane, Zuell, Cornelia (2004): Computerunterstützte Inhaltsanalyse ohne Diktionär? Ein Praxistest. ZUMA-Nachrichten 54.  
[http://www.gesis.org/Publikationen/Zeitschriften/ZUMA\\_Nachrichten/documents/pdfs/54/09\\_Landmann.pdf](http://www.gesis.org/Publikationen/Zeitschriften/ZUMA_Nachrichten/documents/pdfs/54/09_Landmann.pdf)
- Schrodt, Philip A., Deborah J. Gerner (2001): Analyzing International Event Data.  
<http://web.ku.edu/keds/papers.dir/automated.html>
- Stoop, Ineke (2002): Context and Event data. Guidelines for National Coordinators.  
[http://www.scp.nl/users/stoop/ess\\_events/guidelines\\_events.htm](http://www.scp.nl/users/stoop/ess_events/guidelines_events.htm)
- Stoop, Ineke (2004): Event data collection Round 2. Guidelines for ESS National Coordinators  
[http://naticent02.uuhost.uk.uu.net/questionnaire/context\\_event/event\\_reporting\\_guidelines\\_round\\_2.pdf](http://naticent02.uuhost.uk.uu.net/questionnaire/context_event/event_reporting_guidelines_round_2.pdf)
- Stoop, Ineke (2006): Event Reporting Guidelines.  
[http://naticent02.uuhost.uk.uu.net/ess\\_docs/R3/Fieldwork/r3\\_event\\_reporting.pdf](http://naticent02.uuhost.uk.uu.net/ess_docs/R3/Fieldwork/r3_event_reporting.pdf)

## About the Authors

**Cornelia Zuell** and **Juliane Landmann** are members of the Text Analysis group at the Center for Survey Research and Methodology (ZUMA), Mannheim, Germany. Cornelia Zuell's interests include computer-assisted text analysis methodology, text analysis software and statistical issues. Juliane Landmann earned her doctorate in Social Science at the University of Mannheim. Her interests include computer-assisted text analysis methodology and issues of organized interests. Both can be contacted at GESIS-ZUMA, P.O. Box 122155, D-68072 Mannheim, Germany; tel. +49 621 1246147; fax +49 1246100; e-mail [zuell@zuma-mannheim.de](mailto:zuell@zuma-mannheim.de).





# Survey Research in a Wireless World

Aaron Pazurik & Mark Cameron

*Adapted from “The Future is Wireless”, as presented by Aaron Pazurik at the TTRA Europe Conference in Nice, France (April, 2007). A similar paper is published in the TTRA Europe Conference proceedings.*

## 1. Wireless 101

*Connectivity anytime, anywhere...*

Although the term “wireless” elicits thoughts of modern or futuristic technology, wireless communication is hardly a new concept. Wireless technology can be traced back more than a century to the earliest telegraphs and telephones. This type of “analog” communication has since been used for radio and television broadcasts, citizens’ band (CB) radio and early-generation cellular telephones.

Unlike its analog ancestors, modern wireless technology is used to move *data* – ranging from digital voice communications to internet access – providing many new opportunities and transforming the way people interact with one another. Perhaps nowhere is the potential impact of wireless connectivity more evident than tourism research, where mobile access to information presents both exciting new opportunities and challenges.

There are so many wireless technologies available today that it can be very confusing to most people. Wireless connectivity options can primarily be categorized as follows.

Proximity	Close (Personal Area Network)	Near (Local Area Network)	Far (Wide Area Network)
Technology	Bluetooth	Wifi	GSM or CDMA
Typical Usage	Mobile phone headset	Internet “Hotspot”	Mobile Phones
Usage Cost	None	None - low	Network only (pay-per-use or data plans)

Wireless communication has changed very rapidly in recent years. The most significant changes have been improvements to transfer speed, network coverage and availability of devices. This discussion is focused primarily on wide area wireless using data services (internet) where connection is truly “anywhere, anytime.”

## 2. Mobile Wireless Today

Nearly all phones sold and in-use today use a digital network and are capable of supporting digital services such as internet and text messaging. Original “cellular” phone networks using analog signals are being decommissioned and will eventually cease to be in use.

The growth in mobile phones in recent years is astounding. Mobile phone usage is dramatically increasing in developing countries such as China and India.

**Worldwide Mobile Phone Sales (Gartner, Inc)**

1997	2000	2006
107 million	414 million	990 million

Traditionally, phones have been classified as either a “mobile phone” or a “Smartphone”. The “Smartphone” is the convergence of personal digital assistants (PDA) with the mobile phone. While the term “Smartphone” is not clearly defined—a web search will turn up countless definitions with subtle differences—but it is most commonly used to refer to full-featured phones with an open Operating System (OS) allowing you to run third-party software programs much like you would with a personal computer.

### Adoption of Platforms and Standards

There are many major players in the market and as a result, standards adoption has been slow to take effect. There are a number of Operating Systems in the Smartphone race, including: Palm OS, Windows Mobile, Symbian OS, Linux and RIM’s Blackberry. Most other phones, while not running an open OS, have almost universally adopted the Java platform for application development. The mobile version of Java, called Java 2 Mobile Edition or J2ME, runs on virtually every mobile phone sold today, blurring the traditional line between Smartphones and so-called “Dumbphones”.

(Seems like the place to discuss SMS, EMS, WAP and internet but already have it below)

## 3. Mobile Wireless Tomorrow

### *Faster, Cheaper, Better!*

Wireless data speed has increased significantly in recent years and all indications are that it will continue to do so as new networks are installed. High-speed networks (3G, 1xEV-DO) are being installed throughout the world and permit data speeds over 300kbps. Next generation (4G) plans are already being devised and will permit data throughput over 100Mbps<sup>1</sup> – equivalent to a typical hard-wired corporate network today.

All indications are that wireless data access will become less cost prohibitive over time, primarily due to increased demand in a highly competitive market of wireless carriers. Another factor will be the availability of more powerful devices at lower costs. It is conceivable that mobile phones will one day be as cheap as calculators are today, yet capable of running sophisticated applications—like survey software! When compared to PCs, production of smaller devices is simply cheaper. Also, for data access, it is significantly more expensive to install a wired network (phone/cable/fibre) compared to a wireless network.

<sup>1</sup> Source: Wireless World Research Forum

This is one reason why wireless phone subscriptions have grown rapidly compared to wired network access in developing countries such as China and India – in fact, China mobile phone subscriptions surpassed land-line phone installations in Oct 2003<sup>2</sup>.

Wireless data accessibility (network coverage) has grown steadily, starting in metropolitan areas and expanding to smaller centers. While this increased the number of users, it has also more importantly brought more reliability to wireless data for true “anytime, anywhere” access.

#### 4. A History of Technology Adoption in Survey Research

Like most things, survey research methods have undergone a massive change over the past few decades. Prior to the 1950’s, virtually all research was conducted using paper and tabulated manually. Technology entered the research scene in 1951 when the world’s first mass-produced computer, the UNIVAC I, was adopted by the US Census Bureau<sup>3</sup>. Manual tabulation gave way to computerized data entry and automated tabulation, which was introduced on an even broader scale by SPSS in the late 1960’s and SAS in the early 1970’s.

Throughout the 1970’s and 1980’s, Computer-Assisted Telephone Interviewing (CATI) and Computer-Assisted Personal Interviewing (CAPI) systems were developed to provide both questionnaire content and data collection in one system. CAPI took a strong hold in Europe, where high population density and expensive telephone communication made face-to-face research more attractive than telephone interviewing; at the same time, CATI dominated the North American research scene due to inexpensive telecommunications and sparse population density. CAPI and CATI continue to be heavily used today, but traditional approaches to both are threatened by new communication infrastructures.

The Internet and mobile telephones have charted a parallel path of adoption since the mid 1990’s, reaching mainstream status and thoroughly disrupting just about every traditional process. Web-based surveys have quickly emerged as a mainstream research method, and they have also provided both tremendous new opportunities and significant challenges for telephone and personal interviews. And unlike previous generations of research technology, mass adoption of Internet-based research has taken place practically overnight.

Mobile Computer-Assisted Personal Interviewing (MCAPI) became a viable method in the early 1990’s with the advent of handheld computers. Handheld computers were an ideal solution for many face-to-face interviews, providing a low-cost alternative to laptops or tablet PCs and significant efficiency gains over paper-and-pencil collection. But in many ways MCAPI, even during the heights of the “PalmPilot revolution”, was overshadowed by the emergence of the Internet. In hindsight the slow adoption of MCAPI is understandable, as a much greater focus was being placed on the communication infrastructure—the Internet and wireless telecommunications—which would ultimately converge to provide a truly mobile method for inter-personal communication. MCAPI is now, after more than 15 years of use, considered a mainstream replacement for existing paper-and-pencil collection in traditional research methods (e.g. face-to-face interviews, diary studies and mystery shopping).

<sup>2</sup> Source: China Ministry of Information Industry

<sup>3</sup> Source: Wikipedia ([http://en.wikipedia.org/wiki/UNIVAC\\_I](http://en.wikipedia.org/wiki/UNIVAC_I))

## 5. The Current State of Mobile Research Technology

Before we look to the future, let's take a look at what is available today for mobile wireless technology and how it is used by the Research industry.

### Messaging Systems

Short Message Service (SMS) was the first wireless messaging platform widely adopted around the world by nearly all phone manufacturers and networks. SMS is based on a system of sending short (160 characters or less) text messages to a mobile device. The messages are stored until the user reads the message similar to voice mail. SMS in market research is done by sending a user a question or series of questions (separate messages) to which the user is able to respond and in effect "answer" the question using designated text codes or phrases. SMS is ideally suited to very short surveys such as comment cards or exit polls, due to its limited content length and back-and-forth messaging system. Cost is another limiting factor with SMS messages often charged independently of phone minutes or Internet usage.

Enhanced Messaging Service (EMS) is an enhancement to SMS where the message can contain special text formatting (bold, italic, etc), animations, pictures, icons, sound effects and special ring tones. EMS is able to provide a richer experience but still retains the same limitations of SMS regarding cost and back-and-forth process.

Multimedia Messaging Service (MMS) is the evolution of SMS and EMS which allows messages to contain multimedia objects such as images, audio, video and rich text. For the average SMS user, MMS is able to provide a richer experience, such as instantly sending pictures to a friend. MMS has some significant differences, however; technically, the actual MMS message is delivered through an HTTP request similar to browsing a web site. MMS shares many of the same challenges as SMS and EMS, as well as additional ones; for example, not all messages are compatible between different devices, similar to the way difference web browsers display content in different ways.

### Internet Web Based Systems

Wireless Application Protocol (WAP) was introduced into mobile devices as a way of providing interactive experience for users when devices were not able to display and download full desktop designed web sites. WAP uses Wireless Markup Language (WML) which is specially formatted and compact web content designed for mobile WAP browsers. Although WAP was widely adopted by most device manufacturers, it never saw the same take-up as SMS presumably due to the fact that web content providers needed to make huge changes in order to deliver content to these mobile devices. As devices have become more powerful and wireless data speeds increase, most agree that WAP will ultimately be overtaken by the mobile web browser.

Mobile web browsers, available on a growing number of mobile phones, are improving every day. Data speeds and faster processors have made it feasible to conduct mobile web surveys. However, web applications are not persistent, meaning that they generally rely on constant access to the source server typically for each screen display (question), which means they rely heavily on a reliable wireless connection. Unfortunately, slow or lost connections are still all too common.

One challenge that is common to both text messaging systems and mobile web browsers is that they are "server-centric". This means that most of the processing for all users is done on a

server. While this model works well for applications that provide field users with access to a central database—like inventory and customer data—a server-centric approach is not ideal for data collection, where most of the data is flowing *into* a central repository. This principle could also be applied to Internet surveys on the desktop, but it is far more evident in a mobile setting.

### Disconnected or Limited-Connection

As mentioned previously, Personal Digital Assistants (PDA) have been used for face-to-face and self-completed research since at least the early 1990's, but they are only now reaching mainstream status. The convergence of PDA's with wireless communications and the Internet has enabled MCAPI research to utilize Smartphones—providing truly wireless surveys that can run in a disconnected mode, but connect as often as required to submit response data and retrieve new or modified questionnaires. These wireless survey tools are now as powerful as web and CAPI/CATI systems, allowing complex routing logic and even management of sample quotas. However, as with CAPI/CATI systems, PDA or Smartphone hardware must be capitalized, which can be justified for heavy use but is cost-prohibitive for most types of “on-demand” research, i.e. self-completed surveys of the general public.

Tablet computers have been used extensively for CAPI since the mid 1990's. They provide great functionality and are ideal for many face-to-face research situations, particularly surveys with multimedia content. Unfortunately, tablets have never reached a critical mass in the general marketplace, and as such remain very expensive compared to laptops and PDAs.

Ultra-Portable Personal Computers (UMPC) are a new arrival on the scene, attempting to bridge the gap between large, expensive tablets and smaller PDAs. It remains to be seen whether these devices will be widely adopted or play a more niche role as a complement to PDAs and full-sized tablets.

## 6. The Future is Wireless...

We believe that mobile computing will generate a disruptive change in the way that businesses communicate with their customers, forcing us to question many commonly held research practices. Even “new” research methods, such as current web-based surveys, will be transformed by the emergence of ubiquitous wireless communication. While the desktop Internet will most likely continue to grow, it is conceivable that the wireless Internet will become even more widespread, with just about everybody holding some form of digital connectivity in their purse or their pocket.

Timely Point-of-experience (POE) research is a major emerging opportunity. Point-of-experience research means reaching people when and where their feedback is most relevant: in a hotel, on a beach, at a restaurant or an event. The key challenge for POE research is proper sampling. If the invitation process is well managed, this method can combine the benefits of face-to-face interviewing (high response rates and relevance of data) with the cost-effectiveness of web research.

### Benefits:

The key benefit of mobile Point-of-experience research is relevance: providing more accurate results due to less dependence on recollection – i.e. feedback that is “in the experience” instead of “in the mind”. And instant results will enable organizations to immediately turn feedback into action. Survey tools that are easy to use for both the respondent and the

research buyer, if tied to proper sampling and incentives, can open researchers to a whole new world or opportunities.

### Challenges:

In a world where you can reach anyone, anytime, the greatest challenge for researchers may be over-sampling. If every hotel, restaurant and theme park is asking people to respond to surveys everywhere they go, respondents will grow as weary of POE research as they currently are of telephone and web-based research.

Incentives and sampling strategy will be important. One possible remedy is community-based or industry-based collaboration, e.g. a research program offered by a regional government or association to its members rather than separately deployed by each member.

The rate of adoption for new technology will also play a significant role, which will depend heavily on demographic and geographic factors. For example, younger generations are more likely to embrace the latest and greatest technology, and preferences for devices and communication methods will continue to differ greatly throughout the world.

To be effective for “on-demand” research, technology must be standardized across a broad range of devices – e.g. many different mobile phones, Smartphones, and Ultra-Mobile PCs. Mobile technology will only be truly effective for widespread research use when survey tools can seamlessly target the vast majority of people, i.e. when mobile surveys are as ubiquitous as mobile phones are today.

It is impossible to predict the future, but the disruptive effect of mobile technology on survey research seems inevitable. What remains to be seen is exactly how that disruption will unfold, and how long will it take to do so.

### About the Authors

Aaron Pazurik, BA, is the Chief Technology Officer of Techneos Systems, Inc., located in Vancouver, Canada. Aaron has been the chief architect and head of development for Techneos since 1999, and developed one of the first wireless survey platforms in 2001. Mark Cameron, BSc, is President and CEO of Techneos, having co-founded the company in 1995. Mark is a pioneer in the field of Mobile Computer Assisted Interviewing using handheld computers; he has been developing and promoting handheld survey software since 1991, dating back to his earliest exposure to survey technology in the Socio-Economic Research Division of Parks Canada.

## Exploring the Practicalities of Developing Web 2.0 Applications in Online Research

AJ Johnson, Jane Mills, Graham Davies

### Abstract

As an industry we continue to struggle to find respondents to provide us with survey information vital to the success of our industry. To survive the Internet revolution we must use innovative methods to motivate and engage respondents and maintain their long term loyalty. Over the years much has been written about intrinsically motivating respondents through survey design, communication and the use of incentives, however, researchers are still looking towards technology to help improve or even solve the problem. Web 2.0 is the latest 'buzz' word in market research that promises to make the online research more dynamic and interactive. It offers the potential not only to increase motivation, but also to provide new levels of insight from our collected data.

This paper explores the relationship between Web 2.0 and market research using a hands-on approach. Definitions of Web 2.0 are varied and often inconsistent and so we focus on how increasing respondents' interactive communication and collaboration would affect our current models of online market research. The increased availability of broadband Internet and widespread acceptance of social networking among the population lead us to suggest that Web 2.0 may not be a passing fad. Indeed, Ipsos MORI has long been aware of the benefits of community interaction and has been putting these into practise, where appropriate, for many years.

To really enter the world of Web 2.0 you need to hand over control of content and direction to your user community. Our aim was to explore these new approaches in a fast and cost effective manner so we designed and implemented an experimental Web 2.0 research application using freely available open source tools. We used the Drupal content management system as a framework and created an application that allowed users to submit their opinions in the form of statements that others could rate or comment on. Using this approach, rather than using shrink-wrapped vendor solutions or writing the application from scratch, we learned a great deal about the principles behind Web 2.0 technology. That said, we remain keen to work with software vendors in this area and present a summary of our discussions about future Web 2.0 functionality. The site, nicknamed 'The Opinionator', emphasized community interaction and was well received by our panel of testers.

### Keywords

Web 2.0, Research 2.0, Online market research, Free Software, Collaboration



## 1. Introduction

Market researchers have always strived to use the latest technology to improve all aspects of the survey process from data collection to analysis. In some areas, such as online surveys and panels, we have been quick to adopt new techniques and move into new areas; in others, such as Web 2.0, we are in danger of lagging behind. That said, the sheer speed and frequency with which new techniques and technologies are emerging from the Internet is baffling and keeping up with the latest advances is not trivial. Clearly a balance must be struck between blindly adopting every new technique that happens to be hip and stubbornly ignoring possibly revolutionary technologies.

The term 'Web 2.0' has recently entered the market researcher's vocabulary and rumours promise everything from improvements in respondent activity, thanks to intrinsic motivation; to greater insights from research projects, thanks to a blend of qualitative and quantitative research. What's more, all this supposedly comes in at almost no extra cost! If the rumours are true, Web 2.0 has the potential to help solve two key challenges long faced by our industry:

- How do we engage and motivate our respondents to promote increased response rates?
- How do we gain a greater insight from our respondents without increasing costs?

It is little wonder then that the concept of Web 2.0 is currently under the spotlight, offering an industry that still uses some very old methods, the chance to be right at the cutting edge of this new technology. In the last year, discussion about the application of Web 2.0 to the market research industry has become ubiquitous with publication of numerous articles and conference papers by prominent researchers. It is so popular in fact that the term 'Research 2.0' has even been coined to indicate the fusion of Web 2.0 techniques with market research methods. Despite this, the relationship remains in its infancy and little has been done to define its boundaries or implement its concepts on a large scale. The challenge for Research 2.0 is to demonstrate its worth in a real world environment by providing clients with a deeper understanding of the market.

This paper will explore the relationship between market research and Web 2.0, focusing on areas that can be put into practice in the short term and paying particular attention to three key areas:

1. Web 2.0 methods and technologies that can be adapted to market research
2. The plans market research software suppliers have with respect to Web 2.0
3. How well a Web 2.0 market research application succeeds in real world tests

## 2. What is Web 2.0?

The term 'Web 2.0' was introduced as a new concept at a conference brainstorming session between Tim O'Reilly and MediaLive in 2001. In this initial discussion the concepts were exemplified by sites such as Flickr, Wikipedia and the blogosphere in general, where content is added and maintained by site visitors. This contrasts with the traditional 'Web 1.0' model where content is published by site owners in a 'read only' state and is merely consumed by site visitors. Many authors have attempted to pin down an exact definition of Web 2.0, but no

two explanations are the same and as such Web 2.0 remains a nebulous set of concepts and ideas without rigid boundaries and liable to change without warning. The lack of clear definition goes a long way to explain the confusion that surrounds Web 2.0 but throughout this paper we will consider Web 2.0 as a set of methods and techniques that encourage user participation and collaboration and result in the formation of communities that generate content or otherwise add value to a site.

### 3. Market Research and Web 2.0

The potential uses of Web 2.0 principles in the market research sector have been explored by several distinguished authors in the past year and readers are encouraged to explore papers presented at recent MRS and ESOMAR conferences and elsewhere. For example, Ray Poynter 2006 outlined a possible evolutionary path from the 'Research 1.0' world, in which researchers have a parent to child relationship; to a 'Research 2.0' world, where the relationship is more adult to adult and interactive.

Despite all the excitement surrounding Web 2.0, its uptake by market research companies has been less impressive. Demand from clients for Web 2.0 features has also been fairly minimal so far in our experience. That said, several specialist research agencies, including Brain Juicer and Virtual Surveys, have openly embraced Web 2.0 concepts and launched surveys that promote greater user interaction. Brain Juicer, for example, is using concept testing via collaboration to get more thoughtful responses from its respondents. The larger mixed model research agencies, however, appear to be more cautious and sceptical on the advantages of Web 2.0 in market research. Although there is general acknowledgement of the potential benefits of Web 2.0 techniques, many emphasise that solid advantages over traditional market research are yet to be demonstrated. As GFK note on their website, "We explore the theoretical possibilities [of Web 2.0 in market research], but we experiment hands-on with the technologies and gain first hand experience of their strengths and weaknesses. So we maintain a realistic and open-minded position – particularly about the here and now".

### 4. Why the time might be right for Web 2.0 in market research

Several factors have probably led to the recent dramatic rise in interest in Web 2.0, and other Internet-based techniques, within the market research industry. For one, companies have started to build up large panels of respondents willing to carry out online surveys. With methods already in place to gather personal information, provide incentives for responses and recruit new members, it is tempting to start exploring ways to better utilize the potential of these embryonic 'research communities'. Other key characteristics in place to aid the success of Web 2.0 include:

- **Mass adoption of broadband Internet** – The adoption of broadband technology has risen dramatically over the past few years; Technology Tracker 2006 (<http://www.ipsos-mori.com/technology/pdf/statistics.pdf>) states that half of all adults in Britain were using broadband at home by the end of 2006. Broadband connections are mostly 'always on', have very large download limits and are considerably faster in comparison to dial-up. Together these features allow researchers to make more engaging surveys that take longer to complete and have images, media and other features to greatly improve their look and feel.

- **Mass adoption of social networking sites** – Ofcom 2006 note that ‘Over two in five adults with Internet access had used social networking websites for keeping up with friends’ and 70% of 16-45-year-olds have used these types of websites’. This broadening of the demographics of social networking sites should provide the range of responses required by researchers.
- **Integration of the Internet into daily life** – Checking email, browsing websites and keeping up with your online social network has become part of normal daily life. Indeed according to our own internal omnibus survey 40% of those aged 15+ in the UK claim to access the internet roughly once a day.
- **Opening of market research software** - The evolution of market research software towards a more open source model over the last decade will enable researchers to plug Web 2.0 techniques and concepts into existing software without using vast resources.

## 5. Web 2.0 at Ipsos MORI

With more than a decade’s experience of using the Internet to carry out research, Ipsos MORI has a strong pedigree of using online methodologies to deliver better solutions to its clients and we are always on the look out for new techniques. We believe that it is extremely important to move beyond the basic migration of off-line projects to the online environment, to a point where we fully understanding the new opportunities that the internet platform offers.

We are no strangers to using Web 2.0 techniques in the research process and with the support of our ‘Qualitative Hothouse’ Ipsos MORI has been conducting online qualitative research for more than five years. Many of our research projects have benefited from online discussions between normally hard to reach groups. In addition, our ‘Participation Unit’ embraces the opportunity for members of the community to take on the role of researchers in their own environment to provide unique insights otherwise inaccessible to market researchers. Therefore we actively encourage our respondents and communities to collaborate and bring their own ideas to our research projects.

Community interaction and other Web 2.0 techniques allow many of the exciting and beneficial qualitative techniques to be used in the quantitative environment by engaging the respondent to think outside the box rather than ticking one. This is a clear benefit but what is still up for discussion is the level of impact that Web 2.0 will have on mainstream market research. Outside our industry, Web 2.0 is showing real benefits in a number of areas and is being used by an increasing number of people each day. If market researchers are to join the tide of Web 2.0 adoption they may face several challenges. As we see it, researchers and clients need to assess if they are able to:

- Accept a fundamental shift away from traditional parent-child relationships to more parent-parent relationships (Pete Comely 2006).
- Confidently undertake research without tight control of a project's direction.
- Shift focus away from designing questions and collecting data, to providing added value through interpreting large amounts of less structured data from multiple sources.

The belief of the authors is that the movement towards Web 2.0 in market research will certainly not replace traditional research methods but will have a considerable impact on our business, and will blur the lines between quantitative and qualitative research. Some clients will push to adopt the techniques early to gain added value and be seen to be innovative. Others will be more conservative, waiting to see what benefits there are and move more slowly once the techniques become more acceptable. Ipsos MORI will naturally steer clients through this learning curve and maximise the benefits.

To gain further insights into using research 2.0 in the real world we set out to design and create an experimental market research web application that would bring some of the more radical concepts of community-based research into reality for our research teams. The design and concepts behind our application are not meant to be unique within the market research industry but has been specifically designed to combine Web 2.0 techniques with some of our more traditional opinion research. Retaining familiarity would make it easier for our research teams to see its potential for our clients. We believe evolution rather than revolution is the way to migrate towards these new techniques. The rest of this paper looks at what happened when we created a collaborative tool to rate opinion on a range of social and consumer issues. We will cover what we did, how we did it, the learning points and the reaction it received from a group of in house testers.

## 6. Designing and building a Web 2.0 application for market research

### Designing the Concept

The application, nicknamed 'The Opinionator', was designed to display opinions, sorted by topic, in the form of statements that visitors could rate based on their agreement or disagreement with the statement. For example, the front page might include the statement "Street crime is on the rise in my local town", with a score to indicate how many visitors agreed with this statement. A table showing the highest and lowest ranking opinions would be displayed to incite visitors to become members and vote for their preferred opinions or submit their own statements for others to rate. We hoped that inflammatory or biased statements left by users would promote discussion and encourage the addition of comments and competing opinions. However, to prevent publication of offensive material or repetitive statements we decided to pass opinions through a team of researchers to act as moderators or editors before publication. Members would be rewarded using a points system based on how many of their comments or opinions were published, edited or rejected. We wanted to keep the overall look and feel of the site similar to existing online surveys but wanted an AJAX-based voting method to promote site browsing and eliminate the need for page submission and reload. Other desirable features included an RSS feed of new opinions and animated displays of recent voting and submission activity.

### Deciding on the Technology

Research agencies have traditionally looked towards survey software suppliers to provide the technology to carry out data collection and delivery. Some believe that these suppliers are often too slow at keeping up to date with new technology and slow to release new features. Others would argue that the blame lies with the research agencies that spend little on survey technology in comparison to other parts of their business. Both have an element of truth and the result can often be a 'chicken and egg' situation.

To decide how to develop our application we first looked at what off the shelf technology was available from our survey software suppliers. We approached seven of the major companies to examine

- What they understood by the concept of Web 2.0
- What impact they thought it would have
- Whether their software currently possessed Web 2.0 features
- What plans they had to introduce Web 2.0 features

Six out of the seven suppliers responded. Of those who did respond, they expressed some very interesting views on these issues. In summary all are very aware of Web 2.0, all claim to have existing Web 2.0 features and all are working on introducing further features within 3-12 months. However, most are waiting to fully assess its impact before making any strong commitments to large scale development. The word 'hype' was used by half of those who responded. Above all, most are not currently receiving regular demands from their clients to develop Web 2.0 features **as a priority**.

There was some difference around their definitions of Web 2.0 and how they felt it will be developed for market research. This fits with our belief that Web 2.0 is a dynamic term that has no clear definition. Most suggested that it was usability enhancements to the survey experience through Ajax type technologies, the development of portals and developing Web 2.0 technologies to improve the delivery of data. However, not all picked up on the concept of collaboration and user generated content which we see as the core to our prototype application. For this reason we decided that the mainstream survey software suppliers could not currently help us for our proposed application.

With no out of the box solution we had to explore other ways to setup our prototype site. The most flexible option would be to design and write the entire application from scratch. This option would probably be preferred in the longer term, but for this demonstration there simply wasn't the time or budget to take this route. We briefly looked at web application frameworks, like Ruby On Rails, but decided that the learning curve and development time for these was too great. Our next choice was to explore the world of open source software to see what could be adapted to our needs. Unfortunately, most high profile Web 2.0 community sites do not make their code available for download and although there are some clone projects (e.g. pligg.com is attempting to clone the popular news site digg.com), these tended not to be very mature.

#### Which did we choose and why?

We investigated several open source solutions but finally settled on a highly flexible content management system called Drupal, available at [www.drupal.org](http://www.drupal.org), as the basis of our site. Drupal is written by a bright and active developer community in PHP and uses a MySQL database to store data and settings. It has a compact core of modules that perform basic content and user management tasks and has a vast array of contributed modules that can extend its functionality. Recent versions also include a lightweight JavaScript library (jQuery) that provides AJAX functionality.

### How did we develop the research application?

Drupal, as with many open source web applications, is written in PHP and uses a MySQL database to store data and settings. The system requirements recommend using a machine running the GNU/Linux operating system with the Apache web server. This presented us with our first stumbling block since most in house servers use Microsoft Windows and IIS. Although this setup is supported, it is not widely used or tested and therefore may not be as stable. Once a suitable machine was found, installation of Drupal using the online documentation was trivial and the base system provided many of our required features. The core user management system provided simple registration and email support and the profile module allowed us to ask for additional demographic information during registration. We tweaked the registration process to jump straight to a guided tour of the site where members were presented with opinions submitted by other users and asked to rate each based on a five point scale where 1 star corresponded to complete disagreement and 5 stars represented complete agreement. The current average score was displayed next to each opinion to see if this would influence the voting behaviour of users. The voting system was set up using a set of contributed modules with an API, AJAX-enabled star rating system that required no page reloading, and a module to display lists of opinions based on their score. Once users had rated existing opinion they were asked to submit their own opinions on one of the suggested topics. These new opinions were immediately available for others to rate but an email was sent to a moderator in case inappropriate content had been added. A user ranking system was employed to award points to users for submitting their opinions, comments or ratings and a table of the highest ranking users was displayed to try and encourage further contributions. Opinions were divided into topics and for the demo these were limited to crime and the environment but in the real world test we would imagine that users would be able to submit opinions on any topic.

### How did the final version compare to the specification laid out at the beginning (did we meet our objectives?)

It is a testament to the flexibility and power of Drupal that we were able to incorporate most of the features outlined in our specification into the site with almost no PHP programming. Most configuration and maintenance can be done entirely through the web browser using a granular access control system to strictly control access to certain areas of the site. The exception to this was the theming, which required access to the file system and some knowledge of HTML and CSS to get the site to look similar to existing online surveys. Time restrictions meant we were unable to include any animated graphical result displays. Given more time we would also have liked to allow members to vote on the comments of other users and have this tie in with the points system.

### What challenges did we face?

In general, open source and the Drupal system proved to be a very effective way to rapidly setup a site to explore some of the ideas behind of Web 2.0. The core code that runs Drupal is very mature and has been tested by thousands of developers and end users, thus massively reducing development time. However, we had to abandon some of our desired features due to poorly maintained contributed modules that were not quite production ready. The stability, security and scalability of the demo site would also have to be rigorously tested before releasing to a larger audience. A major challenge came not from the software itself but from

company policies on supported programming languages and software vendors. It may prove costly to retrain in house developers in the required languages and skills. The effect of open source licenses, like the GNU Public License, on future development must also be investigated before proceeding further. In the longer term it may be advantageous to rewrite prototypes from scratch in house to meet exact business needs rather than try and adapt third party solutions. Although considerably more time consuming, this allows in house developers to retain control of patches and updates without external delays.

Our original thoughts on this application included many grand ideas for completely opening online market research, allowing evolution of ideas through collaboration and using advanced graphical displays to wow users and researchers alike. Budget restraints have meant that we had to cut back a little to produce a working demo on a very short time-scale. That said, we did achieve our revised goal of getting respondents more involved in the process of gauging other people's opinion on a particular topic. Within our site we explored whether the voting behaviour of our users would be influenced by the ability to view the aggregated real-time results of others.

### What have we learned?

Our experiments using open source software to rapidly develop a prototype community-based market research application have proved very useful. We have learned that by using the right open source project as a base, one can go a long way towards a specified goal without the need for extensive programming. We remain cautious about moving forward with this type of solution into a production environment and are keen to work with software vendors to guide their future development in this area. Our efforts to motivate respondents to join our community were not entirely successful in the trial and we suspect that email or SMS alerts promising monetary rewards for new votes or comments may be necessary to spur members into activity. Future prototypes will be sure to incorporate ideas for improvements from company executives and clients with the aim of gradually evolving a system that will make the best use of our core research skills whilst fostering a sense of community involvement and engagement.

### Feedback on our application

To evaluate the potential of using the Opinionator for market research purposes we conducted an online survey to ask the users their opinions of the tool, its relevance to market research and ways to improve for future development. As the experiment only ran for one week we wanted to gather an overall understanding of their experiences within the set time frame and ways to enhance the tool for further use. It was decided to test the Opinionator with market research professionals first to understand their stance on the tool before inviting our panels to take part. This was to ensure that the research executives saw a suitable research angle to the need technology and were able to see if their clients would be interested in the tool.

As part of the experiment to test the newly created site we invited 20 market researchers from Ipsos MORI in the UK and Ipsos North America to go to the site and experience the new tool as a potential user would. With no previous explanation of the experiment just an email invitation, as a panelist would normally receive. Of those who were invited all logged in and viewed the site and more than two thirds contributed comments to the application. The success of the users actively taking part was the addition of instructions for those who were not computer savvy and were not sure as to where to go and what to do.

The feedback from the users was positive with researchers commenting on how great the idea was and thought it was an interesting concept to be able to create questions as well as answer others. Also most users agreed that the Opinionator was a good tool in engaging what the public thought on a number of different topics.

Furthermore it is interesting to note that two thirds of respondents were not influenced by the results of the polls, and did not feel obliged to vote with the majority. This is captivating in itself as there was a fear in whether respondents would be influenced into voting for what most users had agreed with and voted for. However we were pleased to find out that this was not the case, in fact the contrary happened with users openly debating issues they felt strongly about and posting comments to oppose the comments already posted by others. "I thoroughly enjoyed seeing how my comments had been rated and if people had other views to broaden my view on the crime topics".

The research also proved a useful tool in working out the tools weaknesses and ways to improve it for future users. The participants thought the home page was confusing and they were not sure as to where to go first. "At first glance of the homepage though, I wasn't sure where to go - whether to go to the left or the right...the 'continue' or 'click here to start' buttons were a little hard to see". Even though there were instructions we need to change these and make them a lot clear in version 2. This is so that the users are precisely aware of where they are meant to go, and what to do and the process of navigating the site.

The aim of the experiment was to test the suitability of Web 2.0 technologies in market research and we have been productive in defining some of the vast Web 2.0 technologies and looking at their suitability for market research. Although we only looked at a few of the Web 2.0 tools we have made proactive progress on experimenting with the technology and not theory. The future potential of this project is vast. The experiment has brought up a number of new opportunities to explore, with the introduction of RRS feeds, email updates on new content, giving the user the opportunity to rate the comment and author and the introduction of blogs to the community site. As the potential is so vast we will be concentrating on further versions of the Opinionator to integrate these new and exciting opportunities in our quest to make Web 2.0 suitable and usable in the market research industry.

## 7. Conclusion

It would be an understatement to say that we learned a lot from this experiment as a team. Instead of being bogged down with the magnitude of Web 2.0, its definition and the literature on the matter we have focused solely on the key concepts and their impact on the market research industry. By actively experimenting with the free and available software we have produced a new Web 2.0 tool that can be used to gather market research on a wide range of topics and has the potential to shape new and improved products in the online industry.

Web 2.0 technology has opened a window of opportunity to researchers and we now wait to see how wide this opening becomes. Beyond the hype of Web 2.0 are some clear benefits that will be undoubtedly developed by our industry. It remains paramount to ensure that the high quality demanded by our clients is fulfilled by maximizing the use of new technology whilst being careful not to develop technology for technology sake. Web 2.0 is not a revolutionary new research method but rather a complimentary technique in the researcher's toolkit. The fast increase in the use of social networking sites and blogs cannot be ignored by the research



industry; it should instead be looked at as an exciting opportunity to prepare market research for the decades to come.

## References

GfK Website, March 2007

<http://www.gfknop.com/customresearch-uk/expertise/online/web2/index.en.html>).

Ipsos MORI, Technology tracker, December 2006

Ofcom, Communications Market 2006 <http://www.ofcom.org.uk/research/cm/cm06/main.pdf>

Pete Comley, The games we play, ESOMAR Panel research Conference, Barcelona  
November 2006

Ray Poynter, You & 2.0, MRS Research Magazine, December 2006

Tim O'Reilly, What is Web 2.0, Design patterns and business models for the next generation  
of software, <http://oreilly.com>, September 2005

## About the Authors

**AJ Johnson** Director, Head of Online Research Ipsos MORI

Is responsible for all online data collection and the development of online research at Ipsos MORI. Located in London, AJ works within the industry to raise the profile of technology within the market research industry. This includes committee membership of the Association for Survey Computing as well as chairing the MRS-sponsored Technology Group and judging the annual Technology Effectiveness Award.

**Jane Mills** Online Executive, Ipsos MORI

Is a specialist in online research and is responsible for developing all areas of online research within Ipsos MORI. Located in London Jane has a BA honours in Business Studies and is a member of the MRS.

**Graham Davies**

Is an independent IT consultant located in Manchester, England. Graham has a PhD in Molecular Biology and specializes in researching technological and IT-based solutions to real world problems.

# A new era of Market Research – Real-Time Sampling™ (RTS)

Hugh Davis

## 1. Introduction

In recent years, the Internet has been transformed from a Web surfer's paradise to the world's business district. Today, for much of the world's working population, the Web is now associated with getting their job done. Checking e-mail has become somewhat of a chore, and due to the e-mail volume most people now receive, it has to be done daily, if not hourly, and it takes a lot of time. Rather than strolling leisurely through cyberspace, happily taking surveys to which they are invited, survey respondents have begun to be more selective and it is even more challenging to attract and retain panel respondents and keep high levels of respondent cooperation.

So, as it becomes harder and harder to stand out as a compelling choice for consumers online, market researchers are scrambling to make surveys more appealing, adding one glitzy feature after another to entice consumers to devote a few minutes of their valuable and limited time to their online surveys. In fact, market researchers rely on people to volunteer their time for little to no compensation, amidst a myriad of other online digital distractions.

The result? We're entering a new era of market research. Marketing research professionals understand that they need to provide a "value proposition" beyond the traditional survey and the traditional online research panel to cut through the clutter and engage survey respondents.

Moving beyond the panel, there are other types of respondents – individuals who would be willing to take an online survey, but may not be willing to join a marketing research panel. Greenfield Online is offering market researchers broader access to alternate respondent sources by tapping into a rich base of consumers willing to participate in online surveys and introduced Real-Time Sampling™.

Real-Time Sampling™ (RTS) expands the respondent access capability by reaching out in real-time to potential survey takers who may not choose to join an online panel. These respondents are recruited via a Global Survey Network of pre-recruited partners and interested respondents are screened provided with a relevant survey opportunity assigned by a proprietary Survey Router. This Survey Router intelligently assigns respondents to relevant surveys based on demographic data. Further, the Survey Router synthesizes multiple sample and recruitment sources to ensure high-quality data.

## History

Before Greenfield Online successfully launched its Real-Time Sampling™ other research companies have tried as well to release a similar solution in the past. It didn't work for a couple of reasons. First of all the technology was not sophisticated enough to support such an innovative product. Secondly the economic model was based on payment for traffic which was making the cost per survey too high.

## 2. Background & Objectives

Greenfield Online has developed an approach to fulfill the sample requirements of researchers that differs fundamentally from existing, conventional online Panel methods. As mentioned above, this approach -- Real-Time Sampling -- consists of recruiting individuals in real time from a network of websites with which Greenfield Online has developed referral relationships.

This methodology is innovative in that it leverages the many potential survey takers online who are willing to take a single survey, but who may not necessarily want to join a market research Panel. Our research has shown that real-time respondents also report participation in marketing research studies, and many have joined/are members of online panels.

Potential Real-Time Sampling survey takers are solicited to participate while they are surfing the Internet and are on the websites of Greenfield Online's network of affiliated recruiting websites -- the "Global Survey Network." All sites in this network are pre-screened and approved for the program. Greenfield Online began to develop this network during 2005, and it currently has hundreds of participating sites ranging from small, low traffic sites to higher volume top web properties.

The recruiting "offer" varies depending on the site, but is generally an offer for a chance to win a substantial cash prize in a drawing for participating in the process. The offer is delivered using various media placements and email solicitations on the recruiting sites.

## 3. How Does Real-Time Sampling Work?

Potential survey takers who agree to take a survey are sent from the web sites in the Global Survey Network that they are on at the time they agree to participate to the Survey Router. Based on their responses to the screening questions, recruits are assigned to the available surveys. This is done on an equal probability basis: the recruit is equally likely to end up in any of the currently open surveys. The actual qualification of a recruit takes place in the survey to which the Router assigns him or her.

All recruits provide an email address so that their participation in surveys can be tracked. Each is returned to the website from which they were originated, regardless of whether they qualified and participated in a survey or not. Greenfield Online has the ability to track future visits from the real-time survey takers. This is done using email addresses and cookies.

Real-Time Sample is composed of users driven to the system by the Global Survey Network, as well as Ad Words, MSN Search, and regular search engine traffic.

## 4. Case Study: Panel Sample vs. Real-Time Sample

An independent study commissioned by Greenfield Online and conducted by Anderson Analytics LLC, an independent marketing-research consultancy, found Panel and Real-Time Internet samples to be quite similar, while limited demographic differences between Panel sample and Real-Time Sample exist, these differences do not impact top-line conclusions of typical market research projects, as evidenced in the following analysis.

The study was designed to determine how different, if at all, Panel sample is from Real-Time Sample and what marketing implications exist by sample source.

Via a typical concept test, we measured respondent interest, overall appeal, and quantified basic pricing and market share for two different chocolate bar advertising concepts. The end goal being to determine if the same conclusions will be reached with regard to the advertising concepts, regardless of which sample source used.

Our analysis concluded that product decisions would be the same whether Panel sample or Real-Time Sample were used.

After conducting this typical product concept test, it is apparent that:

- Both Panel and Real-Time Samples report similar chocolate and junk food consumption, including soda, gum, ice cream and/or candy. When further asked about specific category purchases, brands purchased were also correlated.
- Both Panel and Real-Time Samples report that the amount they spent on the specific category in the last month is similar.
- The two ad/product concepts tested were ranked similarly between Panel and Real-Time respondents, and the “go/no go” marketing decision would have been the same no matter which sample source was used.
- Further, the survey asked respondents to provide an acceptable price range for two products. Again, the results provided by both Panel and Real-Time respondents were the same.

## 5. Methodology and Concepts

In order to test two concepts as realistically as possible, with the primary objective being to test the accuracy of two different sample sources, a two concept four cell monadic concept test design was chosen. A monadic design was chosen as it is typically preferred, and considered to be more pure than a sequential-monadic or proto-monadic test when sample costs are not prohibitive. The four cells were set up as follows:

- Panel sample seeing concept A, Romantic Chocolate (n=632)
- Panel sample seeing concept B, Vitamin Filled Chocolate (n=616)
- Real-Time Sample seeing concept A Romantic Chocolate (n=642)
- Real-Time Sample seeing concept B Vitamin Filled Chocolate (n=626)

The product to be tested, two new chocolate bar concepts, were chosen because chocolate bars have a broad appeal among consumers, and are thus an ideal product to test overall sample composition and differences.

It is generally understood that the more the concepts being tested represent the final product or experience, the more dependable the concept test results. However, as the true objective in this study is measuring Panel vs. Real-Time Sample, rather than predicting exact market share of a potential product, the level of concept refinement is irrelevant.

The concepts to be tested were designed by Anderson Analytics to accurately reflect typical concepts tested online by advertising agencies and consumer packaged goods companies. The concepts (full page magazine layout mock-ups), were designed to

have wide general appeal. Concept A, with a more standard offering and message, and concept B, a somewhat more unique/new and different message.

Concept Copy was short but adequate in communicating the differing sales messages. Concept A used a traditional emotive message of love/romance common in chocolate advertising.

Concept B employed a more humorous angle, and also focused on the taste and health benefits (unique & different) of the chocolate product.

## 6. How the Study Was Conducted

Both Panel sample and Real-Time Sample respondents were invited to the survey, Panelists received emailed invitations, and Real-Time respondents were intercepted in real-time and randomly assigned to participate in a survey. Quotas were set to control for gender, age, and region based on Nielsen NetRatings Audience Profiler data. The male/female split was a standard 45%/55%, distribution for age and region were also standard.

Controlling for these three demographic attributes, may self correct many other demographic differences which could exist between the two sample sources.

Many other demographic, techno-graphic, and behavioral attributes were measured, though not controlled for, in order to quantify how similar and different the two sample sources are.

The survey was fielded on July 25<sup>th</sup>, 2006, and was completed by 2,516 respondents. Within each of the two sample sources, a base size over 1,200 gives a confidence interval of  $\pm 2.77\%$  at the 95% confidence level. Throughout this report, significance testing has been conducted between sample sources and between concepts at the 95% confidence level.

## 7. Category Behavior

### Chocolate Purchasing and Junk Food Consumption

In both Panel sample and Real-Time Sample, over three quarters of respondents report having purchased chocolate within the past three months. In addition there are no significant differences with any of the other candy/junk food category items such as gum, soda, and ice cream.

Panel sample respondents were more likely than Real-Time Sample respondents to have purchased \$5 or less worth of chocolate within the past month.

Panel sample respondents' chocolate purchasing was also significantly less frequent than Real-Time Sample respondents. However the average \$ amount spent on chocolate within the past month (\$6.12 for Panel sample, \$7.22 for Real-Time Sample) as well as the average number of chocolate purchasing occasions (2.0 times for Panel sample, 2.3 times for Real-Time Sample) were calculated and deemed acceptable for purposes of the concept test.

### Chocolate Brands Purchased

Respondent experience and exposure to competing products is an important consideration in concept testing. Therefore, as is customary, respondents were asked about their purchases within the past three months on a list of specific chocolate products.

Both sample groups exhibited recent experience with several of the chocolate products on our list, with over half of respondents in each group having purchased Hershey's within the past three months.

Real-Time Sample respondents were significantly more likely than Panel sample respondents to have purchased several of the chocolate items on our list. The difference was largest (9 points) for the most popular product mentioned, Hershey's.

While recent purchases for some of the items are higher among the Real-Time Sample group, the rank order of the products were the same among both samples.

It is important to note, as discussed earlier in this report, there are some demographic differences between Panel and Real-Time Samples such as the somewhat higher incidence of children in Real-Time Sample households. While the concept being tested in this case was thought to be targeted for personal adult consumption, it is reasonable to believe that households with children will purchase more chocolate products. Marketers may wish to consider, therefore, how such subtle sample differences may affect research projects.

## 8. Concept Test Results

Key concept/product dimensions were measured on a 5 point scale and a concept scorecard was created. Concept B, Vitamin Filled Chocolate, appears to be a clear winner for both Panel sample respondents, and Real-Time Sample respondents based on the three key measures: Overall Appeal, New and Different, and Purchase Intent.

An observed difference between Real-Time and Panel Samples is that within the Panel sample Vitamin Filled Chocolate top-box scores are also significantly higher than that of Romantic Chocolate for the Overall Appeal and New & Different measures.

Interestingly, Real-Time Sample seems to yield slightly higher top-two-box scores for these same measures, thus indicating that Real-Time Sample may be somewhat more inclined to select a "4" rather than a "5" on a 5-point scale.

Slight differences in key measures between the two sample sources aside, Vitamin Filled Chocolate is the clear winner for all three key measures on the concept scorecard. Thus a go/no-go decision based on typical key measures would be identical regardless of the sample source, all other things constant.

### Pricing Results

Measuring price is always a challenge, and customarily some pricing segmentation would take place as part of the analysis prior to arriving at a suggested price point for a new product. However, an overall picture of price sensitivity can be useful. In this case using either Panel sample or Real-Time Sample as a source one would draw the same conclusion. Consumers seem to be willing to pay more for Vitamin Filled Chocolate.

Furthermore the inflection point at \$3.50 is the same in both sample sources. There also appears to be very little difference in the actual percentage of customers willing to pay a specific price point for the product regardless of sample source (about 30% would purchase at \$5, and about 15% at \$3.50).

## Future Share

In order to evaluate concept 'lift' or change in the share of consumers future chocolate purchases, respondents were asked to think about their next 10 chocolate purchases and allocate how many of each chocolate product/brand they would be likely to purchase, assuming the product was available. If they would not purchase a particular chocolate bar they were instructed to enter a "0" next to that chocolate bar. Responses were required to add to 10.

Shortly after being exposed to the concept, respondents were asked to redo the exercise with the new chocolate concept brand included in the product list.

Once again Vitamin Filled Chocolate outperformed Romantic Chocolate regardless of sample source. In fact the results were remarkably similar between Panel and Real-Time Sample. Romantic Chocolate was purchased an average of 0.26 out of 10 times among Panel respondents, and an average of .28 out of 10 times among Real-Time respondents. Likewise, Vitamin Filled Chocolate did equally well among both Panel and Real-Time Sample, Concept B (0.73 times and 0.78 times, respectively).

For both sample sources Vitamin Filled Chocolate took most market share away from the same three top competitors (Hershey's, Snickers, and Reese's).

## 9. Detailed Look at Sample Differences and Possible Implications

### Shopping Behavior

There is little difference in Offline purchasing behavior between Panel sample and Real-Time Sample. Real-Time Sample respondents were only significantly more likely to have purchased one of the items we asked about (Music) than Panel sample respondents (34.1% VS 29.5%).

However, Real-Time Sample respondents were significantly more likely than Panel respondents to have purchased a number of items Online (Food, Coffee, Clothing, Movies, Home Goods, Cosmetics, and Music). However, even in the item with the largest relative difference, food, the percentage difference was less than 5 basis points (36.0% VS 31.2%). Still, marketers whose products are primarily marketed online should be aware of these differences.

## 10. Conclusion

To continue to dialog with respondents, researchers will have to adapt to the changing lifestyles of today's connected digital consumer. To remain compelling, market research must fit into the tight spaces that exist between consumers' busy lives of juggling work and family.

In this hyper speed culture, marketers can't afford to wait weeks or months to find out how consumers feel about their products.

Today researchers have a wider choice to reach respondents for their surveys. There are more opportunities than ever to reach difficult target groups and communicate with consumers that don't want to join a panel. This new technology will not replace the panel sample but creates additional benefits to our customers in a similar way Online became a complementary methodology to face to face and telephone.

Greenfield Online's Panel sample and Real-Time Sample exhibit many similarities. While there are some demographic differences that exist between the Greenfield Online Panel and Real-Time Sampling, this concept test analysis has shown that a marketer is likely to make similar marketing decisions with either sample source.

Real-Time Sample represents a valid sample source, and one that enables Greenfield Online to gather opinions from a broad audience of survey takers beyond its panel of pre-recruited respondents.

Since July 1<sup>st</sup> 2006, 2,8 million unique visitors have gone through the Real-Time Sampling™ methodology to complete 1,9 million qualified surveys.

Although it has been assumed that this method only performs well for high incidence studies our data indicates otherwise as 61,6 percent were achieved in low incidence studies (less than 25 percent) incidence rate. Real-Time Sampling™ works equally as well for high incidence as it does for low incidence projects.

Greenfield Online is now looking to export this solution to Europe and probably to the UK as the market has reached a level of maturity that will require finding other ways to attract potential survey respondents.

At Greenfield Online, we understand the importance of innovation, and leveraging the latest technologies to meet the market research challenge of the future. By venturing into fertile new territory like Real Time Sampling, we expect to continue to supply our clients with the kind of critical market data they need to make time critical decisions.





# Video Web Survey - Results of an experimental comparison with a text-based Web survey

Marek Fuchs & Frederik Funke

## Abstract

Since the early days of Web surveying it was recognized as a potential advantage of computer mediated surveying to enrich studies with multimedia stimuli such as graphics, pictures, spoken word or other sounds. But in fact, these possibilities have only seldom put into action. Just in recent years Web surveys have been enriched by graphics and pictures – some of which are content bearing. Methodological evaluations have shown that these pictures can have a serious impact on the perceived question meaning and thus on the responses provided. In this paper we will report preliminary results from an assessment of the next step towards a full audiovisual Web survey. In spring of 2007 we have conducted a field experimental randomized comparison of a traditional text-based Web surveys to a Web survey containing videos of an interviewer reading the questions to the respondents. In the paper, unit non-response, social desirability and social presence are evaluated. In addition, we assess response times as well as the social setting and the location in which the survey was filled in. Finally we will consider ideas on how the underlying question answer process can be examined in greater detail.

## Keywords

Web survey, audiovisual elements, unit non-response, social desirability, social presence.

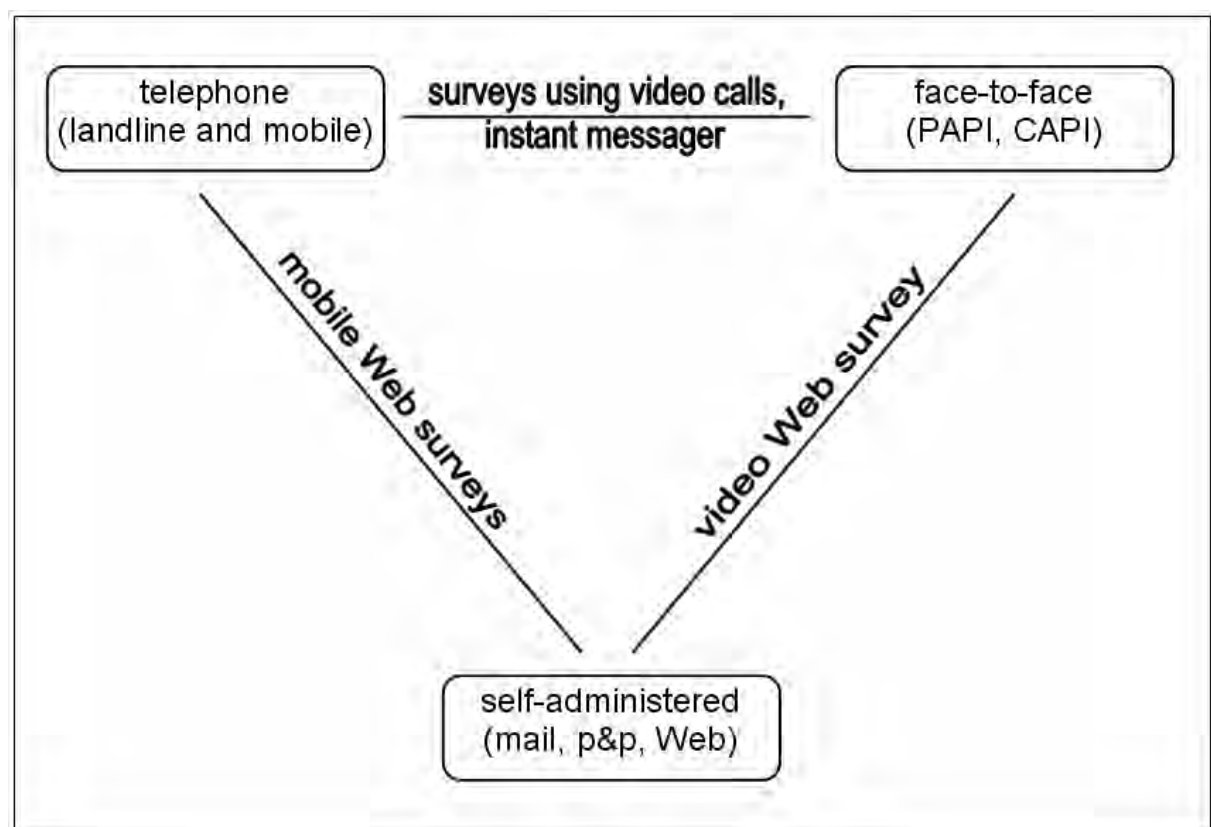
## 1. Introduction and framework

With introduction of computers and modern information and communications technology (ICT), a tremendous change occurred in the way survey data collection takes place. At least two phases can be differentiated: In the early days of computer-assisted data collection the surveys were characterized by an integration of data collection and data entry. Interviewers – or respondents themselves – were supposed to enter the responses at the time of the interview into a user interface. This allowed timely reporting on the results of the study because data were available right after the interview. In the early stages of computer-assisted data collection this feature was predominantly discussed in the light of cost efficiency. Later, it became apparent that big savings were not to be expected, however, an increase in data quality, the ability to design respondent-oriented individualized questionnaires became the key advantages of the integrated data collection process (e.g. Couper et al. 1998).

In the past years, another phase of using ICTs in survey data collection was initiated. Currently, traditional survey modes are in the process of merging (see Figure 1 for graphical rep-

resentation): (1) With the development of mobile Web surveys (Fuchs 2007) researchers combine elements of traditional self-administration (Web survey) and administration by telephone (random digit dialing sampling based on the telephone numbering system of the mobile devices). (2) Surveys using video telephony facilities offered by many cell phone providers draw on sampling and field procedures that have been developed for traditional telephone surveys while at the same time elements of a face-to-face conversation (interviewer and respondents can see each other, including facial expressions and gestures) are being introduced. (3) Finally, with the introduction of live or pre-recorded videos showing human or animated interviewers reading the questions to the respondents, a first step towards a merger of face-to-face interviewing and self-administration in Web surveys takes place. Even though the majority of all interviews is still being conducted using either traditional face-to-face, telephone, or self-administered survey modes, integrated modes are about to emerge.

**Figure 1: Merger and integration of traditional survey modes**



In this paper, we focus on the integration of face-to-face interviewing and self-administration. While this process is still in its infancies, we will describe results from an experimental evaluation of a video enhanced Web survey compared to a traditional text-based Web survey. Obviously, this is just a first step towards an integration of face-to-face interviewing and Web survey data collection. Given the various phases of the interviewer-respondent interaction throughout an interview and considering the various stages of the cognitive question-answer process for each particular survey question we could think of various ways of integrating Web surveys and in-person interviewing: e.g. recruiting respondents with a pre-recorded or live video showing an human interviewer or animated agent (see emarsys 2005 for a test of video e-mails), asking questions (like in this experiment), providing explanations and definitions for respondents who are having problems with a particular question (see Conrad et al. 2007 for a concept of how to provide definitions), using system initiated video clips with motivational

content in order to avoid item non-response or break-offs, using audio recordings of the ongoing interaction throughout the interview (CARI, e. g. Thissen/Sattaluri 2006).

In the experiment reported in this paper we restricted the design to a pre-recorded human interviewer reading the questions to the respondents. The responses are still being collected from the respondent using standard computer input devices (mouse, keyboard). We will evaluate the effects of this approach in a field-experimental comparison with a traditional text-based Web survey; a comparison with a face-to-face interview is not part of this assessment.

## 2. Previous research

In the past, Web surveys have restricted themselves to imitating paper-based questionnaires. The question meaning was conveyed using characters and written text. Also, when response options were offered together with a question – which was most often the case, because open-ended questions weren't used to a great extent – they were also displayed using standard text-based language. From a measurement point of view this was a plausible design decision because one can assume that text-based language promotes uniform question understanding by all respondents.

Like for paper-based questionnaires, several design issues have been analyzed for this traditional text-based Web survey approach: the use of radio buttons vs. check boxes vs. pull-down menus (Smyth et al. 2006), the spacing and positioning of elements on the screen (Christian/Dillman 2004), font size and color (Dillman et al. 2001). Also, the use of progress indicators (Couper et al. 2001) and other interactive elements has been assessed (see Conrad et al. 2003 for an overview). Even though it was shown that those formal aspects of the questionnaire might affect the question-answer-process on the respondent's side – and thus the responses obtained – these characteristics are not content-bearing in the first place.

Some extensions of the traditional approach were to be observed: increasingly pictures and other graphical elements were used to enhance the visual appearance of the questionnaire. E.g., Bandilla and Bosnjak (2000) made use of pictures and animated object in Web surveys, predicting deeper cognitive processing of the underlying tasks when using such multimedia-based elements. Also, to a limited extent, pictures were used to transmit the concepts addressed by a particular question (Couper/Tourangeau/Kenyon 2004).

In this paper we will report results from an experiment where the traditional text-based approach has been overcome: We are using video-clips built into a web survey in order to convey the question meaning. In these videos an interviewer is reading the questions to the respondents using the multimedia capabilities of the respondent's computer. Responses are collected using traditional input devices provided by every computer (keyboard and mouse).

In the past, many studies have analyzed the impact of technology on survey data collection in general and on various aspects of the total survey error. Thus, usually in the literature results from comparisons of paper-based and computer-based surveys are reported. As a consequence only little empirical evidence to draw on is available when it comes to a comparison of text-based and video-enhanced Web surveys. The only line of research we can refer to is the methodological evaluation of audiovisual computer-assisted self-interviewing (V-CASI, also AV-CASI). In AV-CASI an electronic self-administered questionnaire includes video clips of a pre-recorded interviewer reading the questions to the respondent (Gerich et al. 2003; Katz et al. 2007).

AV-CASI has been developed in order to collect sensitive information. When responding to sensitive questions in the presence of a live interviewer or with third persons being present during the course of the interview, respondents might be reluctant to answer with a valid response. By contrast, they tend to report in a socially desirable way or refuse to answer altogether. In order to overcome such resistance, respondents are equipped with headphones which prevent others to overhear the questions, like drug use, sexual preferences, and crime. In addition, responses were collected self-administered in order to prevent respondents from reacting to the presumed interviewer expectations in the process of delivering the response. Several studies could demonstrate the positive effects of A-CASI. In 1998, Turner and colleagues (1998) were studying adolescent high-risk sexual, drug use, and violent behaviors. Respondents admitted up to 17 times more frequently high-risk sexual behavior when subjects were interviewed via A-CASI. As an extension of A-CASI, video clips have been used which show an interviewer reading the questions to the respondent. Several studies have assessed the effects of AV-CASI in contrast to traditional face-to-face interviewing. Katz and colleagues (2007) have evaluated AV-CASI in the context of a blood donor survey and found superior data quality for AV-CASI. Similarly, Weisband and Kiesler (1996) found a more pronounced self-disclosure in the computer-assisted self-administered mode compared to face-to-face interviewer administration – especially for sensitive topics.

Based on these findings from the literature, we assume that video-enhanced Web surveys make use of the advantages of both, interviewer administered interviews and traditional text-based self-administered Web surveys, thus providing better data quality: The pre-recorded interviewer resembles to a certain degree the face-to-face interview setting which is assumed to be advantageous compared to mere self-administered surveys (O'Reilly et al. 2004). Also, a video Web survey offers some practical advantages compared to AV-CASI since interviewer and respondent no longer need be present at the same location. At the same time we can draw on the advantages of self-administration when it comes to the collection of sensitive information.

Compared to traditional face-to-face interviews a video Web survey – like V-CASI (Gerich et al. 2003; Gerich/Lehner 2006) – should promote standardized interviewer behaviors. Since the interviewer is pre-recorded, no situational variation or differences due to interviewer mood are likely to occur. Also, the interviewer cannot react to personal characteristics of the respondent or to the responses provided. In addition, because skip instructions and branching as well as the sequence of the questions are programmed into the AV-CASI system, no accidental or intentional omission of questions can occur and all questions are administered exactly as worded.

So far, the use of video clips showing interviewer reading a question to the respondents is being evaluated compared to traditional face-to-face interview methods (e.g. Katz et al. 2007). Our approach however, assesses the use of pre-recorded interviewers in Web surveys. Videos are being used in order to enhance the respondent's experience when answering the questions, to support and deepen cognitive processing and thus to foster question understanding. As a result we expect to reduce a certain negative effect of Web surveys which has been demonstrated in contrast to a paper-based self administered questionnaire: In a previous study on mode effects of Web surveys compared to an identical paper and pencil survey (Fuchs 2003) it could be demonstrated that respondents showed smaller question order effects in the Web survey mode. This was interpreted as a sign of greater segmentation of each question, which in turn would endanger complete question understanding. Using video segments showing an

interviewer is thus assumed to reduce this segmentation effect since spoken language should increase the impact of the questions in the respondents mind.

However, at the same time the introduction of a human speaker in Web survey might increase social desirability bias. Based on a meta-analysis of mode effects on social desirability (Richman et al. 1999) it is save to assume that interviews yield larger such effects compared to self-administered surveys. Social interface theory suggests that humanizing cues in a computer interface can stimulate responses from users similar to human-human interaction (see Couper/Tourangeau/Steiger 2001 for a summary). Based on this reasoning one could expect greater degrees of social desirability in the video mode compared to the text-based Web survey. By contrast, the literature on the survey interview suggests that computer administration of surveys on highly sensitive topics reduces social desirability effects, even when such humanizing features as voice are used (Tourangeau/Smith 1998; Turner et al. 1998). Whether the effect is reduced to a level that is typical for self-administered surveys is not yet clear. However, because video enhanced Web surveys do not allow direct interaction of interviewers and respondents, the increase in social desirability compared to the text-based Web survey should be rather moderate.

A study from a related field of research is in support of this hypothesis: owever Staples and Luzzo (1999) compared a paper-and-pencil version of a vocational assessment form and its multimedia counterpart including pictures and voice-overs. Results indicate that the multimedia version yields scores similar to those obtained by traditional paper-and-pencil mode. We have to consider that test forms are not fully comparable to survey questionnaires and also one should take into account that our comparison includes a text-based Web survey which will be compared to a multimedia enhanced Web Survey. Nevertheless, the results of Staples and Luzzo (1999) provide some support for the assumption of no differences across the two modes.

Based on this preliminary review of the literature we focus on the following research questions:

- (1) Is it possible to administer a video enhanced Web survey? What are the limitations in terms of the hardware and software available on the respondents' side? How would respondents evaluate a video enhanced version of a questionnaire?
- (2) To what extent does social desirability bias interfere with the measurement process? Is social desirability a problem due to the social presence of the interviewer?
- (3) How does the introduction of the interviewer change the question answer process? Even though we will not be able to assess the question-answer process in greater detail we will evaluate the extent to which the respondents are focused on the survey questions.

### 3. Design of the experiment

In order to explore the effects of video usage in Web surveys we conducted a Web experiment comparing a traditional text-based Web survey to an identical video enhanced Web survey. The survey focused on drug usage among university students enrolled at the home university of the authors. 962 students registered at the online access panel of the University of Kassel (maintained by the authors) were randomly assigned to one of the two experimental groups and invited by e-mail. The initial e-mail was sent on March 13th 2007, the reminder e-mail was sent 16 day later, on April 1st. We offered one out of five USB sticks as an incen-

tive to respondents selected in a lottery drawing at the end of the field phase. There was no difference in the text of the e-mail invitations, however, in the video condition there was a note included that short video clips would be presented throughout the survey and that speakers or headphones were required to participate. 44 emails could not be delivered (account expired, mailbox full, invalid email address), resulting in a net sample of 918 persons (459 were invited to the video condition and 459 to the text condition). To participate, students had to log in to the online access panel using their personal accounts.

In video mode (see Fig. 2) pre-recorded video clips of an interviewer reading the questions to the respondents were included. The interviewer also called all possible response options respectively the instruction on how to answer the question (e.g. "Enter the appropriate number in the field below"). In addition, for closed-ended questions the appropriate response categories (e.g. "yes" or "no") were displayed on the screen. The interviewer shown in the video clips was a female German native speaker, aged 28. She was not a student of the university and should not be known personally by the respondents.

The size of all videos was 240 pixels in width and 180 pixels in height. The videos were presented as flash movies, so that the only technical requirement was that the free and widely spread flash player had to be installed on the client's computer. As soon as the participant entered a particular page throughout the survey the respective video would start playing automatically (push mode). Participants had the opportunity of re-playing the videos as often as they liked.

The text-based version of the survey resembled a traditional Web survey. All necessary information was conveyed using standard text-based language. In order to compare the two questionnaire versions we paralleled as many characteristics of the questionnaire as possible: The Web page had exactly the same dimensions, the question text replaced the video and all other elements (logo, rating scale and submit button) were exactly in the same position as in video mode. Both questionnaires can still be accessed online:

[http://www.onlinepanel.org/surveys/mumeon\\_01\\_2/](http://www.onlinepanel.org/surveys/mumeon_01_2/) (video enhanced version),

[http://www.onlinepanel.org/surveys/mumeon\\_01\\_1/](http://www.onlinepanel.org/surveys/mumeon_01_1/) (text-based version)).

Even though the content of the survey was identical for both conditions, the introductory page differed. In the video condition, a sample video was looped, telling the participant that there were no problems on a technical level. If the technical requirements on the respondent's computer were not sufficient, a text was displayed, telling the participant to go to another computer, capable of presenting multimedia content.

**Figure 2: Screenshots from the video enhanced Web survey (left) and from the text-based Web survey (right)**



In both survey conditions the same questionnaire was used applying an interactive Web survey mode. The questions were presented on 20 pages in the text-based version and 21 pages in the respective video enhanced survey. The questionnaire was organized in three sections: In part 1, the experimental treatment (video enhanced Web survey versus text-based Web survey) took place. In part 2, the setting in which the questionnaire was answered was assessed using text-based Web survey methodology for both conditions. Part 3 was only presented in the video condition, evaluating the quality of the video – also using text-based questions. In both conditions, participants were offered the possibility of commenting on the interview in the end.

Part 1: In both experimental conditions, 17 thematic items were presented in a “one page one item” design. The items were taken from standard inventories on health risk behavior. Some newly developed questions on drug usage were used as well. 7 items were on the use of and attitude towards alcohol, 3 items on smoking and 3 items on the use of other drugs. Additionally, 4 items on social desirability (taken from the German General Social Survey inventory) were asked.

Part 2: 6 items on the perceived interview situation were presented on another single page. Finally, 4 items on the social setting (third parties, location) and 1 item on the speed of the internet connection had to be answered.

Part 3: In addition, in the video condition only, 5 items on the quality of the video and the perceived time needed for loading the video were asked on a separate Web page.

In order to explore the impact of the video enhanced mode we used a variety of dependent variables: First, we assessed survey participation. In addition to the proportion of unit non-response we looked for lurkers and break-offs. Also, the location and the social setting of the respondents when answering the questions were assessed (alone vs. in the company of others, public vs. private place).

For all respondents who completed the questionnaire, we assessed response time for each page of the Web survey applying a client-sided measurement via JavaScript. In the analysis of the response times outliers (defined as mean plus or minus 2 standard deviations) were removed. In addition we assessed item non-response for each substantive question. Also, item non-response was accumulated case wise. Finally, three constructs were compared across the experimental versions: Social desirability was measured using a 4 item standard inventory taken from the German general social survey. In addition, perceived social presence was measured using a 4 item inventory. Also, the personal involvement was measured using 3 items (details on the items will be presented together with the results).

## 4. Results

### 4.1. Unit and item non-response

In both modes, the gross sample consists of 459 potential respondents. In video mode, 44% (204) participants reached the start page of the questionnaire. By comparison, in the text-based mode, 51% (233) accessed to the start page of the survey. This difference is most likely due to the note in the invitation e-mail that the computer has to be capable of playing video and sound. Since members of the gross sample were treated identical regardless of the version they were assigned to – with the exemption of the note mentioned above –, no other reasonable explanation comes to mind. So far, the reason for the lower rate of respondents accessing



the first page of the survey cannot be determined for sure. Three possible explanations arise: (1) Even though we assumed almost every student computer being capable of playing videos it might well be that we overestimated the respective proportion. As a consequence, all selected respondents who have actually no multimedia computer available or who assumed that their computer would not play video and sound did not bother to access the survey. (2) Many students check their e-mail on various occasions during the day using computers at various places on campus. Many of these computer in the library, the university's computing center, public Internet cafes and so on might not be capable of playing videos and sound. If the computer that they used to read the invitation e-mail was not able to play audiovisual content, students were told to access the survey from another computer, e.g. from their own computer at home. It might well be that several of selected respondents might not have taken that extra effort to access the survey. (3) The topic of the survey gave a hint that some sensitive questions would be asked. The third possible explanation is based on the perceived burden arising from sensitive questions being asked using video and sound. Some selected respondents might have feared that it would be burdensome to answer questions on alcohol, nicotine, and other drugs in public if others could overhear the questions. Thus, even though they would have a multimedia computer available they did not access the survey.

So far we have no data to exclude one of the above possible explanations. The only hint we have is related to the location where respondents filled in the questionnaire: in the video-based mode the proportion of people being in a private surrounding using their own or a friend's or relative's computer while answering the survey is far higher.

In future studies we will explore other means of inviting respondents without letting them know that audiovisual content will be used. Also we will ask all selected respondents a few follow up questions on their thoughts while considering survey participation.

In the video condition, 17% (35) of participants who entered the survey dropped out during the questionnaire, most of them (14) already on the starting page of the survey and another large group (9) on the first thematic pages. The other break-offs occurred in later parts of the video-enhanced part of the survey. In the text-based condition, only 1% (2) of participants dropped out in the same part of the study. This suggests that some respondents accessed the survey to check whether or not they can see and hear the content of the survey. When they determined that they would not be able to take part in the survey due to a lack of multimedia facilities they dropped out. Unfortunately, they did not return from another computer, which suggests implement future experiment in studies with more interesting topics or more valuable incentives.

No lurkers were to be observed. Also, after deleting break-offs no incomplete questionnaires (less than 60% of all items answered) were to be noticed. After excluding break-offs a net sample 37% (168) in video mode and 50% (230) in the text-based mode was available. For all further analyses only complete questionnaires were regarded.

Overall, item non-response was quite low in both conditions, ranging from 0.21 unanswered items in video mode to 0.27 in text mode (not significant). Only for 1 out of the 18 questions that were experimentally varied, a significant difference in the proportion of item non-response was to be noticed where we get about 4% item non-response in the text mode while no item non-response occurred in the video mode ("How many cigarettes did you smoke in the past one week?" <open input field>;  $p < .05$ ).

#### 4.2. Location and social setting

As mentioned above, we assumed that respondents in the video mode would be more reluctant to answer the questions in public locations compared to the respondents in the text-based mode. Thus, we thought that certain places would be preferred either for text mode or for video mode. Not every place might be suitable for answering sensitive questions, as there are different levels of privacy connected with different locations. In video mode the interviewing situation can be less private when loudspeakers are used, so we assumed that the participants would avoid certain locations. The rationale for this assumption builds on the perceived burden of answering sensitive question in a location where others can overhear the questions (and might get interested in the responses as well). Table 1 shows the responses of a respective question:

**Table 1. Location**

	video enhanced Web survey	text-based Web survey
at own room in a shared flat in Kassel	19% (32)	14% (33)
at own apartment in Kassel	27% (46)	26% (59)
at a friend's home	4% (6)	4% (10)
at parents' home	33% (56)	30% (70)
at the university's library	0% (0)	0% (1)
at the university's computer center	1% (1)	1% (3)
at an internet café	0% (0)	0% (0)
at another place	17% (28)	24% (54)
Total	100% (169)	100% (230)

Most locations have been used by nearly the same percentage of participants in both modes. The high share of people having answered the questionnaire from their parents' home is most likely because the survey took place during a holiday season. In the video mode, a combined 46% of respondents answered the questionnaire at their own apartment or in their room in a shared flat (a common living arrangement for students in Kassel). By contrast, the respective proportion in the text-based mode is 6 percentage points smaller. If we consider the difference

in the proportion of people answering in their parent's home (which shows a similar pattern) a far larger proportion of respondents answered the video enhanced survey in a privat setting compared to the text-based survey. By contrast, more respondents answered the text-based survey "at another place". We assume that those other places would be work places or hot spot location with WLAN access. Based on this, we assume that people avoid taking part in video enhanced surveys from those other places. Only 17% participated from another place in video mode, whereas 24% did so in text mode ( $p = .058$ ).

**Table 2. Social setting**

	video enhanced Web Survey	text-based Web Survey	total
Were you alone or in the company of others while participating?			
Alone during the whole study.	85% (143)	85% (196)	85% (339)
In part other persons were present in the same room.	8% (14)	7% (17)	8% (31)
Others in the same room present during the entire survey.	5% (9)	5% (11)	5% (20)
I participated in the survey to- gether with another person.	1% (2)	3% (6)	2% (8)
total	100% (168)	100% (230)	100% (398)
While answering the survey, did you discuss the questions with another person in front of the computer screen?			
yes	4% (6)	3% (7)	3% (13)
no	96% (163)	97% (223)	97% (386)
total	100% (169)	100% (230)	100% (399)
Could anyone besides you also hear or see the video/read the questionnaire?			
yes	14% (24)	5% (12)	9% ** (36)
no	86% (143)	95% (211)	91% (354)
total	100% (167)	100% (223)	100% (390)

\*\* =  $p < .01$ .

In addition to the location investigated social setting while answering the questions (see table 2). First, we asked for any company while responding to our questions. Results show that in both conditions (no significant differences) the majority of all respondents were on their own when answering the survey. Under both conditions 85% or respondents had no company. Smaller groups answered with other people being present in the same room, or responded to

the questions jointly with others. As a consequence only a minority discussed the questions with others while answering the questionnaire. 4% in the video enhanced mode and 3% in the text-based mode did so (not statistically significant). Considering these results one could come to the conclusion that respondent behavior was pretty much identical in both conditions.

However, we have to remember that respondents seemed to avoid public locations while responding. Nevertheless, our video Web survey led to a far higher proportion ( $p < .01$ ) of people who told us that others could either see (5%, text-based mode) or see/hear (14%, video enhanced) the questions while they responded to them. Therewith we can record the finding that it was more difficult to obtain privacy in the video enhanced survey than in the text-based survey.

With 55% (93), most participants listened to the video using the computer's build-in speakers, another 35% (58) chose external speakers and only 4% (17) chose the most private situation when using headphones. We assume that most respondents had no choice than to use whatever equipment was available at the computer that they were using. However, we would predict that some of the result would be changed dramatically if respondents could chose to make use of headphones.

#### 4.3. Social desirability, social presence and involvement

In order to evaluate the quality of the data obtained we have measured social desirability bias using a four item scale taken from the German General Social Survey (Table 3, top section). Only for one of the four items we observe a significant difference according to which the video mode seemed to yield more socially desirable answers. For the three other items no difference occurred. As a consequence, the additive social desirability index computed based on the four items shows no difference in the degree of social desirability.

As far as social presence of the interviewer is concerned we get a surprising result. For two of the three items we get higher values for the video mode indicating less (!) social presence. In the second block respondents were supposed to express the degree to which they would have answered similarly in an interview with a live interviewer. Results indicate that respondents in the video mode were less likely to answer in the same way compared to the text-based mode ( $p < .05$ ). Also, respondents in the video mode felt less personally addressed by the questions ( $p < .05$ ). Even though, a pre-recorded interviewer was reading the questions to them we get larger values from the respondents in the video mode which indicates that they agree with the statement to a smaller extent. This indicates that respondents experienced the presence of an interviewer/researcher less dominant in the video mode.

**Table 3. Social desirability, social presence and involvement**

	video enhanced Web survey	text-based Web survey	total
To what extent do the following statements apply to you?			
I always tell the truth. (% yes)	53	55	54
I sometimes get angry when my wishes are not fulfilled. (% no)	18	28	22 *
I am always willing to admit failure (% yes)	77	78	77
Occasionally I have said something intentionally that has offended someone else. (% no)	50	42	47
Mean index social desirability	2.0	2.0	2.0
What was your impression while answering the questions?			
I thought I was talking to a live person.	3.9	3.9	3.9
In a survey with a live interviewer I would have answered similarly.	1.6	1.9	1.7 *
I felt personally addressed by the questions.	2.5	2.7	2.6 *
Mean index social presence	8.0	8.6	8.3 *
I have focussed very much on the single items.	2.0	2.1	2.0 *
I could imagine very well the topics addressed in the questions.	1.8	1.9	1.8
I paid very much attention to the details in the questions.	1.9	2.0	1.9
I have carefully considered my responses.	1.8	1.9	1.8
Mean index involvement	7.4	7.9	7.6 *

Notes: This is a raw translation of the items done by the authors. Scales ranging from 1 = "completely agree" through 5 = "completely disagree". The social desirability items were asked in the respective experimental mode; all other items were asked in text-based methodology for all respondents. Indices were computed by simply adding the values from the respective items. Larger values of an index indicate that the concept in question is less pronounced. \* =  $p < .05$ .

The last four items were concerned with the involvement of the respondent. We wanted to assess the degree to which they would focus on the survey as a whole and the single questions in particular. Results indicate that respondents paid attention to a similar degree in both modes. Only in one out of four items we get a significant difference. Respondents say that

they paid somewhat less attention to the single questions in the video mode. Since there are several small – but non-significant – differences all in the same direction the additive index measuring involvement shows a more pronounced effect. Respondents in the text-based Web survey were more concentrated on the content of the survey. However, we have to take into account that these questions were asked using text-based methodology. Also, one has to consider, that we did not display a written version of the questions together with the video.

#### 4.4. Technical aspects of the video

At the end of the survey respondents in the video enhanced mode were asked to rate the quality of the video presentation. First, respondents assessed the size von the video which was 240 x 180 pixels. On a scale running from 1 = “too large” through 5 = “too small” we get an average of 3.5 indicating that the majority of respondents was more or less satisfied with the size of the video. However, for future use we are encouraged to increase the size slightly (see Table 4).

**Table 4. Perceived quality of video**

Size of the video (1) (mean)	3.5 (SD = 0.6)
Picture quality of the video (2) (mean)	2.7 (SD = 0.9)
Sound quality of the video (2) (mean)	2.2 (SD = 0.9)
Loading speed of the video (3) (mean)	1.2 (SD = 0.5)

Notes: (1) scale running from 1 (“too large”) over 3 (“just right”) to 5 (“too small”), (2) scale running from 1 (“very good”) to 5 (“very bad”), (3) scale running from 1 (“without any delay”) to 4 (“with great delays”).

In addition, we asked for an assessment of the quality of the picture und the sound of the video. While the sound seems to be about right (2.2 on average on a scale from 1 = “very good” through 5 = “very bad”) we have to improve our recoding equipment in order to achieve better picture quality. Even though respondents rate the quality of the picture above scale midpoint, we need to consider improvements. One of our most serious concerns related to the use of videos in a Web surveys were related to the latencies introduced into the survey due to download time for the videos. To our surprise we get a very high rating of the perceived loading speed of the video (1.2 on a scale ranging from 1 = “without any delay” through 5 = “with great delays”). A closer look on the Internet connection revealed a difference between people having a slow connection (i.e. modem or ISDN) and a fast internet connection (i.e. DSL or some kind of LAN). Respondents with a fast connection (91%) rated the loading speed on average with 1.2 and respondents with a slow connection (9%) rated the loading speed on average with 2.2. Even though this difference is statistically significant ( $p < .001$ ), the loading speed is still acceptable with slower Internet connections.

#### 4.5. Response time

For the analysis of response time we removed cases with less than 1 second per page from the data set, as a serious answer could not be given in such a short time. Furthermore, for each

Web page, we excluded all cases from analysis where the response time falls apart more than two standard deviations from the mean.

Because there was an additional introduction in video mode, the start pages differed from each other to a great extent. In video enhanced mode, the mean response time was 28 sec in text-based mode about 8 sec ( $p < .001$ ). However, the mean response time for the remaining 18 thematic pages of the survey did not differ to a great extent (19 sec in video mode and 18 sec in text mode, not significant).

## 5. Discussion

Our preliminary results indicate, that the video enhanced Web survey suffers from a lower response rate, which is in part due to technical requirements that client computers did not fulfil. In addition, we assume that audiovisual content of a Web survey is less likely to be accessed in public. Both effects reduce the response rate; whether or not this involves a possible non-response bias remains unclear. Also, the lower response rate in the video-enhanced Web survey is related to survey costs, which might outweigh the assumed gains in data quality. The introduction of video clips seems lead to a greater dropout because of technical requirements which in turn reduces the size of the net sample.

The technical requirements of the video-enhanced mode might have forced respondents to answer the questions more often in the privacy of their homes. Because the survey invitation states that speakers or headphones are necessary to take part, the survey respondents might have chosen more often a location where these features were available (at home). Alternatively, respondents might have feared that it would be a burdensome experience to answer questions on nicotine, alcohol and drug consumption in public – especially if the questions were read to them aloud. In future experiment we will test this assumption in greater detail. We will develop some items in order to explore the respondents' considerations when choosing the location. Also, we will test whether this effect occurs also with less burdensome survey topics.

In terms of the feared social desirability bias, we detect no differences. Both survey modes yield similar levels of social desirability (as measured by our 4 item scale). In the light of the theoretical reasoning mentioned in the introduction, we come to the preliminary conclusion that the introduction of a pre-recorded human interviewer does not increase the social desirability distortion. We will have to replicate this finding with other speakers (males and females, humans and animated agents) and other sensitive topics.

Interestingly, the introduction of the human-like administration of the interview has decreased the perceived social presence. Unfortunately we have measured this concept with just a few items; in future studies we will have to use an extended version of a respective scale. Also, we have to experiment with both survey modes – so far we have used text-based methodology to assess the social presence indicators.

Even though we are going to proof the equivalence of the text-based and the video-enhanced Web survey mode in terms of the social presence in future studies, we have some preliminary thoughts on a possible reason for this effect: In the video-enhanced mode the interviewer was reading the questions to the respondent, however, she did not collect the responses using spoken language. Respondents still used the keyboard and mouse to provide responses and thus the mode on the reporting stage of the questions answer process was still self-administered – even though the interviewer might be visible while answering the question. In our view, an-

swering our video-enhanced Web survey was more like sitting in front of a TV set: Even though stranger might be on the screen delivering verbal and non-verbal messages the viewers would not interrupt a private conversation or desist from any personal matter. People are used to live their lives in front of the TV, thus, the perceived social presence in the video-enhanced Web mode and its impact on social desirable answers might be rather small. However, none of this speculation has been proven or even tested so far.

In addition to the small differences in terms of social presence and social desirability another finding comes to our astonishment: respondents were less focused on the survey questions in the video mode compared to the text-based survey mode. So far we do not have a compelling explanation for that phenomenon. It might well be that this is the consequence of the fact that we have restricted the video enhanced survey to the verbal communicational channel (not written versions of the questions were displayed on screen). The omitted text in the video mode might have compensated for the improvements introduced by the video channel. Another explanation arises from the discussion of channel richness: Even though a video clip is by far richer in terms of the communicative channels it might well be these additional signals do not lead to a more focussed respondent behaviour. Because only a small subset of the cues conveyed by a video clip is directly related to the question wording and question meaning the richness of the video channel could in fact distract the respondents from the question answer task.

Looking at the respondents' subjective evaluation of their experience while working through the video-enhanced survey we get some mixed results: The size of the video gets a good rating however, it should be slightly larger. Definitely we need to improve our technical equipment when recording the interviewer. Respondents evaluate the quality of video and audio as rather moderate. In the planning stage of the project we considered several video formats for the clips; after some test we have chosen flash format mainly because of the small file size of the video clips and the wide spread use of the free flash player. Even though, we have not experimentally examined this decision it seems reasonable: Respondents are quite happy with the duration between download and start of the video. Even though respondents preferred higher quality recordings in general video enhanced Web surveys seemed to be doable.

Long since the invention of the Web survey and other survey mode that allow for multi channel communication with our respondents, the claim has been made that researchers should make use of the multi media elements in surveys because this would yield data of superior quality. Our preliminary results do not provide compelling evidence that this is really the case. Even though, the findings reported in this paper mark just the beginning of the line of research on video clips in Web surveys, we still lack convincing arguments in favour of the this methodology. In fact, one could argue that it is not worth the effort to use multi medial elements in Web surveys because the benefits are minimal at best. Further, such critique would state that Web surveys should restrict themselves to rather puristic text-based designs.

Even though we cannot decline this line of arguments at this time, we are reluctant to adopt this view without further testing. In addition, given the current changes in the design of many Web pages we predict that Web surveys have to change their appearance substantially. In a Internet world that consists of Web sites that include sound, animated gifs, video clips and the like on a standard bases, Web surveys cannot stay behind and still mimic the pure text-based approach of paper and pencil surveys without been ignored or overlooked. Thus, Web survey researcher will have to embed sound and video just because Web pages will consist of such elements in the near future, regardless of the benefits in terms of data quality. However, in



addition to this anticipated constraint we are still convinced that video-enhanced Web surveys will demonstrate advantages on their own account.

## References

- Bandilla, W., & Bosnjak, M. (2000). Online-Surveys als Herausforderung für die Umfrageforschung – Chancen und Probleme. In P. Mohler & P. Lüttinger (Eds.), *Querschnitt – Festschrift für Max Kaase* (pp. 9-28). Mannheim: ZUMA.
- Christian, L. M., & Dillman, D. A. (2004). The influence of graphical and symbolic language manipulations on responses to self-administered questions. *Public Opinion Quarterly*, 68, 57-80.
- Conrad, F. G., Couper, M. P., & Tourangeau, R. (2003). Interactive Features in Web Surveys. Paper presented at the Joint Meetings of the American Statistical Association, San Francisco, CA, August 6, 2003.
- Conrad, F. G., Schober, M. F., & Coiner, T. (2007). Bringing features of human dialogue to Web surveys. *Applied Cognitive Psychology*, 21, 1-23.
- Couper, M. P. (2005). Technology Trend in Survey Data Collection. *Social Science Computer Review*, 23(4), 486-501.
- Couper, M. P., Baker, R. P., Bethlehem, J., Clark, C. Z. F., Martin, J., Nicholls, W. L., et al. (Eds.). (1998). *Computer Assisted Survey Information Collection*. New York: Wiley.
- Couper, M. P., Tourangeau, R., & Steiger, D. M. (2001). Social Presence in Web Surveys. Paper presented at the Conference on Human Factors in Computing Systems, Seattle, WA, March 31-April 4.
- Couper, M. P., Traugott, M., & Lamias, M. (2001). Web survey design and administration. *Public Opinion Quarterly*, 65(2), 230-253.
- Couper, M. P., Tourangeau, R., & Kenyon, K. (2004). Picture This! Exploring Visual Effects in Web Surveys. *Public Opinion Quarterly*, 68(2), 255-266.
- Couper, M. P., Singer, E., & Tourangeau, R. (2004). Does voice matter? An interactive voice response (IVR) experiment. *Journal of Official Statistics*, 20(3), 551-570.
- Dillman, D. A., Phelps, G., Tortora, R. D., Swift, K., Kohrell, J., & Berck, J. (2001). Response Rate and Measurement Differences in Mixed Mode Surveys Using Mail, Telephone, Interactive Voice Response and the Internet. Unpublished manuscript.
- emarsys. (2005). Video-E-Mails - kommen sie an? Zustellbarkeit und Darstellung von Multimedia-E-Mails. Wien, AT: emarsys eMarketing Systems AG.
- Fuchs, M. (2002). Computer-Assisted Self Interview (CASI). Möglichkeiten für Zeitbudgeterhebungen. In M. Ehling & J. Merz (Eds.), *Neue Technologien in der Umfrageforschung. Anwendungen bei der Erhebung von Zeitverwendung* (pp. 67-84). Baden-Baden: Nomos.
- Fuchs, M. (2003). Kognitive Prozesse und Antwortverhalten in einer Internet-Befragung. *Österreichische Zeitschrift für Soziologie*, 28(4), 19-45.

- Fuchs, M. (2007). Mobile Web Survey. A preliminary discussion of methodological implications. In F. G. Conrad & M. F. Schober (Eds.), *Envisioning the Future of Survey Interviews* (pp. in press). New York: Wiley.
- Gerich, J., & Lehner, R. (2006). Video Computer-Assisted Self-Administered Interviews for Deaf Respondents. *Field Methods*, 18(3), 267-283.
- Gerich, J., Lehner, R., Fellingner, J., & Holzinger, D. (2003). Animaqu - eine computerbasierte Befragung als Möglichkeit zur Erfassung besonderer Zielgruppen. *ZUMA Nachrichten*, 52, 35-54.
- Katz, L. M., Cumming, P. D., & Wallace, E. L. (2007). Computer-Based Blood Donor Screening: A Status Report. *Transfusion Medicine Reviews*, 21(1), 13-25.
- Kissinger, P., Rice, J., Farley, T., Trim, S., Jewitt, K., Margavio, V., et al. (1999). Application of Computer-Assisted Interviews to Sexual Behavior Research. *American Journal of Epidemiology*, 149(10), 950-954.
- O'Reilly, J. M., Hubbard, M. L., Lessler, J. T., Biemer, P. P., & Turner, C. F. (1994). Audio and Video Computer Assisted Self-Interviewing: Preliminary Tests of New Technologies for Data Collection. *Journal of Official Statistics*, 10(2), 197-214.
- Richman, W. L., Kisler, S., Weisband, S., & Drasgo, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology*, 84(5), 754-775.
- Schneider, S. J., Cantor, D., & Hagerty-Heller, T. (2005). Interactive Voice Response (IVR). In S. J. Best, Radcliff, B. (Ed.), *Polling America: An Encyclopedia of Public Opinion* (Vol. 2, pp. 349-356). Westport, CT: Greenwood.
- Smyth, J. D., Dillman, D. A., Christian, L. M., & Stern, M. J. (2006). Comparing Check-All and Forced-Choice Question Formats in Web Surveys. *Public Opinion Quarterly*, 70(1), 66-77.
- Staples, J. G., & Luzzo, D. A. (1999). Measurement comparability of paper-and-pencil and multimedial vocational assessments. Iowa City: American College Testing Program.
- Thissen, R. M., & Sattaluri, S. (2006). Research and Development in Audio Recorded Interviewing. Paper presented at the International Field Directors and Technologies Conference, Montreal, CA, May 2006.
- Tourangeau, R., & Smith, T. (1998). Collecting Sensitive Information with Different Modes of Data Collection. In M. P. Couper, R. P. Baker, J. Bethlehem, C. Z. F. Clark, J. Martin, W. L. Nicholls & J. M. O-Reilly (Eds.), *Computer Assisted Survey Information Collection*. New York: Wiley.
- Turner, C. F., Forsyth, B. H., O'Reilly, J. M., Cooley, P. C., Smith, T. K., Rogers, S. M., et al. (1998). Automated Self-Interviewing and the Survey Measurement of Sensitive Behaviors. In M. P. Couper (Ed.), *Computer Assisted Survey Information Collection* (pp. 455-474). New York.
- Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L. D., Pleck, J. H., & Sonenstein, F. L. (1998). Adolescent Sexual Behavior, Drug Use, and Violence: Increased Reporting with Computer Survey Technology. *Science*, 280, 867-873.

Weisband, S., & Kiesler, S. (1996). Self Disclosure on Computer Forms: Meta-Analysis and Implications. Paper presented at the CHI 96, Vancouver, BC, Canada.

Prof. Dr. Marek Fuchs, Department of Social Sciences, University of Kassel, D-34109 Kassel, Germany, +49.561-804-3102, [marek.fuchs@uni-kassel.de](mailto:marek.fuchs@uni-kassel.de)

Frederik Funke, M.A., Department of Social Sciences, University of Kassel, D-34109 Kassel, Germany, +49.561-804-3245, [email@frederikfunke.de](mailto:email@frederikfunke.de).

# Watch What I Do! - Using graphical input controls in web surveys

Nicola Stanley & Stephen Jenkins

## Abstract

This paper considers the pragmatic aspects surrounding the deployment and use of graphical image-based controls for collecting web survey response data. It compares the relative merits of graphical control surveys in terms of the respondent experience of survey completion including usability, engagement and enjoyment of taking part in the survey. The design effort required of the survey development team is also considered. In particular the paper explores whether interactive web graphical scales can be a direct replacement for traditional web surveys or have more specialised application such as brand perceptual mapping techniques and use with specific respondent groups including established internet panels and young people.

## Keywords

Graphical scales, usability, engagement, web surveys, click and drag, brand / perceptual mapping, gap analysis, numerical scales, descriptive scales, visual control, Internet panels.

## 1. Background

Following an initial rush of enthusiasm to exploit the new technology and explore new ways of reaching respondents, most web questionnaires published are simply electronic equivalents of their paper counterparts. Whilst they often incorporate sophisticated control of the interview by using routing, text-substitution, answer masking and other techniques, respondents' response controls have largely remained unchanged. Multi-choice questions are represented either as lists of check boxes or radio buttons (or sometimes a mixture of the two). Open questions typically are characterised by a text box into which the respondent enters their reply.

At the other end of the survey process, survey results presented to lay audiences often make use of charts and pictures and other graphical elements. For example, ratings may be shown on a graphical scale, proportions of two parts may be shown as a single-element stacked bar chart and aggregate respondent perceptions may be shown on a multi-dimensional map.

If the presentation of results is aided by the construction of a suitable graphical image, then why should not the interview process be similarly improved by using suitable graphical images as a means of inputting response data? The purpose of this paper is to challenge the current status of web survey design and to explore the added value that graphical control scales can have on the survey process.

## 2. About graphical control scales

The main motivation for researchers to use graphical control scales in surveys is to increase respondent involvement in repetitive and low interest surveys through the incorporation of interactive game-type elements (Luetters et al, 2007). Some researchers (for example, Couper et al, 2006 and Luetters et al, 2007) have observed that the use of interactive visual analogue and graphical control scales in web surveys can result in higher rates of non-completion, higher rates of missing data and longer completion times compared to standard radio button scales.

Our approach was to develop experimental surveys to explore the use of a variety of graphical input controls on web surveys conducted with a UK based internet consumer panel. We specifically wanted to try a number of different scale types in the same survey to see if we could detect a consistent non-response bias. We were also interested in the possibility that respondents may find our questionnaire more engaging simply because we had used graphical controls throughout.

Controls we explored included:

- Abstract sliders with and without reflected numerical values – stepped sliders corresponding to single-response choice questions; continuous sliders corresponding to open quantity questions.

- Slider controls such as a grab-able thermometer meniscus – corresponding to open quantity questions.

- Multi-dimensional mapping graphics onto which the respondent drags icons representing items (brand logos), dropping them at appropriate locations – corresponding to batteries of open or closed-choice grids.

All were implemented using standard JavaScript technology and thus did not require the downloading of any special applets or controls onto respondents' computers.


Two important considerations were taken into account in the survey development (as seen in Figure 1). Firstly the need for 'no reply' options on the graphical scales. This is a critical design element that is often overlooked in graphical control scales. Secondly the need for some guidance directions for respondents on the use of the graphical control scales. The team had no prior understanding of how intuitive the use of graphical scales would be to an established internet consumer panel.

Figure 1: Graphical scale showing 'no reply' start position and guidance on scale use

**Your beliefs**

*Move the pointer on the scale upwards for agreement and smiley faces appear or downwards for disagreement and sad faces.*


**Please state how much you agree or disagree with each of the following statements:**

I care about food miles when I buy products  

---

I think that global warming is NOT a real issue  

---

I believe that supermarkets should stop giving customers free plastic bags  

---

I feel that food manufacturers put too much packaging on their products  

### 3. Methodology of the trial

Two web surveys were designed in order to compare the relative merits of graphical image-based web survey ratings against those of traditional standard web survey scales. Environmental responsibility was chosen as the topic of the survey because of its current high profile within the media and the possibility to use a variety of scales to explore many aspects of consumers' claimed habits and attitudes and their perceptions of brands.

A script for a standard ten minute survey was produced. Following that a half day project team meeting was held in order to identify ways in which the survey questions could be designed as interactive graphical control scales.

Dimensions that were explored as part of the study were:

Engagement - how motivated respondents are to interact with the graphical components.

Usability - how easy and intuitive graphical control interfaces are for respondents to use.

Accuracy - comparison of results with those produced using traditional web survey scales.

A number of demographic questions covering dimensions which the team felt might make a difference to how respondents interacted with the graphical control survey were included in both surveys; these were age, gender, hours spent on personal internet use per week and level of qualifications achieved.

A web developer and an image designer worked up the initial drafts of the questionnaires which were then subject to a number of revisions as the team's understanding of graphical scales developed.

The links to each survey were sent simultaneously by e-mail to equivalent samples of an internet consumer panel. The samples were composed of 300 respondents per survey and matched for age and gender. The rate of returned responses was the same for the two surveys over the two days/nights that they were in field. The fieldwork was stopped when 112

responses to each survey were received. All surveys were checked and those that were very incomplete or had suspect answers were omitted, resulting in a total of 105 responses from each survey going forward for analysis.

## 4. Discussion of results

The results are discussed under the main headings of engagement, usability and accuracy.

### 4.1. Engagement

Although the sample of the respondents was matched for age and gender, the actual profiles of the completing respondents was skewed for age (Table 1). The graphical survey appears to be very appealing to the 25 to 34 age group and of lesser interest to 17 to 24 year olds. Over three quarters (20/24) of the 25 to 34s who received the graphical survey responded compared to only just over a third (9/24) who were sent the standard survey. There was a lower response to both surveys for the younger age groups than for the older age groups. This lower response rate is more marked in the case of the graphical control survey where only a third (7/21) of the 17 to 24 year olds responded to the graphical survey compared to close to half (10/21) of those who received the standard survey. This goes against one hypothesis of ours that graphical surveys are more appealing to younger age groups. There were also no clear indications in the data that the decision to take part in the survey or not was affected by respondents' gender, level of internet usage or educational attainment.

Table 1: Respondent profiles

Age Bands	Graphical Control	Standard	Panel Sample
17 to 19	1%	4%	4%
20 to 24	6%	6%	10%
25 to 34	22%	8%	16%
35 to 44	17%	18%	20%
45 to 54	20%	27%	24%
55 to 64	7%	10%	6%
65 or more	28%	28%	20%
Gender			
Male	46%	52%	48%
Female	53%	47%	52%
no reply	1%	1%	

In terms of time taken for respondents to complete the questionnaire, the graphical survey showed an average response time of just less than 11 minutes and the standard version as close to 10 minutes (Table 2). Each survey also had one respondent taking just over 70 minutes (72 and 71 minutes) to complete - these responses were removed from the final analyses.

The use of graphical controls in surveys has often been criticised for resulting in longer response times and higher drop out rates (Couper et al, 2006). This view was not however supported by these findings. The minimum time for completion of the graphical survey was 4 minutes compared to 2 minutes for the standard survey and we could hypothesize that respondents are spending extra time reading the brief guidance notes on how to use the scales. They could also be spending this small amount of additional time considering their responses because they are more engaged with the subject of the survey when questions are presented to them in a colourful and interactive way.

Table 2: Statistics of response times (minutes)

Survey	Mean	Mode	Median	Minimum	Maximum
<b>Graphical</b>	10.8	6.0	9.8	4.3	34.2
<b>Standard</b>	9.7	4.3	7.5	1.7	27.4

Some basic indices of engagement were asked directly of respondents and overall the graphical control survey was scored significantly higher than the standard survey (95% confidence level) for:

The subject being of interest (87% scored 'yes' for the graphical survey versus 76% for the standard survey)

The question style being enjoyable (86% scored 'yes' for the graphical survey versus 72% for the standard survey).

Those aged 65 or older were less likely to find the question style of the graphical survey to be enjoyable (i.e. 77% compared to 89% for those aged below 65). Of the 5 respondents who answered 'no' when asked if they found the question style enjoyable, 4 were aged 65 or above and all 5 were male. These negative scores for graphical control scales accounted for the views of 4 out of the total of 14 male respondents aged over 65 who had responded to the graphical control survey.

A significantly higher number of women (93%) rated 'yes' when asked whether they found the graphical control question style to be enjoyable compared to men (77%). This difference was not observed for the standard survey.

Scores for these engagement indices across respondents' personal internet usage levels (hours/per week) showed no significant differences within either survey. There were also no significant differences for age, gender or qualification classifications for the standard survey across all questions.

The final question on the surveys asked respondents to include any open text comments they had on the survey itself. Some 35 (33%) respondents from the graphical survey added a comment and only 21 (20%) from the standard survey, these figures again demonstrating a greater level of respondent engagement with the graphical survey. These comments were evenly provided by males and females across the age groups. An analysis of the comments is shown in Table 3 below.



Table 3: Analysis of open comments covering engagement topics

Comment Classification	Graphical	Standard
Interesting / entertaining / enjoyable style	15%	5%
Educating	6%	1%
Relevant topic / interesting subject	4%	3%
Comments about recycling / green issues in general	3%	2%
Importance of using the results	1%	-

Overall the comments conclude that there was a high level of interest and engagement with the graphical control survey questions and the subject area. The top four comments made by respondents from the graphical survey were firstly that it was an interesting/enjoyable/entertaining style (15%). Six of these 16 comments were from those aged 65 and over showing some level of appreciation of the graphical design by this age group. The second most commented topic was that the survey was educating (6%) and made them think about the topic. All the comments about educational value were made by women. Thirdly respondents described the survey as a relevant/topical/interesting subject (4%); again, these comments were made only by women. Some of the actual responses that demonstrate the main views are shown in Table 4.

Table 4: Engagement open text comments made about the graphical survey

**Some of the engagement advantages of the graphical control survey were seen to be:**

#### **Interesting/enjoyable/entertaining style**

"I thoroughly enjoyed answering the questions. It was an original and entertaining way to do it. Hope that more surveys come along this way."

"I have never done a survey which used these types of question styles before, even though I find the subject area interesting the styling also helped maintain interest."

"Fantastic interactivity one of the best I've taken part in."

"I love this type of survey; it makes it so much more enjoyable."

#### **Educating**

"This is a very important matter and we need to educate people!!! So this survey is a great thing to get out to people!"

"I think it's a good thing and really gets you thinking more about our everyday actions!"

"This survey has certainly made me think about what our family do and don't do to help the environment. It has actually made me feel quite guilty! Time for some changes I think!"

## **4.2. Usability**

Although the standard survey was scored slightly higher for being easy to respond to (91% versus 89% for the graphical survey) and for being quick to load (97% standard versus 92% graphical) these differences were not significant.

When asked about whether they needed to refer to the scale use guidance notes on the graphical survey, over two thirds (73%) of the respondents answered either 'yes' or

'somewhat'. Just over a quarter (27%) found the scales to be intuitive with no need for them to refer to the explanations.

Trends in the data show that respondents with qualifications from university/college were more likely to refer to the explanations on how to use the scales (43% said 'yes' of those with university/college qualifications (59) compared with only 24% of those without (46)).

Scores for these usability indices across respondents' personal internet usage levels (hours/week) showed no significant differences within either survey.

An analysis of the last question on the surveys which asked respondents to write any comments that they had about the survey (Table 5) demonstrated the importance of guidance notes for the graphical survey (mentioned by 4% of respondents). These were all respondents who had received college/university qualifications.

Table 5: Analysis of open comments covering usability topics

Comment Classification	Graphical	Standard
Importance of the guidance notes	4%	-
Good survey / good design	2%	6%
Difficulty / dislikes	2%	1%
Specific about scales (negative)	2%	5%
It's easy	1%	-

It is of interest to note that the standard survey had more negative comments made about the scales used than the graphical survey. These negative comments all referred to concerns that would also have been of relevance to the graphical survey such as the fixed format postcode question.

Some of the actual responses made that demonstrate the main views on the usability for the graphical survey are shown in Table 6.

Table 6: Usability open text comments made about the graphical survey

**Some of the usability disadvantages of the graphical control scales were described as:**

Need for guidance notes	Difficulties/dislikes	Specifics about the scales (negative)
"It did confuse me at first so the explanations helped."	"Slightly more complicated at times than other surveys."	"The 'temperature bulb' approach offered no improvement over older established methods."
"It took a moment or two to get used to the style. The guidance was very helpful."	"I feel that the style of the questionnaire was infantile - almost imbecile! As an experienced statistician I cannot see how any reliable results can be expected from it."	"On the two axis grid question there was no option for brands I was unaware of so therefore had to try and squeeze them all into a neutral space in the middle which is not such a good judge of true responses."
"Although some questions weren't immediately obvious about how to answer them, the accompanying description		[Note: respondents did not require to

helped."  
 "You do have to be careful  
 though to ensure you  
 understand what the scaling  
 means."

'click and drag' all the images. This  
 clearly demonstrates the importance  
 of highlighting 'no reply/don't know'  
 areas on graphic scales]

We gained a general view of the popularity (or 'easiness') of each question by undertaking an analysis of 'no replies' to each comparable question. This analysis demonstrated that overall there is not a great deal of difference in reply rates across the majority of questions within each survey. The majority of scales were completed by 98% of respondents and above. However there was a lower reply rate at 93% and 95% for two of the graphical scales and in each case 100% of respondents completed the comparative questions in the standard survey. Consideration should be given to make these questions more intuitive and simple to use for respondents in the future. The more difficult scales were:

Dual variable slider bar (100/105 respondents completed)

When asked how much volume of their household waste they recycled the graphical control survey provided respondents with a sliding bar that automatically calculated the relative percentages of recycled and non-recycled waste. The percent figures chosen were shown on the image (Figure 2).

Figure 2: The dual variable slider bar

**What proportion (volume) of your overall waste would you say is recycled and what is put in the refuse bin?**  
*Click & drag the scale to show the relative amounts.*



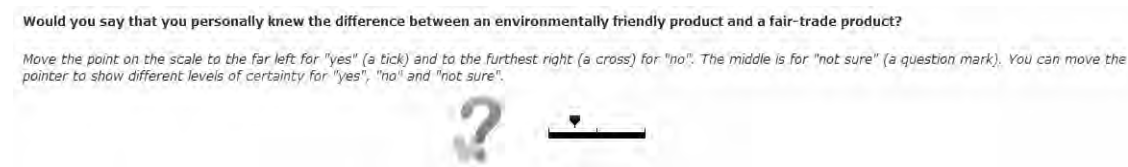
In contrast the standard survey required respondents to type the quantity of '% recycled' in a numeric box. The results from the standard survey show that all but two of the respondents typed in figures for '% recycled' that ended in either 0 or 5. As the graphical control slider bar enabled respondents to allocate % scores in increments of 1, the amounts inputted using this scale were more varied, however it was also more difficult to slide the bar accurately to a specific amount.

Opacity faded slider (98/105 respondents completed)

The opacity faded slider on the graphical control survey provided a gauge that enabled respondents to choose 'yes', 'not sure' or 'no' and any position of certainty/uncertainty across this scale (Figure 3). They were asked to use this scale to show whether they knew the difference between an environmentally friendly product and a fair-trade product. The levels of certainty/uncertainty chosen coincided with overlaid images of a tick, question mark or cross becoming either bolder or fainter. Analysing the scores showed that 23% chose 'yes' and no one chose 'not sure' or 'no'. The rest of the results were allocated in varying degrees of certainty between 'not sure' and 'yes' (53%) and 'not sure' and 'no' (16%). This contrasts with the standard survey where respondents had to choose one of the three fixed answers, here 52% ticked 'yes', 36% 'not sure' and 11% scored 'no'. The level of certainty of those feeling strongly that they do know the difference is clearer in the fixed choice option where 52%

ticked 'yes' however the truer picture is likely to be reflected in the graphical scale where only 23% were completely certain that they meant 'yes'.

Figure 3: The opacity faded slider



#### 4.3. Accuracy

In comparing the use of the scales across the two surveys there is an indication that respondents are more likely to use the upper end of the scale in some graphical control questions compared with the comparative standard scales. This results in a trend for higher mean scores in the graphical compared with the standard scales. Similar observations were made by Thomas and Couper, 2007 who found that for interactive visual analogue scales, end-anchored scales yielded higher mean values than fully anchored scales.

Examples of higher mean scores for the graphical scales are seen in the purchasing choice 'thermometer gauge' and the two dimensional brand perceptual mapping exercise.

The importance scale used for the 'thermometer gauge' question was an anchored sliding scale. Anchor point descriptions showing only when respondents moved the pointer along the scale (Figure 4). The comparative standard scale used was a fully anchored drop down box with scale descriptions.

The graphical control scores for the sliding 'importance' gauge resulted in respondents recording significantly higher mean scores for their importance to purchasing decisions than for those who used the drop down box in the standard survey (Table 6).

These results suggest that respondents using the graphical scales are more motivated to using the higher end of this graphical scale than those using the standard drop-down box scale. In effect they are forced to move the meniscus across all points of the scale in order to view the scale description at each level.

Figure 4: Grab-able thermometer meniscus

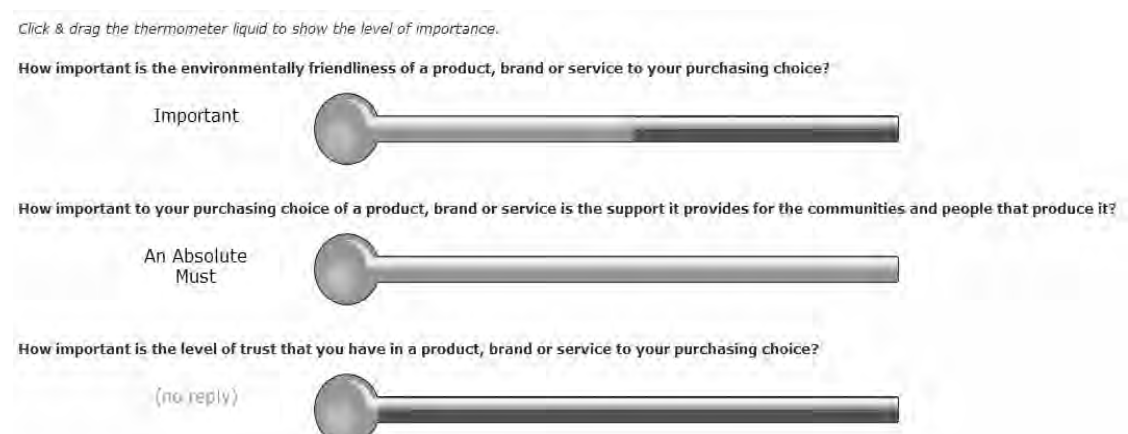


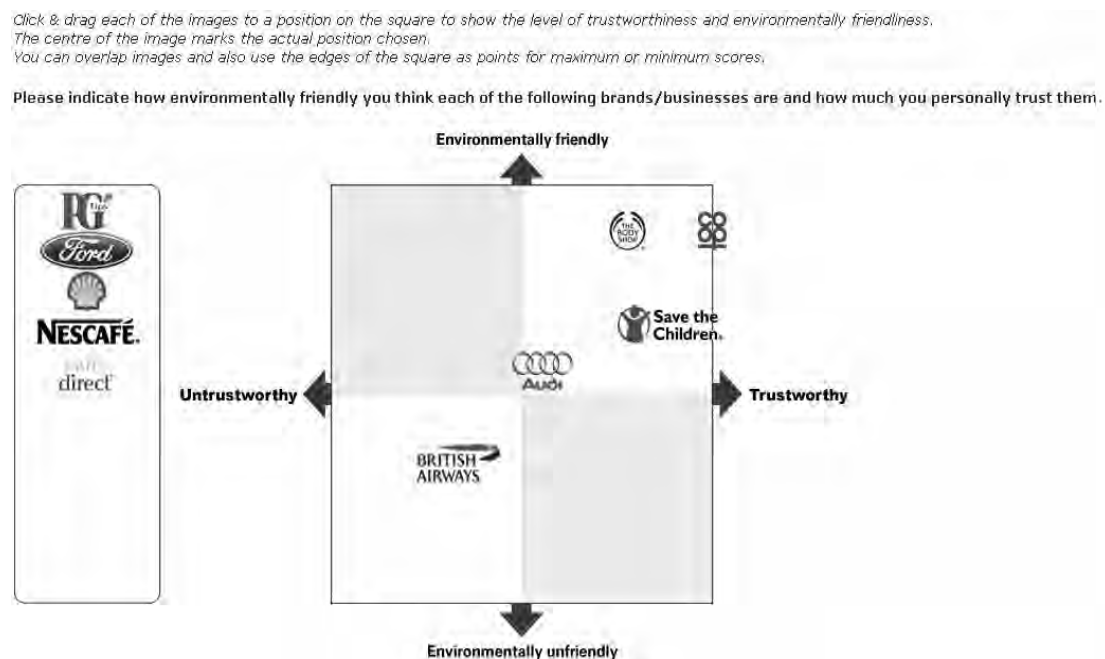
Table 6: Comparison of mean scores for importance in purchasing choices

Rating	Graphical	Standard
Environmental friendliness	3.3	2.9
Support for producer communities	3.2	2.9
Level of trust	3.8	3.5

*Note: A mean score was calculated by assigning a score of 5 for 'an absolute must', 4 for 'very important', 3 for 'important', 2 for 'of little importance' and 1 for 'unimportant'.*

An analysis of the perceptual brand mapping for the respondents using a 'click and drag' graphical control two dimensional map (Figure 5) versus a standard radio button scale also showed that there was a tendency for those using the graphical control scale to use higher levels in the scale than those using the standard scale (see Figure 6).

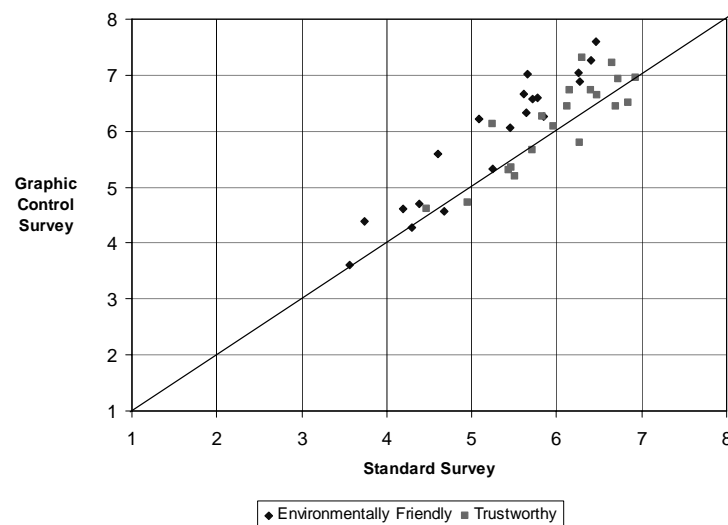
Figure 5: Interactive brand mapping



The two dimensional brand mapping space used end-anchored bipolar scales, with the axes intersecting at the neutral midpoint. The standard 10 point radio button scale was end-anchored (e.g. completely untrustworthy/completely trustworthy) with an anchored neutral mid-point (e.g. neither untrustworthy nor trustworthy).

Whether the observed increased use of the higher end of the scale in these graphical scales is a more accurate reflection of the true value of a score compared to standard scales, or whether it is introducing a bias is open to debate and requires further work.

Figure 6: Score comparisons for graphical brand mapping and standard brand rating



## 5. Further work

We would see that our project has only scratched the surface of the possibilities that graphical input methods for questionnaires promises. There is, consequently, a vast amount of experimental work that could be carried out. In order to keep the work focussed we consider that there are two overall goals:

Reduce potential set-up time, probably by introducing more standardisation into controls.

Provide guidelines for questionnaire designers on the application of these controls.

As part of this we would specifically want to:

Compare the use of different types of graphical control in the same situation with the aim of coming up with some heuristics as to which form of graphical control is appropriate when.

Compare use of abstract images with the use of subject-related images. For example, if one is asking a question relating to proportion of time, which is a better way to represent time passing: an abstract scale, a container indicating percentage full, or a clock face?

## 6. Conclusions

We developed experimental surveys to explore the use of a variety of graphical input controls on web surveys. Controls we investigated included continuous and stepped sliders, and controls that involved the respondent dragging and dropping brand logos onto one- and two-dimensional image maps.

One fairly significant point to be made about graphical controls is how to design the control with a default response of 'no reply' and making it clear to the respondent that a response has not yet been given. This is especially true of slider controls where the slider needs to be somewhere in order for the respondent to recognise it. A similar problem arises with the use of image maps and the solution here seems to be to allocate a special place to where respondents drag icons to indicate 'don't know'.

Another potential problem is the novelty of graphical input methods for survey respondents. In our opinion, care does need to be taken to ensure that respondents know how to interact, either by making the control intuitively 'obvious' or by including appropriate instructions.

Currently the production of a graphical control survey requires more investment in development time and closer working with the developer, designer and researcher, however as graphical control scales become more standardised the additional time required on survey development will reduce.

Overall, we were encouraged that many respondents across all ages from the internet consumer panel found the graphical inputs acceptable, enjoyed completing the questionnaire and were looking forward to more surveys of this type in the future. The increased level of respondent engagement with the subject area through the use of interactive and colourful graphical scales is an important advantage over standard scales. We found no significant disadvantage of graphical scales for response rates or completion times. We therefore consider that the use of graphical controls such as those used in our survey could improve response rates and levels of interest from established consumer internet panels.

## References

- Couper, M.P., Tourangeau, R., Conrad, F.G., Singer, E. (2006), "Evaluating the effectiveness of visual analog scales: A web experiment", *Social Science Computer Review* 2006, 24
- Luetters, H., Westphal, D., Heublein, F. (2007), "*SniperScale: Graphical scaling in data collection and its effect on the response behaviour of participants in online studies*", German Online Research Conference (GOR)
- Thomas, R.K., Couper, M.P. (2007), "*A comparison of visual analog and graphic rating scales*", German Online Research Conference (GOR)

## Acknowledgements

The rest of the project team, all from Snap Surveys Ltd: Ralph Sutcliffe, Senior Researcher, Andrew Checkley, Web Software Developer and David Mitchell, Web/Image Designer.

## About the Authors

Dr Nicola Stanley is Managing Director of Silver Dialogue Ltd., a stakeholder engagement consultancy. She is a market research and communications specialist, with specific interest in novel approaches to motivate and empower individuals and groups through dialogue techniques. She has over 18 years' of experience gained through consumer/business research and corporate communication roles. She can be contacted at Silver Dialogue Ltd., PO Box 6470, Rushden, Northamptonshire, NN10 0ZW, UK, or by email [nicola\(at\)silverdialogue.com](mailto:nicola(at)silverdialogue.com).

Dr Stephen Jenkins is Technical Director of Snap Surveys Ltd. and is responsible for the direction and development of Snap survey software. He has over 30 years' experience in the design, development and use of survey software. He is a founding member of The Triple-S Group and an author of the Triple-S survey interchange standard. He can be contacted at Snap Surveys Ltd., Mead Court, Thornbury, Bristol BS35 3UW, UK, or by email [sjenkins\(at\)snapsurveys.com](mailto:sjenkins(at)snapsurveys.com).

Copies of the full surveys can be requested from the authors.

# Factors Affecting Return Rates to On-Line Surveys

John Cooper

## Abstract

This paper reports on the return rates achieved in a number of internal surveys performed at Bath Spa University over the last twelve years and discusses factors that may have influenced these rates. The surveys examined have used paper questionnaires, on-line web-based questionnaires and a mixture of the two, all produced with the help of the SNAP survey analysis software package.

Factors that are considered include the age, gender and IT (Information Technology) skill level of the respondent. The use of reminders and financial incentives is also examined. The time scale for respondents to reply to on-line surveys is analysed, in particular for the timing of reminders. The paper concludes by presenting a list of factors that appear to influence the return rate and a list of those factors that appear not to have an influence.

## Keywords

Survey response rate; on-line; reminder interval; IT skills, interview mode comparisons

## 1. Introduction

### Historical background

Research in the 1960's indicated that postal surveys could achieve return rates of between 75% and 96%, depending on the number of follow-up letters sent (Scott, 1961). Scott had used three separate follow-up letters.

Research performed by the author in 1986-7 produced an overall return rate of 51%. The respondents were all teachers in a selection of secondary schools. The return rate from individual schools varied from 29% to 80%, with an inverse relationship (correlation coefficient -0.58) between return rate and number of teachers in the school (Cooper, 1989).

The survey methodology used a local contact in each school to distribute the paper questionnaires and encourage completion. Return was via a "ballot box" located in the Staff Room (or equivalent) in each school. It is suggested that the larger schools with split sites and other internal divisions were more difficult to manage even with an on-site facilitator.

Research elsewhere has indicated an increased return rate through financial inducement. A 25c coin attached to one sample of postal questionnaires achieved a return rate of 83% compared with 63% of those not sent the 25 cents. It should be noted that the researchers also



employed a single reminder letter and up to two telephone reminders (Amour and Bedell, 1978).

The appearance of the questionnaire has also been cited as a factor affecting return rates: 85% for questionnaires “attractively” printed compared to 62% where they had been plain copied (Gillespie and Perry, 1975).

### Current Research

The almost exponential growth of the use of the internet has seen the mechanism for delivering surveys at Bath Spa University move from paper to electronic. As with all surveys, achieving a high return rate is a common goal. The choice of a sample frame could also be a factor.

The Bath Spa University surveys differ in some respect in that the target audience is known and reachable through email. In general no sampling is needed.

So what factors affect their willingness to participate in surveys? The use of the SNAP survey analysis package throughout the whole research cycle has simplified longitudinal comparisons. Recent versions have enabled better targeting of the questionnaire to match the audience.

## 2. Return Rates from Captive Audiences

### Induction Sessions 1996 – 2000

Induction sessions were held for new undergraduates to introduce them to the computing facilities. The sessions were scheduled as part of a comprehensive week-long programme and every student was allocated a specific session.

During the session a paper questionnaire was issued to be completed during the session. All questionnaires were collected by hand before the students left the room.

These questionnaires asked inter alia, students to self assess their own IT skills in a number of separate areas. Three typical areas were *use of a mouse*, *word processor* and *spreadsheet*. As technology developed the topics were broadened to include *email* and the *internet*.

To provide a reasonable degree of comparison, the possible skill levels were limited to three choices: “no ability”, “some ability but not confident” and “confident”.

The observed return rate was virtually 100% of those present. Subsequent analysis indicated this was only 70% of those expected to attend!

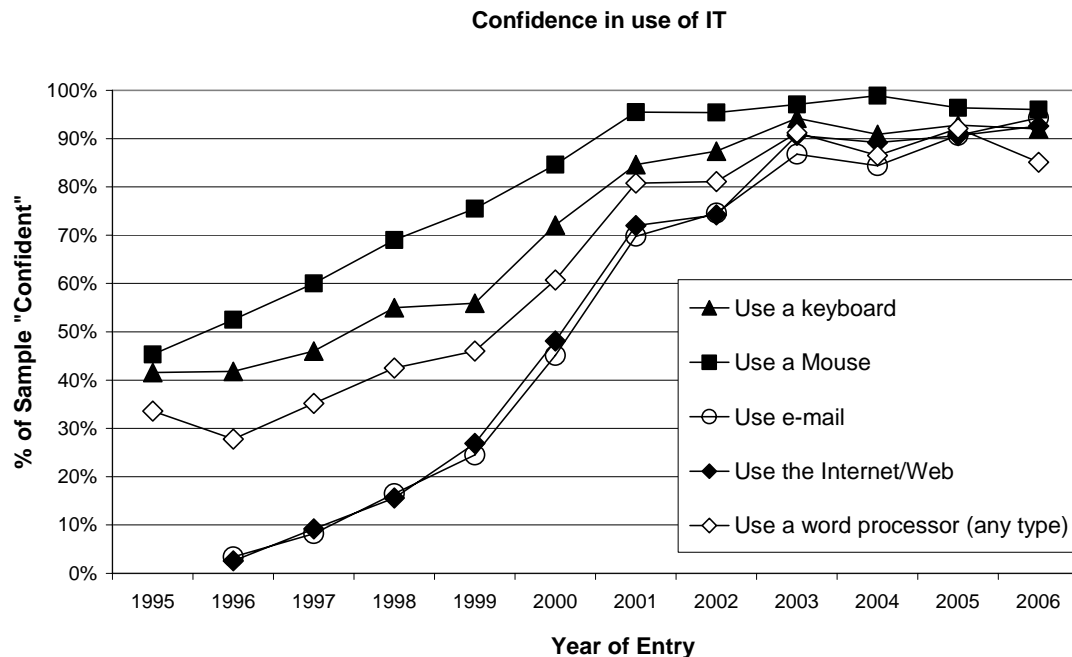
## 3. The Growth of IT Skills in the population

### Induction Sessions 1997 – 2006

After five years the delivery mode was changed. The number of new students made individual sessions impractical but all students were being allocated University email addresses. So in the next two years the questionnaire was emailed in html format to lists of new students.

Tighter restrictions on the format of emails acceptable to the various filters and firewalls led to a change of delivery in 2003. The survey was now mounted on a “hidden” web page and invitations containing a hyperlink to the survey were emailed to the respective lists.

The chart in Figure 1 shows that, in 2006, over 90% of respondents rated themselves as “Confident” to use a mouse, keyboard, email and the internet/web. Word processing has ‘hovered’ around the 90% mark in recent years.



**Figure 1: Confidence in use of IT**

### Age and Gender Profile of the Respondents

The nature of courses offered at Bath Spa University has always resulted in a greater proportion of female students. The respondents to the survey reported here consistently mirrored the overall distribution at 25% male: 75% female. The University actively encourages widening participation which results in a healthy proportion of older students. This is reflected in the age profile of respondents, where about 20% were aged 30 and over and a further 35% were in the 20-29 age range.

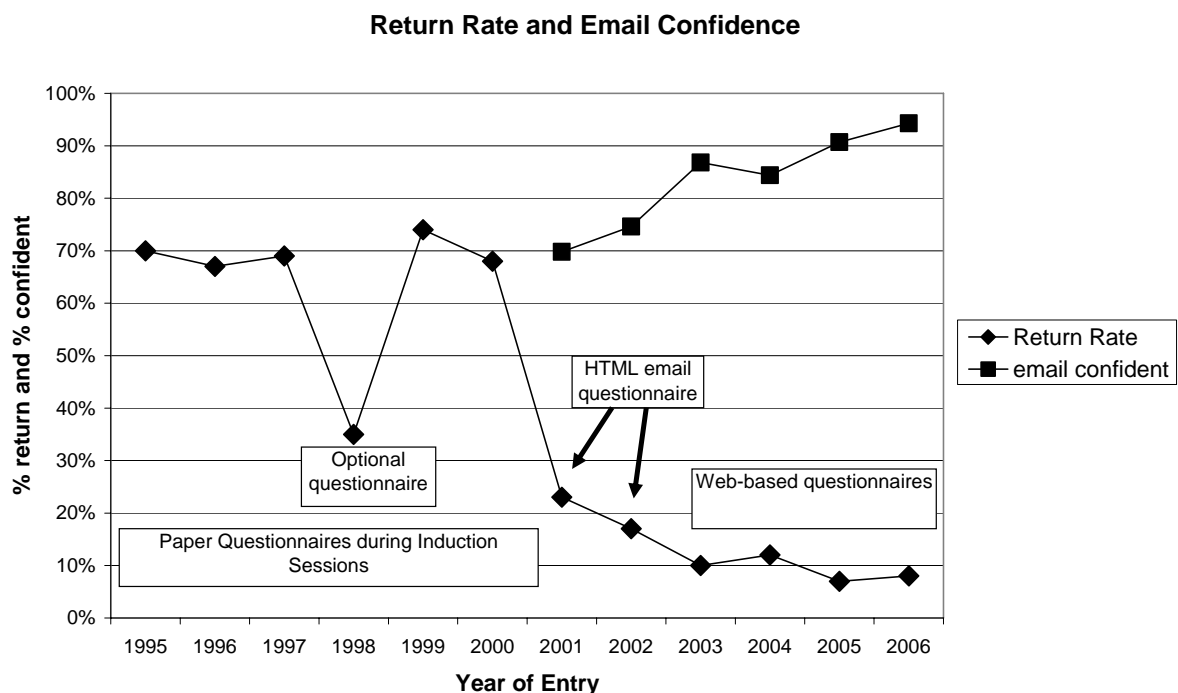
A recent statistical report released by the European Commission’s Eurostat Press Office (Eurostat, 2007) indicated that just 19% of women aged 16-74 and 35% of men in the same age group in the UK were considered to have “High” computer skills, based on six separate tasks including “use basic arithmetic formula (add, subtract, multiply, divide) in a spreadsheet”. By way of comparison, the “confident to use a spreadsheet” data for the current survey is 35% overall, but 53% of males and just 23% of females. When analysed by age group the greatest percentage of “confident” was in the 20-29 age group (42%), with 35% of the “under 20” and 32% of the “30-39” groups choosing this option.

The conclusion is that whilst the Bath Spa University population does not mirror the population as a whole there is sufficient spread of age and IT skills to make tentative extrapolations from the results to be reported in the following paragraphs.

### Variation with Return Rates as Technology Changed

The chart in Figure 2 shows the estimated return rates for the whole period of the research. The annotations indicate the method used to deliver the questionnaire. In general the closed Induction sessions achieved almost a 70% return – though as the sessions were classified as ‘compulsory’ one would have hoped for nearer 100%!

In 1998 the Induction sessions were run as ‘open house’ with the questionnaires available for students to complete as they moved around the open access computer rooms. This experiment was not successful and the formal sessions were reintroduced until increasing student numbers made them impractical.



**Figure 2: Overall return rate by year of survey with email confidence for comparison**

### Email Questionnaires – Novelty Value?

The use of an email questionnaire (2001, 2002) achieved a response comparable with the optional survey in 1998. However Figure 2 shows a steady downward trend suggesting that even the novelty of having a University email account (and perhaps for many an email account for the first time) could not encourage a better response. The chart also shows an increasing confidence to use email amongst the respondents. Subsequent web-based surveys have shown that achieving a double figure percentage return is becoming increasingly difficult.

## 4. Do IT skills affect the Response rate?

### Comparison between on-line and paper response rates

A concern was that those with low IT skill levels were being excluded from the on-line surveys. A small scale project was undertaken in 2002 to check for the effect of IT skill on completion rates.

A number of first year modules were selected and paper questionnaires issued to all present in a series of seminars.

The paper questionnaires contained the same self-assessed questions on confidence in a total of ten separate areas of IT. From their responses a total 'skill score' was calculated ranging from zero ("no ability" in all ten) to twenty ("confident" in all ten).

A total of 112 responses were received, this being about 6% of the overall first year intake. Of the responses 9% had also done the on-line survey. When the Skill Scores were calculated the "had done the email survey" group averaged 12.6 whilst the "had not done" averaged 13.9, with no significant difference between these scores. There were also no significant differences between the scores achieved according to gender or age group.

## 5. Do Financial Inducements Work?

Bath Spa University undertook a wide-ranging satisfaction survey of all its students in Spring 2005. To encourage participation and completion three iPods (or cash equivalent) were offered through a random draw. The overall response rate was an estimated 13%, which compares favourably with the 7% achieved through the IT Skills survey held the same year.

The survey was available both on-line and through paper questionnaires available in both main Campus Libraries. 13% of the total responses (85 out of 658) were from paper questionnaires. However these represented just 3% of the returns from the main campus but 68% of those from the Art & Design site (both sites showing an overall return of 13%). It was known that Library staff at the latter site had a closer working relationship with their students and that they had promoted the survey at every opportunity. This appeared to compensate for the previously observed reluctance of students at this campus to complete on-line surveys. Work reported earlier in this paper had also suggested that personal promotion had a greater effect in a smaller institution - the Art & Design site comprised just 14% of the total student numbers.

It is concluded that the financial inducement did result in an increased return rate but that the total increase of about 6% was due in the greater part to the prospect of winning an iPod but to a lesser part in the provision of an alternative paper format which was personally promoted.

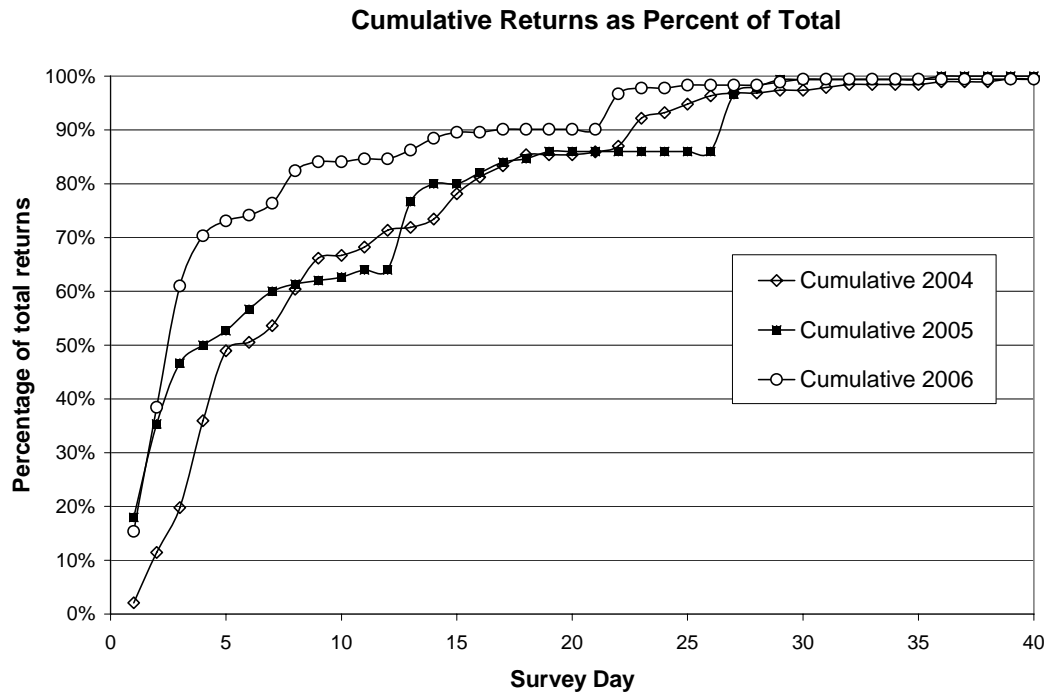
## 6. How long does it take? Are reminders necessary?

### Tracking Returns over Time

In addition to the ease of importing the returns, another valuable feature of both the email and web-based surveys was that the returns were date and time-stamped. This enabled the return rate to be analysed against time.

Figure 3 shows the cumulative return rate expressed as a percentage of the final total for the three most recent IT Skills surveys (shown as "Cumulative 2003" to "Cumulative 2006"). It can be seen that all the surveys follow a similar pattern with 50% of the total returns being received within the first week (in electronic surveys of students weekends are just as productive as weekdays!).

The chart also shows steps in each trace where a reminder had been sent. The effect of reminders is discussed in the next section.



**Figure 3: Returns as a function of time**

With mail surveys reminder times were measured in weeks – with on-line surveys it has become days. This data suggests that more than three-quarters of the returns will be in within two weeks, so reminders should start to be sent from about this time.

#### Reminders – How Many and How Often?

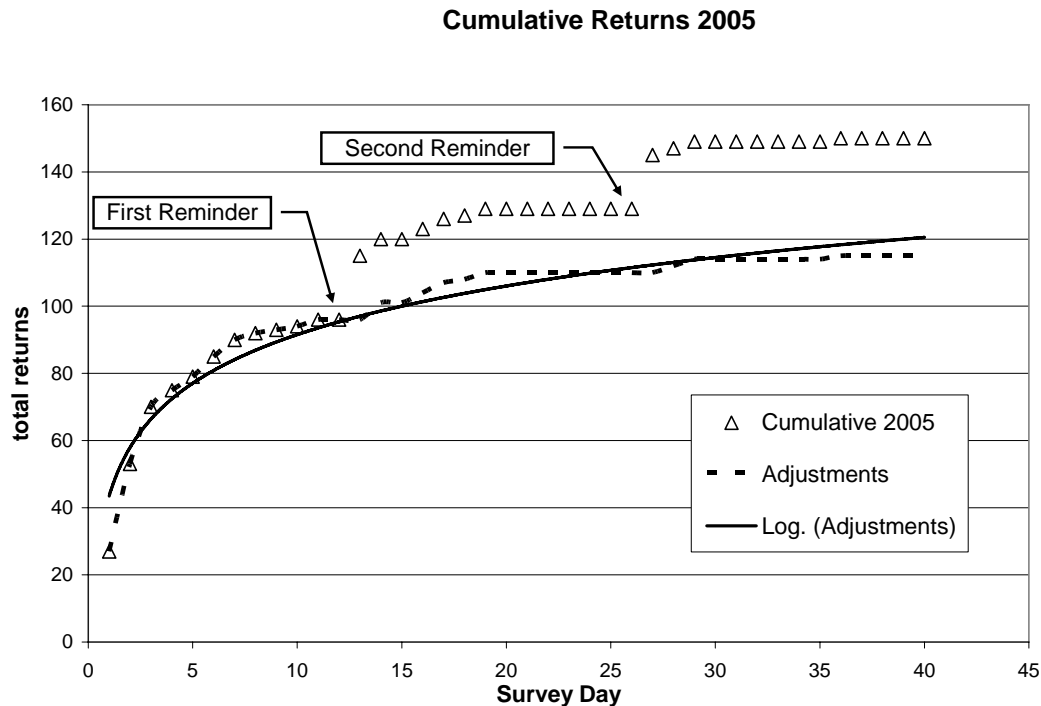
Research has confirmed that giving reminders do result in an increase in the total completion rate. But when should reminders be given? How many reminders should be used? How does the “new technology” influence the answers to these questions?

Reminders have been used in all the electronic surveys reported in this paper. There has been no set pattern of issuing these – it has been more an exercise in opportunism.

The effect of the reminders is illustrated in the typical chart shown in Figure 4. Here the actual cumulative returns for the 2005 Entry survey are plotted as separate points against the survey day (day one being the day the invitations were sent out).

The effect of each reminder is calculated by taking the difference between the total returns on the days before and after the day the reminder was issued. These differences are subtracted from the cumulative total for all subsequent days to give the “adjustments” data which is plotted as a dashed line.

This adjusted data has then had a logarithmic best fit curve applied, indicated as a solid line. This showed a reasonable match to the adjustments line, supporting the hypothesis that the reminders caused an increase in the total number of responses rather than merely ‘bringing forward’ responses that would have been in received at a later date.



**Figure 4: Effect of Reminders**

The same process has been applied to all the surveys considered here. Though not all show such a good fit to a logarithmic plot they all show a similar pattern. It is proposed that the method of determining the increase as a result of a reminder though not mathematically rigorous can be taken as a general indicator

The following tables summarise the pattern of reminders used in a number of surveys all targeted at Bath Spa University students. The “Reminder Day” is a count from the day the survey was launched.

The “Percent Increase” was calculated from the difference between the cumulative return on the day before the reminder and the return on the day after the reminder was issued by email. This was divided by the total number of invitations sent to arrive at an overall percentage increase.

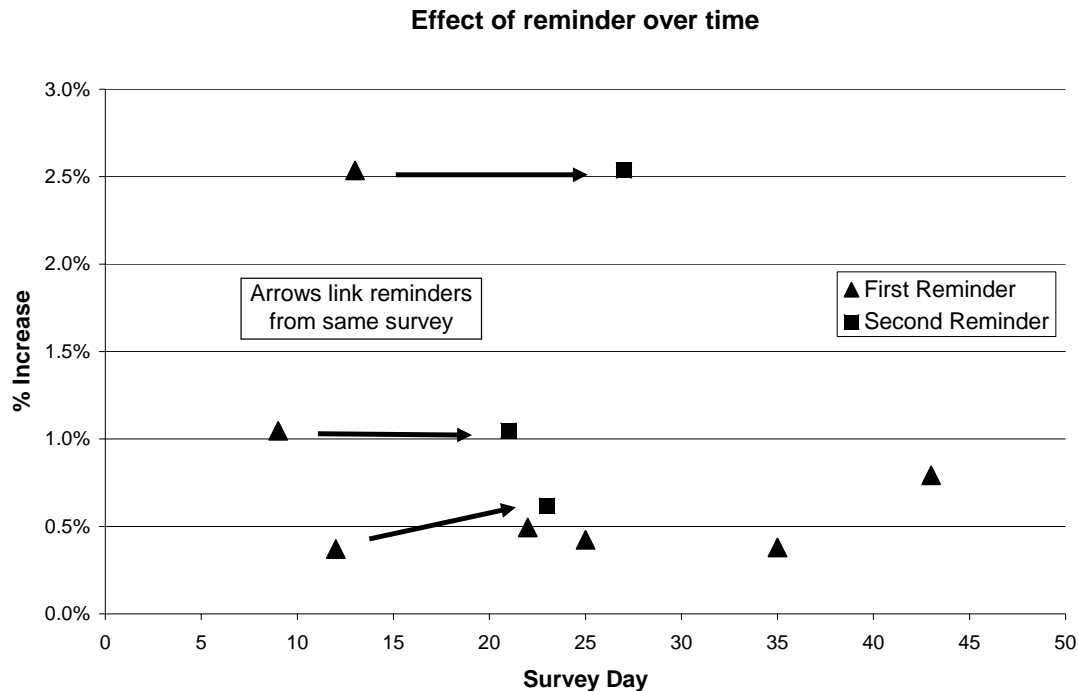
The first table shows those where just a single reminder was given.

Survey Year	Number of Reminder	Reward Offered?	Reminder Day	Percent Increase
2001	First	No	43	0.8%
2002	First	No	55	0.4%
2003	First	No	25	0.4%
2006	First	No	22	0.5%

The second table shows those surveys where a second reminder was given. This table includes the 2005 satisfaction survey (StuSat05) where financial rewards were offered.

Survey Year	Number of Reminder	Reward Offered?	Reminder Day	Percent Increase
2004	First	No	12	0.4%
2004	Second	No	22	0.6%
2005	First	No	13	2.5%
2005	Second	No	27	2.5%
StuSat05	First	Yes	9	1.0%
StuSat05	Second	Yes	21	1.0%

The percentage increases caused by the reminders, as determined by the method described above, have been plotted against the day in the survey that the reminder was issued (Figure 4). The chart differentiates between first and second reminder but there is no discernable pattern (correlation coefficient = -0.2). The average increase from all reminders was 1.0%.



**Figure 5: Increase in response rate due to reminders**

The chart also shows, as an arrow, a link between the data for first and second reminders for a given survey. In two of the surveys both reminders achieved the same increase, whilst in the third survey the second reminder was slightly more successful.

From this limited data it is suggested that neither the time a reminder is given nor whether it is the first or second reminder has any influence on the increased response rate achieved by the reminder.

## 7. Summary and Conclusions

This paper has considered a number of factors affecting return rates to surveys delivered to students at Bath Spa University. It covered the progression from paper questionnaires, though html questionnaires delivered by email, to the current method of web-based surveys with email invitations.

The results show an increasing difficulty to obtain double-figure percentage returns without the use of financial inducements. Reminders with on-line surveys appear to have a consistent effect and are independent of time into the survey. There is also evidence that a second can bring about a similar percentage increase to the first.

### Factors that appear to influence return rates

- Use of a Reminder
- Use of a Second Reminder
- Financial Inducements
- Personal Promotion (in smaller survey centres)



### Factors that do not appear to influence return rates

- Level of IT Skills (but all Bath Spa University students are expected to use email)
- Gender of Respondent
- Age of Respondent (for Bath Spa University students)

### References

- Amour J and Bedell J R (1978) Facilitating Return Rates Of Mailed Program Evaluation Questionnaires, *Evaluation & the Health Professions*; Summer 1978, Vol. 1 Issue 2, p81-88.
- Cooper J W (1989) An Investigation into the Use of Computers in the Curriculum, M.Phil thesis, University of Bath.
- Eurostat (2007) 8 March 2007: International Women's Day A statistical illustration of the situation of women and men in the EU27 [online] Press Release STAT/07/32, Luxembourg: European Commission, available from:  
<[http://epp.eurostat.ec.europa.eu/pls/portal/docs/PAGE/PGP\\_PRD\\_CAT\\_PREREL/PGE\\_CAT\\_PREREL\\_YEAR\\_2007/PGE\\_CAT\\_PREREL\\_YEAR\\_2007\\_MONTH\\_03/3-05032007-EN-AP.PDF](http://epp.eurostat.ec.europa.eu/pls/portal/docs/PAGE/PGP_PRD_CAT_PREREL/PGE_CAT_PREREL_YEAR_2007/PGE_CAT_PREREL_YEAR_2007_MONTH_03/3-05032007-EN-AP.PDF)>. [Accessed 20th April 2007].
- Gillespie D F and Perry R W. (1975) Survey Return Rates And Questionnaire Appearance, *Australian & New Zealand Journal of Sociology*, Oct75, Vol. 11 Issue 3, 71-72.
- Scott C J (1961) Research on Mail Surveys, *Journal of the Royal Statistical Society*, XXIV, Series A, 124, 143-205.

### About the Author

John Cooper *MA, MSc (Oxon), MPhil (Bath), CertEd (FE) (Birmingham), MBCS, FHEA* is Head of IT Training and Support at Bath Spa University. He has been involved with data analysis from his time as a research student at Oxford University in the 1970's. In recent years he has mainly used SPSS, Minitab and SNAP for data analysis and has taught the use of these to undergraduates and postgraduates. He currently uses SNAP to deliver on-line surveys to the University population and supports an increasing number of staff across the University who are also using SNAP. He can be contacted at Bath Spa University, Library and Information Services, Newton Park, Newton St. Loe, Bath BA2 9BN U.K.; tel: 01225 875456; e-mail [j.cooper@bathspa.ac.uk](mailto:j.cooper@bathspa.ac.uk).

# The Effect of Reminder Intervals on Response Rates for Web Surveys

John S. Lemon

## Abstract

The University of Aberdeen has used questionnaire design software for a number of years to produce both paper and web based surveys. The number and frequency of web based surveys to staff and students increases with every year, some are for research purposes while others are to audit the facilities and services of the University.

The view for web based surveys on reminders has always been the same as for paper ones; namely that there should be at least a week between each of the two reminders and that the survey closure should be another week after that. Examination of the pattern of responses to a number of web based surveys indicated that the majority of responses are within 36 hours of the initial invitation, or a reminder, which indicates that people will respond immediately or not at all.

The latest release of the software used provides the opportunity to automate invitations and reminders, with reminders only sent to those who had not finished the survey. It was decided to test whether shorter reminder intervals would enable surveys to be conducted over shorter periods of time without affecting response rates.

## Keywords

Web surveys, response rates, reminder intervals

## 1. Introduction

The University of Aberdeen, like many other education establishments uses many surveys. These may be carried out by the University administration, staff, or students and the survey population may be staff, students or external. The surveys normally fall into one of four main areas:

- Course evaluation – this is almost entirely paper based and is carried out towards the end of every course, or module. It is paper based to ensure high return rates as the forms are handed out at a lecture towards the end of the teaching period and collected in before the students leave the lecture room. This ensures response rates remain high. The completed forms are scanned to reduce staff workload as at least 50,000 course evaluation forms are processed each year.

- Audit of service provision – the move from paper to web based has been largely driven by the increase in the number of surveys required by the University administration to ensure we provide a consistent and increasing level of service to the students in the customer orientated environment. Provision of support services to staff are also assessed in this way.
- Undergraduate elective and dissertation projects – the move has been from paper based to web based for students conducting their project in Aberdeen, while those who go overseas to conduct their research still tend to use paper based forms.
- Staff and postgraduate research – in the past this has been largely paper based but increasing costs and lower response rates have forced a move towards web based surveys.

The move to the web has not always resulted in a new approach to the way in which the surveys are handled, especially those auditing service provisions. In many cases the staff administering the survey transferred the procedures that had worked with paper based projects over to web based ones with little thought as to how the new technology could be improved. One example of this was to use samples of the student population for an audit of one of the services provided by the University administration. When asked why a web based survey was being sent out to only 10% of the potential respondents the answer was that it had always been done this way to “reduce costs”. When it was explained what the actual costs of processing a web based survey were as compared to a postal one then they reluctantly agreed to survey the entire student body. As this approach produced the largest number of responses for many years the decision was made to survey the entire student population in future.

The biggest resistance to change in procedure was the timing of reminders but even though it was pointed out that people respond to E-mails quickly, or not at all, there was a marked reluctance to change the frequency and speed of reminders from that which had worked for paper surveys over the years. The reluctance to change was due to the fact that “we must give people time to respond”. Explaining that that the delays in the response to paper surveys were largely due to the postal service had little effect. In order to effect changes either published or research evidence was required.

A search of the web and published literature produced only one paper or article on response rates and optimal reminder intervals (Crawford et al., 2001). There were many articles on the effect on response rates of other aspects of web surveys; for example salutation (Heerwegh, 2005; Pearson & Levine, 2003), click boxes or drop-downs (Healey, 2007), progress bars (Conrad et al., 2003) or incentives (Cobanoglu & Conanoglu, 2003) but only this one paper on reminder intervals or patterns. In this paper reminder intervals had been examined as one of four aspects in a major study on response rates but in this case the overall length of the survey had been the same for the two study groups with one group receiving an extra reminder shortly after the initial invitation. We wanted to reduce the overall time taken for a survey so needed research data to prove the hypothesis.

There is an unwritten policy within the University administration that “all students” surveys should not be run concurrently, they should not be started within a week of any previous survey and as far as possible should avoid revision and exam periods. It is also recognised that despite the fact many students have access to the web and therefore E-mail during vacations it is not worthwhile conducting any surveys outside term time.

The computing service conducts an annual survey of all students to assess their views on the facilities provided; this is normally carried out during the latter part of the Spring term (mid-late March). In 2006 due to a change in staff and re-organisation of the department the preparations for the survey were delayed and two other service departments within the University started their own surveys. This meant that the computing services survey would have to be delayed until after the Easter vacation and couldn't start until the second week of the summer term. If the reminder intervals were then adhered to the survey would impact on the revision / exam period and therefore could not be run. The alternative proposition to delay it to after the summer vacation was not really be practical as it would conflict with all the "start of academic year" surveys and could not be run until late in the Winter term.

This was not acceptable in terms of getting an accurate picture of student opinions and how they might have changed from previous years. The patterns of responses for previous years were examined to see whether the reminder intervals and the duration of the survey process could be shortened. For the previous two years the computing services survey had been conducted solely on the web using snap and included a number of items of what the developers of snap ([www.snapsurveys.com](http://www.snapsurveys.com)) call "paradata", the two that were relevant to this study were the date and time the respondent started the survey. The file required a little cleaning to remove the extreme values from the data as the version of snap used for these previous surveys had collected the date and time data from the users' machine instead of from a server side script. Almost 2% of the cases had dates set outside the survey dates, some by a number of years.

The paradata had not been available for the earlier years (pre-2004) despite using snap but the results of the analysis of the data for 2004 were interesting as they showed that the majority of responses came within 36-48 hours of the initial invite or a reminder (Fig. 1)<sup>1</sup>.

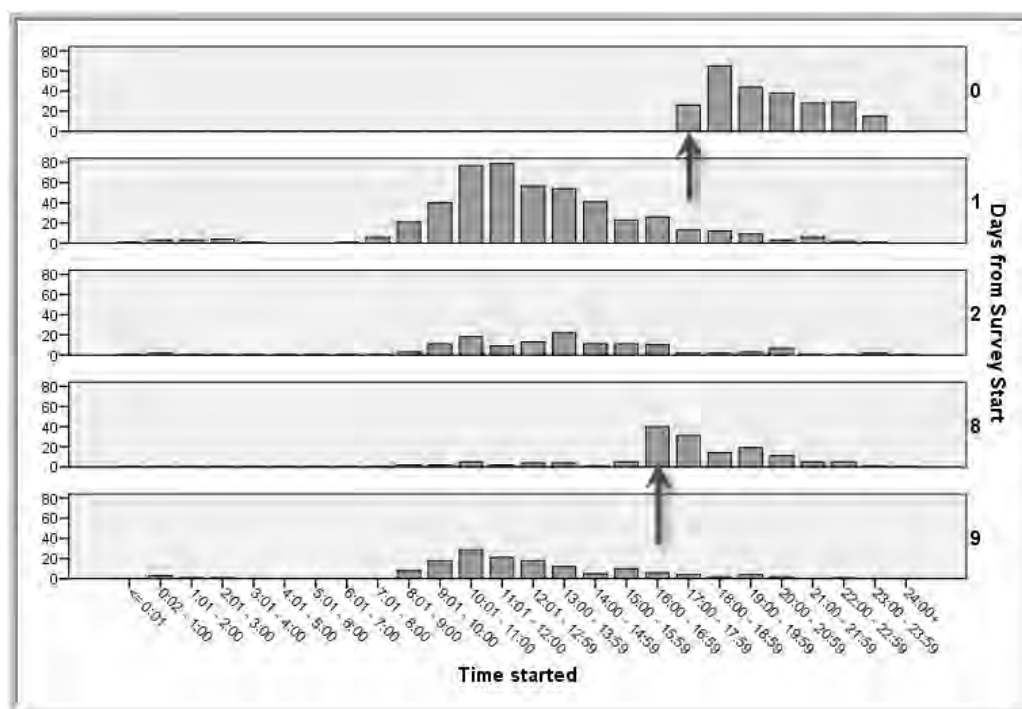


Fig. 1

<sup>1</sup> The arrows indicate when the invite, or reminder, was sent out.

Only the relevant days have been shown in the chart as for all the other days of the survey (days 3-7 and 10-24) there were only one or two responses for each time period. For various logistical reasons the E-mailing of invites and reminders was not undertaken using the facilities within snap and were therefore sent out late in the afternoon of the relevant day.

When the data for 2005 was compared with 2004 (Fig. 2) the pattern of responses was similar thus reinforcing the opinion that responding to web surveys is something that people do immediately on receiving the E-mail invitation, or not all.

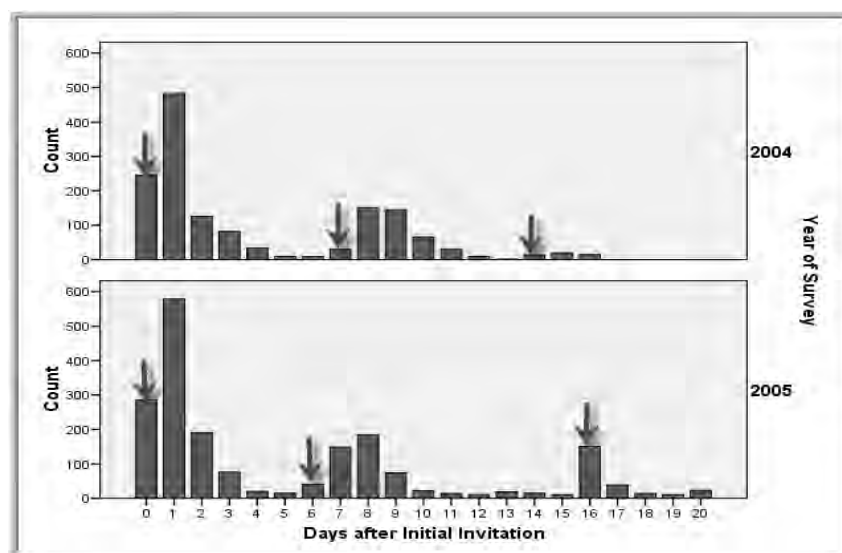


Fig. 2

Based on the observed pattern of responses for 2004 and 2005 it was decided that for 2006 the reminder intervals would be reduced in order to meet the tight time scale imposed on the survey. The pattern adopted was:

- Initial invite early on Tuesday morning
- First reminder two days later on Thursday morning
- Second and final reminder four days later on Monday Morning
- Close the survey 5 days later

The pattern of responses for the 2006 survey is shown in Fig. 3.

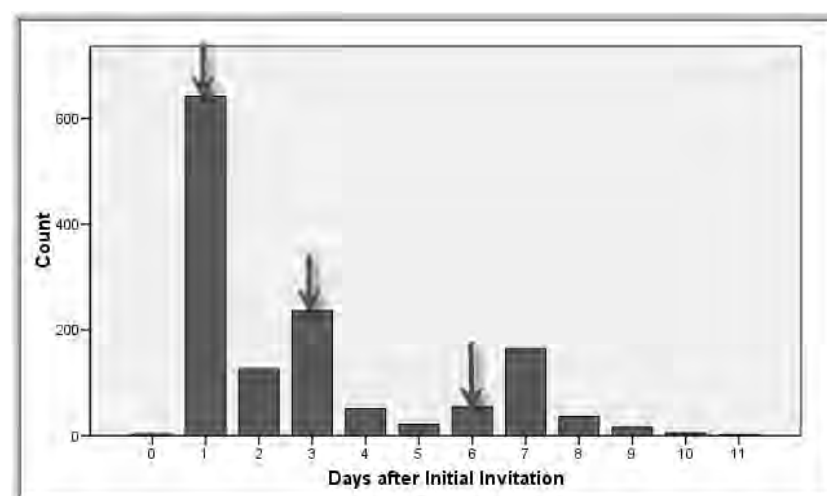


Fig. 3

Unfortunately when the number of responses from 2006 was compared with the two previous years (Table 1) there was a decline in the overall number of responses; the total number of

students surveyed (12,000) was almost the same for the three years so something had caused the decline.

Year of Survey	Number responding	Response Rate
2004	1472	12.27
2005	1944	16.20
2006	1358	11.32

Table 1.

The immediate reaction was that the cause of this decline was the speed of reminders, ignoring the facts that the survey timing, after the Easter vacation instead of before, was different and that prior to the Easter vacation there had been two long “all student” surveys with more substantial incentives to complete them.<sup>2</sup> It looked as though all future surveys were going to revert to the slow reminder intervals which would then exacerbate the problem of coordinating the increasing number of surveys to staff and students.

It was decided that it would be necessary to prove that the decline in response rates had been due to other factors than the shortened reminder periods.

## 2. Experimental process

As the frequent reminder rate had been held responsible for the decline in response rate any experiments which made changes to the established, slow, reminder pattern could only be done with surveys that were not deemed to be important to the auditing or administration process of the University. Fortunately a number of such surveys were found and a standard procedure for administering these surveys was established. The experimental process was helped by the fact that the University of Aberdeen was a beta test site for the new version of snap with the snaponline option which includes the facilities to automate the survey process. This meant that all the surveys used in this experiment to establish whether reminder rates affected response rate would be administered in a similar fashion.

All the surveys in the experiment used the University E-mail address of the student to issue the invites and reminders to ensure consistency. To ensure students were allocated at random to the two experimental populations the list of students was sorted in ascending order on the E-mail addresses and then the even numbered rows were allocated to the ‘fast’ reminder group and the odd numbered rows to the other, ‘slow’ group. If the data set which provided the E-mails also contained other information about the student then the data was pre-sorted on these criteria. One study had information on the year of study and the sex of the student so this dataset was sorted initially by year of study, then by gender within year of study and finally on E-mail address before allocating to the study group on the same “odd to slow, even to fast” basis.

The surveys were always set up so that the initial invitation for both groups was sent out early on Tuesday morning. The snaponline software can be set up with specified start, finish and reminder dates; all these actions happen as soon as possible after midnight. The patterns of reminders for the two groups are shown in table 2 below.

<sup>2</sup> Experience at the University of Aberdeen has shown that without an incentive, or prize, awarded randomly to one of the students who complete a survey, the response rates are extremely poor.

	Invite	Interval	Reminder 1	Interval	Reminder 2	Interval	Close	Total
<b>Fast</b>	Tues	+2	Thur	+4	Mon	+2	Thu (2359)	8 days
<b>Slow</b>	Tues	+8	Wed	+8	Thur	+8	Fri (2359)	24 days

Table 2.

Analysis of the response rates of the four surveys conducted in this way shows that there is no deterioration between the frequent and infrequent pattern of reminders and in fact in some surveys there is a slight increase in response rates from the frequent groups. The comparative numbers responding are shown in Table 3 and Fig. 4 below.

Survey	Number Surveyed	Speed	Response Rate	Number responding
<b>Computing Introduction</b>	818	Fast	31.54	258
		Slow	31.17	255
<b>Medical Students</b>	966	Fast	29.40	284
		Slow	27.02	261
<b>Freshers Week Experience</b>	2172	Fast	17.54	381
		Slow	18.00	391
<b>First semester Experience</b>	2202	Fast	18.85	415
		Slow	24.07	530

Table 3.

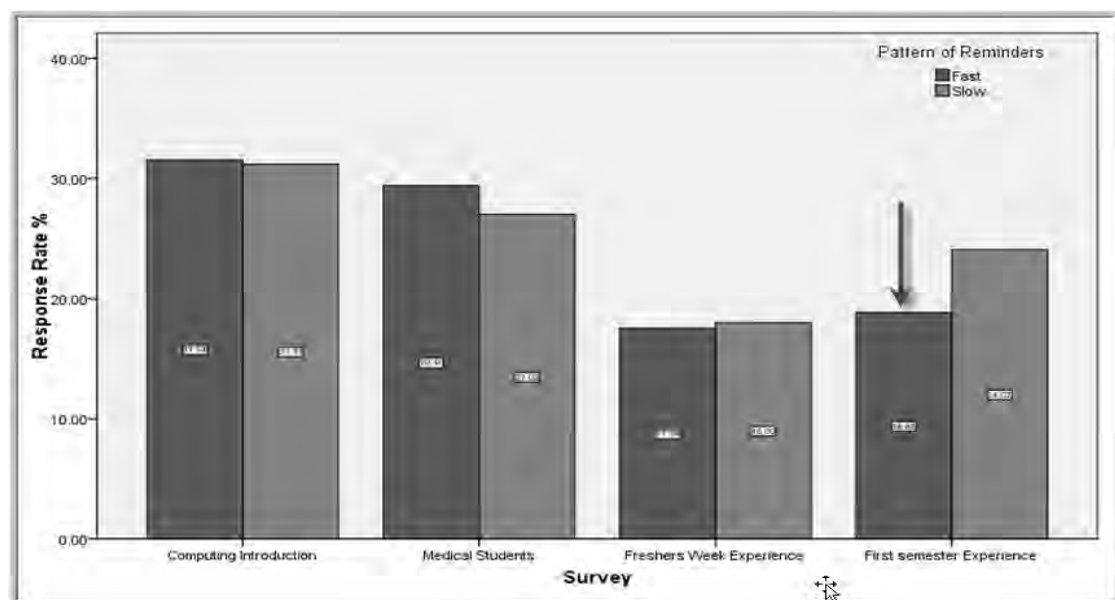


Fig. 4

Although the chart indicates that on the whole there is no difference between the patterns of responses between the two groups there is one survey group, “First Semester Experience”, which does not follow the trend as there is a marked difference between the fast and slow groups. The cause of this difference was most probably due to a reaction on the part of the respondents in the “fast” group to a technical problem with the server, which meant they were

unable to start the survey. Although both survey groups had been set up in a similar manner on the same server the fault only affected the survey sent to the “fast” group. Even when the fault was resolved after a short time it appears that resistance to completing the survey had already been established and couldn’t be restored. Modifying the reminder message to explain what had happened may have reduced the impact of the temporary software failure but it still resulted in a lower response rate for the group affected. This problem only serves to highlight how sensitive response rates are to a large number of interacting factors and that it is important to ensure that those that can be controlled are optimised. It is interesting to note that the decline in responses to the “fast” group of the “1<sup>st</sup> Semester experience” study does not appear to be reflected in the numbers responding to the reminders as shown in Fig. 5 and Table 4.

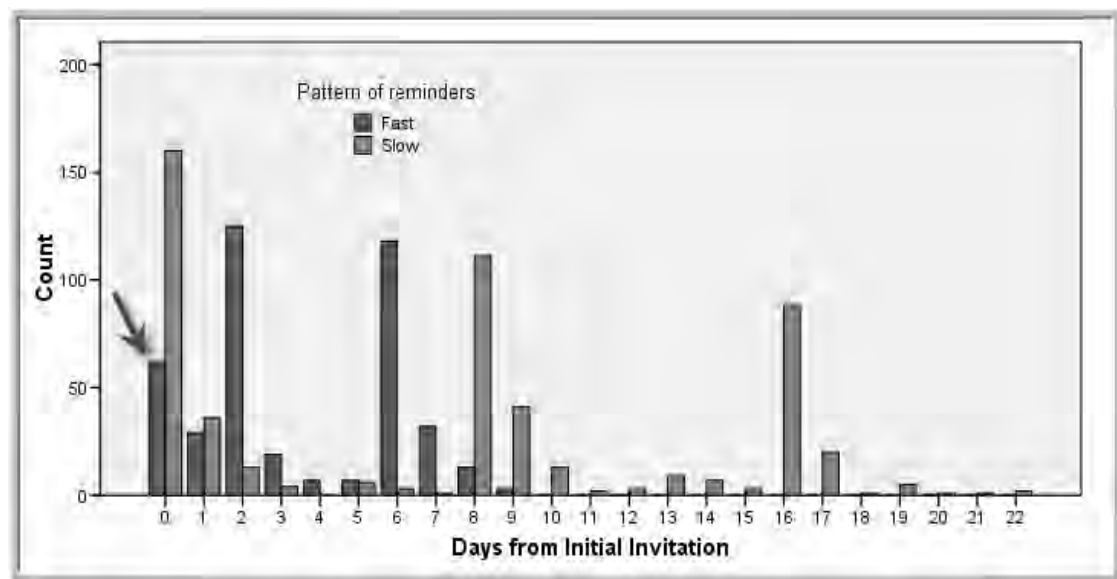


Fig. 5

	Fast		Slow	
<b>Before 1<sup>st</sup> Reminder</b>	91	4.13%	223	10.13%
<b>Between 1<sup>st</sup> and 2<sup>nd</sup> Reminder</b>	158	7.18%	189	8.58%
<b>Before 2<sup>nd</sup> Reminder ( Cumulative )</b>	249	11.30%	412	18.71%
<b>Between 2<sup>nd</sup> Reminder and Close</b>	166	7.54%	118	5.36%
<b>Final figures</b>	415	18.85%	530	24.07%

Table 4.

It may be that once respondents have been deterred from responding to a survey for any reason they are unlikely to respond at all, even though there have been reminders. This could be the basis for further experimental work on response rate but how this could be measured would be hard.

As mentioned before the only surveys that were available to be used in the experiment were non-critical ones to students. This meant there was a high probability that the results might not be reproducible to staff surveys or external research. Fortunately a survey to staff was offered as part of the study and the same criteria on allocation to the experimental groups was



applied, using the staff number instead of the student number and sorting into faculty / department groupings instead of year and course. This survey had the most marked difference between the ‘fast’ and ‘slow’ groups in terms of response rates (16.23% vs. 14.55%) – Fig. 6.

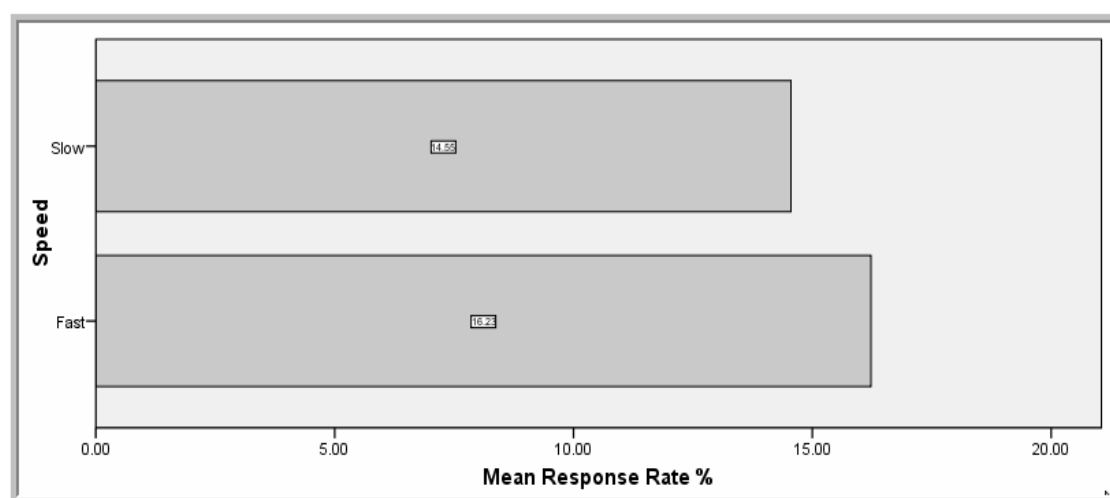


Fig. 6

### 3. Conclusions

From the limited information available it appears that the shorter time interval between E-mail reminders for web based surveys has little effect on the overall response rate, in fact it may show a slight improvement in some surveys. This means more surveys can be conducted within a given time period. However, it is unlikely that the experiment can be repeated at the University of Aberdeen as most people have now adopted the “fast” reminder pattern.

### References

- Crawford, S. C., Couper, M. P., Lamias, M. J. 2001. Web Surveys: perceptions of Burden. *Social Science Computer Review*, 19 146-162
- Cobanoglu, C. & Cobanoglu, N. 2003. The effect of incentives in web surveys: Application and ethical considerations. *International Journal of Market Research*, **45**, 475-488+520.
- Conrad, F., Couper, M. P., Tourangeau, R. & Peytchev, A. 2003. Effectiveness of Progress Indicators. It's what's up front that counts. *ASC2003: The Impact of Technology on the Survey Process*, 333-342
- Healey, B. 2007. Drop downs and scroll mice: The effect of response option format and input mechanism employed on data quality in Web surveys. *Social Science Computer Review*, **25**, 111.
- Heerwegh, D. 2005. Effects of personal salutations in e-mail invitations to participate in a web survey. *Public Opinion Quarterly*, **69**, 588-598.
- Pearson, J. & Levine, R. 2003. Salutations and Response Rates to Online Surveys. *ASC2003: The Impact of Technology on the Survey Process*, 351-362

## About the Author

I have worked as a Senior Computing Adviser at the University of Aberdeen for nearly 30 years supporting staff and students from all disciplines in the collection, storage and manipulation of their data for research. I was one of the people responsible for introducing snap to the University in 1999 and have acted as a beta tester for the software for a number of years. I try to maintain a research interest in trying to improve response rates to surveys but this has to be balanced with increasing demands for assistance on the design and administration of surveys. I would like to thank the brave people who allowed me to experiment with their surveys, the staff at SnapSurveys who helped resolve the problems I encountered, family and friends who gave advice on the paper and the Directorate of Information Technology for their support in conducting this research.

I can be contacted via E-mail at [j.s.lemon@abdn.ac.uk](mailto:j.s.lemon@abdn.ac.uk)



## Online Panels – A Paradigm Theft?

Jelke Bethlehem & Ineke Stoop

### Abstract

Decreasing survey participation leads to increasing survey costs and smaller precision of survey estimates. In the likely case that noncontacts and refusals differ from respondents, it may also increase nonresponse bias. Online panels are promoted as the solution to these problems. Online panels are relatively inexpensive as large samples can easily be drawn from panels containing hundreds of thousands of willing respondents. With an increasing internet penetration, even specific groups such as the elderly will have internet access and therefore they can be included in online panels.

Survey methodology has been developed over a period of more than 100 years. The paradigm of probability sampling has shown to work well in social research, official statistics and market research. It has allowed researchers to produce well-founded and reliable survey results. Often the impression is created that this paradigm also applies to online panel research. Sadly, general survey quality criteria such as sample size and response rate cannot be generalized to panel studies. Under-coverage and self-selection may seriously limit their value for scientific and policy making purposes.

Some survey researchers claim that the problems mentioned above can be reduced by applying some kind of weighting adjustment procedure, e.g. using weighting variables measured in a reference survey. We argue this is a too optimistic point of view. The presentation will outline why online panels do not solve the problems caused by decreased participation. Will discuss online panels within a quality framework and will present examples of possibly useful applications of online panels in academic and governmental research.

In market research online panels are considered to be the future. There are signs, however, that despite the rapidly increasing market share of web research, the rising star of online panels is already on the wane.

### Keywords

Access panels, online research, coverage, probability sampling, nonresponse.

## 1. The changing landscape of survey research

The survey research landscape has undergone radical changes over the last decades. First, there was the change from traditional paper and pencil interviewing to computer-assisted

interviewing. And now face-to-face, mail and telephone surveys are increasingly replaced by online surveys.

The popularity of online research is not surprising. An online survey is a simple means to get access to a large group of people. Questionnaires can be distributed at very low costs. No interviewers are needed, and there are no mailing and printing costs. Surveys can be launched very quickly. Little time is lost between the moment the questionnaire is ready and the start of the fieldwork. And online surveys offer new, attractive possibilities, such as the use of multimedia (sound, pictures, animation and movies).

However, there is another side to this coin. Online research is not without methodological problems. These problems have an impact on the quality of the survey results. The cause of these problems can partly be found in the incorrect application of the principles of sample survey theory that have been developed more than a century ago.

The theory of survey sampling is heavily based on the probability sampling paradigm. By selecting random samples probability theory can be applied, making it possible to quantify accuracy of estimates. This paradigm has been successfully applied in official and academic statistics in the 1940's, and to a much lesser extent also in more commercial market research.

At first sight, online surveys seem to have much in common with other types of surveys. It is just another mode of data collection. Questions are not asked face-to-face or by telephone, but over the Internet. What is different for many online surveys, however, is that the principles of probability sampling have not been applied. This can have a major impact on survey results.

It is not always clear what is meant by online research and how online surveys fit in the framework of survey sampling theory. There is also confusion about what it means. Online research is often used as a synonym for online panels, but this is a too limited view. There are also large online cross-sectional surveys. An example is the *21minuten.nl*, a survey supposed to supply answers to questions about important problems in Dutch society. Within a period of six weeks in 2006 about 170.000 people completed the online questionnaires. A similar survey was conducted in Germany (*Perspektive Deutschland*).

Panels need not necessarily be online panels. Panels existed already long before the Internet emerged. Panel members can very well complete questionnaires face-to-face, by telephone or by mail. An interesting example of an online panel 'avant la lettre', is the Telepanel, see Saris (1998). It started in 1986 and used home computers placed in the homes of panel members. Questionnaires were downloaded, and answers uploaded, by means of telephone and modem.

Online research is often claimed to be representative because of the high number of respondents or as a result of advanced adjustment weighting procedures. The term representative is rather confusing. It can have many meanings and is often used in a very loose sense to convey a vague idea of good quality. A high number of respondents are often considered to ensure validity and reliability. There are serious doubts, however, whether a large sample size as a result of self-selection of respondents has the same meaning as a large sample size in probability sampling. Similarly, a high response rate in a sample among cooperative panel members is unlikely to have the same impact on the quality of outcomes as a high response rate in a random sample of the population.

This paper discusses online research, web surveys and access panels from the perspective of classic survey quality criteria and compares the different paradigms behind probability sampling and other forms of sample selection. It ends with a plea for transparency.

## 2. So you have a representative sample?

In her overview of the history of survey research in the United States, Converse (1987) sketches the permanent controversy between market research organizations with their tradition of non-probability samples such as quota sampling and 'juries', and official statistics with probability sampling founded in inferential statistics and probability theory. Fowler (2002, p. 53) describes the gap between both traditions as follows: 'The federal government will not fund survey research efforts designed to make estimates of population characteristics that are not based on probability sampling techniques. Most academic survey organizations and many non-profit research organizations have a similar approach to sampling. At the same time, most of the major public opinion groups, political polling groups, and market research organizations rely heavily on non-probability sampling methods'.

The controversy goes back to the roots of survey sampling, see e.g. Kish (2003). Anders Kiaer, the director of the Norwegian Statistical Bureau, can be seen as the founder of the survey method that is now widely applied in official statistics and social research. In 1895 he published his *Representative Method*. It was a partial inquiry in which a large number of persons were questioned. This selection should form a 'miniature' of the population. Persons were selected arbitrary, but according to some rational scheme based on general results of previous investigations. Anders Kiaer stressed the importance of *representativeness*. His argument was that, if a sample was representative with respect to variables for which the population distribution was known, it would also be representative with respect to the other survey variables. A basic problem of the Representative Method was that there was no way of establishing the accuracy of estimates. The method lacked a formal theory of inference. It was Bowley (1906), who made the first steps in this direction. He showed that for large samples, selected at random from the population, the estimate had an approximately normal distribution.

From this moment on, there were two methods of sample selection. The first one was Kiaer's Representative Method, based on purposive selection, in which representativeness played a crucial role, and for which no measure of the accuracy of the estimates could be obtained. The second was Bowley's approach, based on simple random sampling, and for which an indication of the accuracy of estimates could be computed. Both methods existed side by side for a number of years. This situation lasted until 1934, in which year the Polish scientist Jerzy Neyman published his now famous paper, see Neyman (1934). Neyman developed a new theory based on the concept of the confidence interval. By using random selection instead of purposive selection, there was no need any more to make prior assumptions about the population.

The contribution of Neyman was not only that he invented the confidence interval. By making an empirical evaluation of Italian census data, he could prove that the Representative Method based on purposive sampling failed to provide satisfactory estimates of population characteristics. The result of Neyman's evaluation of purposive sampling was that the method fell into disrepute in official statistics.

The concept of 'representativity' plays a crucial role in the discussion about the foundations of survey sampling. Kruskal and Mosteller (1979a, 1979b and 1979c) present an extensive overview of what representative is supposed to mean in non-scientific literature, scientific literature excluding statistics and in the current statistical literature. They found the following meanings for 'representative sampling': (1) general acclaim for data, (2) absence of selective

forces, (3) miniature of the population, (4) typical or ideal case(s), (5) coverage of the population, (6) a vague term, to be made precise, (7) representative sampling as a specific sampling method, (8) as permitting good estimation, or (9) good enough for a particular purpose.

Kruskal and Mosteller (1979b, p. 125) recommended not using the word *representative*, but instead to specify what one means. This did not catch on. Kish (2003, p. 12) thought that *representative sampling* was a term that can be avoided and assumed in 1995 that it was disappearing from the technical vocabulary. Schnell (1997, p. 12) called it an immeasurable, unscientific concept, both with little success. ‘*Representative*’ in The Netherlands is still a household word in describing samples and sampling even when it is not clear whether this means a miniature of the population, or good coverage of the population, or a random sample or just “trust me!”.

The problem is, of course, that both in probability sampling and other forms of sampling claims are made samples are representative, often with quite different meanings and sometimes with no concrete meaning at all besides conveying a vague sense of good quality.

### 3. By a small sample we may judge the whole piece?

The basics of probability sampling as it is applied now in e.g. official statistics are laid down by Horvitz and Thompson (1952) in their seminal paper. They state that unbiased estimators of population characteristics can always be constructed, provided samples are selected by means of probability sampling and every element in the population has a known and strictly positive probability of being selected.

Moreover, under these conditions standard errors of estimates, and thus confidence intervals, can be computed. Therefore it is possible to establish the accuracy of estimates. The Horvitz-Thompson approach can be used also in surveys with complex sampling designs, like stratified random samples, cluster samples and two-stage samples.

Notwithstanding the advantages of probability sampling, there are many forms of non-probability sampling. Kalton (1983) distinguishes three types of non-probability samples:

- *Convenience samples*. Sample selection is mainly based on easy availability or accessibility of elements. An example is conducting survey interviews in a shopping mall on Saturday afternoon. Also samples composed of respondents volunteering to participate are convenience samples.
- *Purposive samples*. The sample is chosen by a subject matter expert in such a way that is ‘representative’ in his/her opinion. The expert will usually attempt to include elements that cover all various aspects in the population.
- *Quota sampling*. Interviewers are given quotas of different types of people with whom they have to conduct interviews. Quotas are often based on demographic characteristics like gender, age, marital status and neighbourhood. Quota sampling is similar to the original Representative Method developed by Anders Kiaer in 1895.

The weakness of all non-probability methods is that there is no theoretical framework although efforts have been made to develop this (Deville, 1991). Therefore it is not possible to establish accuracy of estimates other than by subjective assessment. Nevertheless, despite its theoretical weakness, non-probability sampling is widely used, particularly in market

research. Usually, the reasons to use it are convenience and relatively low costs. Increasingly, market researchers seem to feel that the problems facing regular surveys (declining participation, people being at home less often, legal restrictions, rapidly increasing costs) are such that non-probability sampling – especially in the form of online panels – is actually to be preferred above traditional surveys or at least considered as an equal and more manageable alternative (see for instance Thomas, 2006) in his editorial to the recent ESOMAR Panel Research Conference.

Morton-Williams (1993, pp.31-35) has shown that the claim that quota samples are representative is based on two assumptions, namely ‘... that the behaviour and attitudes to be measured are related primarily to the variables used as quota controls; secondly, that they are not associated independently of these controls with factors underlying nonresponse nor with the characteristics of those likely to require more than one call to obtain an interview’ (p.32). Or, as Sudman (1966) said a long time ago: ‘In probability sampling with quotas the basic assumption made is that it is possible to divide the respondents into strata in which the probability of being available for interviewing is known and is the same for all individuals within the stratum, although varying between strata.’ These assumptions cannot be put to the test. Similar claims, however, are made with respect to access panels.

*Online panels*, also called *access panels*, and for short just *panels*, are becoming increasingly popular in the western world. A panel consists of persons who have agreed to regularly participate in surveys run by a specific organization, generally a market research organization. There are various ways to set up an online panel. Both probability and non-probability sampling techniques can be used.

Ideally, a panel is constructed using a random sample from a population. One way to select such a sample is using Random Digit Dialling. Another is to invite respondents in a probability based survey to become a member of a follow-up panel. If the original selection probabilities and response rates are available (which is rarely the case), a survey among (a random sample from) an access panel can theoretically be considered as a case of probability sampling. In this case, an access panel is very similar to a host survey.

Unfortunately, many online panels are based on some form of non-probability sampling. Major opinion polls in The Netherlands rely on self-selection of respondents. The same is true for the large 21minutes.nl web survey. A study across 19 online panels of Dutch market research organisations shows that most of them use self-registration, links and banners on websites or snowballing, see Vonk et al. (2006). This all means that most online research has two fundamental methodological flaws:

**Keywords**      *Under-coverage.* People without Internet will never participate in online research. This means research results can only apply to the Internet population and not to the complete population.

**Keywords**      *Self-selection.* Researchers have no control over the selection mechanism. Selection probabilities are unknown. Therefore, no unbiased estimates can be computed, nor can the accuracy of estimates be established.

These two problems are analysed in more detail in the subsequent sections.

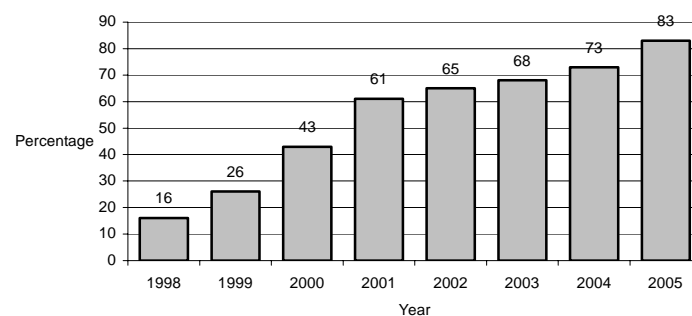


#### 4. Coverage problems

Online research suffers from under-coverage because the target population is usually much wider than the Internet population. Bethlehem (2007) has analysed the situation in the Netherlands.

The percentage of persons having an Internet connection at home increases from year to year, see figure 4.1. In seven years time, the percentage of Internet connections increased from 16% to 83%. Still, it is clear that not every household will have access to Internet in the near future.

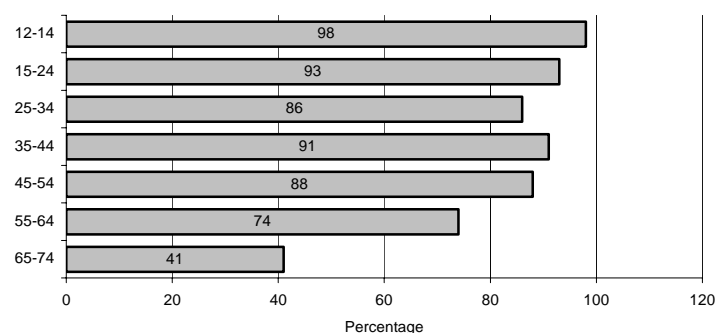
Figure 4.1. Percentage of persons having Internet



Internet access is unevenly distributed over the population. More males than females have access to the Internet. Figure 4.2 shows the percentage of people having Internet access at home by age group (in 2005). This percentage decreases with age. Particularly, the elderly are much under-represented when the Internet is used as a selection mechanism.

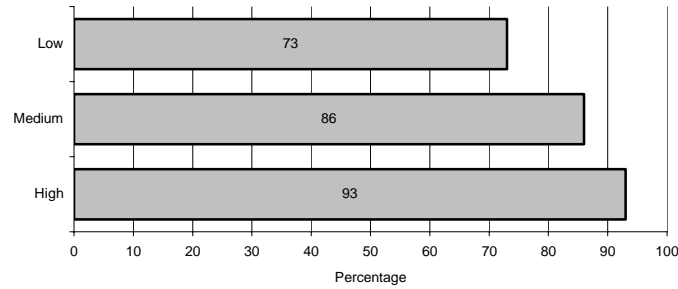
Figure 4.3 contains the percentage of people using the Internet by level of education (in 2005). It is clear that people with a higher level of education more frequently have Internet than people with a lower level of education.

Figure 4.2. Having Internet, by age.



Internet access among non-native young people is much lower than among native young people: 91% of the young natives have access to Internet. This is 80% for young people from Surinam and Antilles, 68% for young people from Turkey and only 64% for young people from Morocco. The results described above are in line with the findings of authors in other countries. See e.g. Couper (2000), and Dillman and Bowker (2001).

Figure 4.3. Having Internet by level of education.



To see what the impact of under-coverage on estimates can be, we analyse the situation in which a proper random sample is selected from the Internet population. Let the target population of the survey consist of  $N$  persons, which are labelled  $1, 2, \dots, N$ . Associated with each person  $k$  is a value  $Y_k$  of the target variable  $Y$ . Aim of the online survey is assumed to be estimation of the population mean

$$Y = \frac{1}{N} \sum_{k=1}^N Y_k \quad (4.1)$$

of the target variable  $Y$ .

The population  $U$  is divided into two sub-populations  $U_I$  of persons having access to Internet, and  $U_{NI}$  of persons not having access to the Internet. Associated with each person  $k$  is an indicator  $I_k$ , where  $I_k = 1$  if person  $k$  has access to the Internet (and thus is a member of sub-population  $U_I$ ), and  $I_k = 0$  otherwise. The sub-population  $U_I$  will be called the *Internet population*. Let

$$N_I = \sum_{k=1}^N I_k \quad (4.2)$$

denote the size of sub-population  $U_I$ . Likewise,  $N_{NI}$  denotes the size of the sub-population  $U_{NI}$ , where  $N_I + N_{NI} = N$ . The mean of the target variable for the elements in the Internet-population is equal to

$$\bar{Y}_I = \frac{1}{N_I} \sum_{k=1}^N I_k Y_k. \quad (4.3)$$

A random sample selected without replacement from the Internet-population can be represented by a series  $a_1, a_2, \dots, a_N$  of  $N$  indicators, where the  $k$ -th indicator  $a_k$  assumes the value 1 if person  $k$  is selected, and otherwise it assumes the value 0, for  $k = 1, 2, \dots, N$ . Note that always  $a_k = 0$  for elements  $k$  outside the Internet-population. The sample size is denoted by  $n_I$ . The sample mean

$$\bar{y}_I = \frac{1}{n_I} \sum_{k=1}^N a_k I_k Y_k \quad (4.4)$$

is an unbiased estimator of the mean  $\bar{Y}_I$  of the Internet population, but not necessarily of the mean  $\bar{Y}$  of the target population. The bias is equal to

$$B(\bar{y}_{HT}) = E(\bar{y}_{HT}) - \bar{Y} = \bar{Y}_I - \bar{Y} = \frac{N_{NI}}{N} (\bar{Y}_I - \bar{Y}_{NI})$$

(4.5)

The magnitude of this bias is determined by two factors. The first factor is the relative size  $N_{NI} / N$  of the sub-population without Internet. The bias will increase as a larger proportion of the population does not have access to Internet. The second factor is the *contrast*  $\bar{Y}_I - \bar{Y}_{NI}$  between the Internet-population and the non-Internet-population. The more the mean of the target variable differs for these two sub-populations, the larger the bias will be.

Since Internet coverage is steadily increasing, the factor  $N_{NI} / N$  is decreasing. This has a bias reducing effect. However, it is not clear whether the contrast also decreases. To the contrary, it is not unlikely that the (small) group of people without Internet will be more and more different from the rest of the population. As a result, substantial bias may still remain.

## 5. Self-selection problems

The problem of self-selection is that the researcher has no control over the selection mechanism of the survey. Respondents are those people who happen to have Internet, visit the website and decide to participate in the survey. Therefore, no unbiased estimates can be computed nor can the accuracy of estimates be determined.

Table 5.1. Dutch Parliamentary elections 2003  
Outcomes and the results of various opinion surveys

	Election	Kennisnet	RTL4	SBS6	Nederland 1
Sample size		17,000	10,000	3,000	1,200
Seats in parliament:					
CDA (christian democrats)	44	29	24	42	42
LPF (populist party)	8	18	12	6	7
VVD (liberals)	28	24	38	28	28
PvdA (social democrats)	42	13	41	45	43
SP (socialists)	9	22	10	11	9
GL (green party)	8	26	9	6	8
D66 (liberal democrats)	6	4	7	5	6
Other parties	5	14	9	7	7
Mean Absolute Difference		12.5	5.3	1.8	0.8

The effects of self-selection can be illustrated using an example related to the general elections in The Netherlands in 2003. Various organisations made attempts to use opinion polls to predict the outcome of these elections. The results of these polls are summarised in table 5.1.

A typical example of a self-selecting survey was the survey on the Dutch website *Kennisnet* (Knowledge net). This is a website for all those involved in education. More than 11,000 schools and other educational institutes use this website. The survey was an opinion poll for the general elections of 22 January 2003. Everybody, also those not involved in education, could participate in the poll. Table 4.1 contains both the official results (seats in parliament) of the election (column *Election*) and the results of this poll on the morning of the Election Day (column *Kennisnet*). The survey estimates were based on votes of approximately 17,000

people. No adjustment weighting was carried out. Although this is a large sample, it is clear that the survey results were no way near the true election results. The Mean Absolute Difference (MAD) is equal to 12.5, which means that the estimated number of seats and the true number of seats differ on average by an amount of 12.5. This survey could certainly not be used for predicting election results.

Another example of a self-selection web survey was the election site of the Dutch Television channel RTL 4. It resembled to some extent the Kennisnet survey, but was targeted at a much wider audience. Again, the survey researcher had no control at all over who was voting. There was some protection, by means of cookies, against voting more than once. However, this also had the draw-back, that only one member of the family could participate. Table 5.1 shows the survey results at noon on the day of the general elections (column *RTL4*). Figures were based on slightly over 10,000 votes. No weighting adjustment procedure was carried out. The results are better than that of the Kennisnet survey (the MAD decreased from 12.5 to 5.3). However, deviations between estimates and true figures are still substantial, particularly for the large parties. Note that even a large sample size of over 10,000 people does not help to get accurate estimates.

The Dutch commercial television channel *SBS6* used an access panel. Values of basic demographic variables were available for all panel members. A sample of size 3,000 was selected. Selection was carried out such that the sample was representative with respect to the social-demographic and voting characteristics. Table 5.1 shows the results (column *SBS6*). The survey took place on the day before the general elections. Although attempts have been made to create a 'representative' sample, the results differ still from the final result. The MAD has decreased to 1.8, but is still substantial.

A better prediction was obtained with a true probability sample. The table shows the results of a survey based on such a probability sample. It was carried out by the television channel *Nederland 1* in co-operation with the marketing agency *Interview-NSS*. A sample of size 1,200 was selected by means of random digit dialling. The MAD was reduced to 0.8.

Table 5.2. Dutch Parliamentary elections 2006  
Outcomes and the results of various opinion surveys

	<b>Election result</b>	<b>Politieke<sup>1)</sup> Barometer</b>	<b>Peil.nl<sup>2)</sup></b>	<b>De Stemming<sup>3)</sup></b>	<b>DPES 2006<sup>4)</sup></b>
Sample size		1,000	2,500	2,000	2,600
Seats in parliament:					
CDA (christian democrats)	41	41	42	41	41
PvdA (social democrats)	33	37	38	31	32
VVD (liberals)	22	23	22	21	22
SP (socialists)	25	23	23	32	26
GL (green party)	7	7	8	5	7
D66 (liberal democrats)	3	3	2	1	3
ChristenUnie (christan)	6	6	6	8	6
SGP (christian)	2	2	2	1	2
PvdD (Animal party)	2	2	1	2	2
PvdV (Conservative)	9	4	5	6	8
Other parties	0	2	1	2	1
Mean Absolute Difference		1.27	1.45	2.00	0.36

1) Politieke Barometer, Interview-NSS: sample from online panel (N=1000)

- 2) Peil.nl, Maurice de Hond: representative sample from online panel (weighted) (N=2500)
- 3) De Stemming, TNS-NIPO: sample from online panel (N=2000)
- 4) Dutch Parliamentary Election Study, fieldwork 10 October-22 November 2006, two-stage random sample from population register, N=4000, response rate=65%.

A more recent comparison is presented in table 5.2. Sample sizes are similar in this case and the differences between MAD based on three samples from online panels and one random sample are much smaller. As in 2003 the random sample, in this case the Dutch Parliamentary Election Study conducted by Statistics Netherlands, clearly outperformed the other surveys.

The conclusion from the analysis above is that a probability sample is a vital prerequisite for making proper inference about the target population of a survey. Even with a probability sample of only size 1,200 better results can be obtained than with a non-probability sample of size 10,000 or more. To explore the effect of self-selection on estimates, we assume that each person  $k$  in the Internet-population has unknown probability  $\rho_k$  of participating in the survey, for  $k = 1, 2, \dots, N_I$ . The responding persons can be denoted by a series  $r_1, r_2, \dots, r_N$  of  $N$  indicators, where the  $k$ -th indicator  $r_k$  assumes the value 1 if person  $k$  participates, and otherwise it assumes the value 0. Note that sampling without replacement is assumed. The expected value  $\rho_k = E(r_k)$  is called the *response propensity* of person  $k$ . For sake of convenience we have also introduced response propensities for non-Internet-population elements. By definition the values of all these probabilities are 0. The realised sample size is denoted by  $n_S = r_1 + r_2 + \dots + r_N$ .

A naive researcher assuming that every person in the Internet-population has the same probability of being selected in the sample, will use the sample mean

$$\bar{y}_S = \frac{1}{n_S} \sum_{k=1}^N r_k Y_k \quad (5.1)$$

as an estimator for the population mean. The expected value of this estimator is approximately equal to

$$E(\bar{y}_S) \approx \bar{Y}_I^* = \frac{1}{N_I \bar{\rho}} \sum_{k=1}^N \rho_k I_k Y_k \quad (5.2)$$

where  $\bar{\rho}$  is the mean of all response propensities in the Internet-population, see Bethlehem (2007).

Generally, the expected value of the sample mean is not equal to the population mean of the Internet-population. The only situation in which the bias vanishes is that in which all response propensities in the Internet-population are equal. Indeed, in this case, self-selection does not lead to an unrepresentative sample because all elements have the same selection probability.

Bethlehem (1988) shows that the bias of the sample mean (5.1) can be written as

$$B(\bar{y}_S) = E(\bar{y}_S) - \bar{Y}_I \approx \bar{Y}_I^* - \bar{Y}_I = \frac{C(\rho, Y)}{\bar{\rho}}, \quad (5.3)$$

in which

$$C(\rho, Y) = \frac{I}{N_I} \sum_{k=1}^N I_k (\rho_k - \bar{\rho})(Y_k - \bar{Y})$$

(5.4)

is the covariance between the values of target variable and the response propensities in the Internet-population. The bias of the sample mean (as an estimator of the mean of the Internet population) is determined by two factors:

- The average response propensity. The more likely people are to participate in the survey, the higher the average response propensity will be, and thus the smaller the bias will be.
- The relationship between the target variable and response behaviour. The stronger the relationship, the higher the bias will be.

Three situations can be distinguished in which this bias vanishes:

- 1) All response probabilities are equal. Again, this is the case in the which the self-selection process can be compared with a simple random sample;
- 2) All values of the target variable are equal. This situation is very unlikely to occur. If this were the case, no survey would be necessary. One observation would be sufficient.
- 3) There is no relationship between target variable and response behaviour. It means participation does not depend on the value of the target variable.

If it is the objective of the survey to estimate the mean of the total population (and not just the mean of the Internet population), two factors contribute to the bias: under-coverage and self-selection. Although it is theoretically possible that these two effects compensate one another, it is more likely in many practical situations that they enforce each other.

## 6. Does weighting adjustment help?

Weighting adjustment is a family of techniques that attempt to improve the quality of survey estimates by making use of auxiliary information. *Auxiliary information* is defined as a set of variables that have been measured in the survey, and for which information on their population distribution is available. By comparing the population distribution of an auxiliary variable with its sample distribution, it can be assessed whether or not the sample is representative for the population (with respect to this variable). If these distributions differ considerably, one must conclude that the sample is selective. To correct this, adjustment weights are computed. Weights are assigned to all records of observed elements. Estimates of population characteristics can now be obtained by using the weighted values instead of the unweighted values. Weighting adjustment is used to correct surveys that are affected by nonresponse, see e.g. Bethlehem (2002).

*Post-stratification* is a well-known and often used weighting method. To carry out post-stratification, one or more qualitative auxiliary variables are needed. Together they divide the target population into a number of strata (i.e. sub-populations). Identical adjustment weights are assigned to all elements in the same stratum. The bias of the estimate based on weighted data will be small if there is (on average) no difference between elements with and without Internet within the strata. This is the case if there is a strong relationship between the target variable and the stratification variables. Then the variation in the values of the target variable manifests itself between strata but not within strata. In other words, the strata are

homogeneous with respect to the target variable. Unfortunately, such auxiliary variables are not very often available, or there is only a weak correlation.

If proper auxiliary variables are not available, it might be considered to conduct a *reference survey*. Such a reference survey is based on a small probability sample, where data collection takes place with a mode different from the web, e.g. CAPI (Computer Assisted Personal Interviewing, with laptops) or CATI (Computer Assisted Telephone Interviewing). Under the assumption of no nonresponse, or ignorable nonresponse, this reference survey will produce unbiased estimates of the population distribution of auxiliary variables.

Using an estimated population distribution in post-stratification results in the same expected value for the estimator. So, the conditions under which the bias is reduced are the same as those for the normal post-stratification estimator.

An interesting aspect of the reference survey approach is that any variable can be used for adjustment weighting as long as it is measured in both surveys. For example, some market research organisations use 'webographics' or 'psychographic' variables to divide the population in 'mentality groups'. People in the same groups are assumed to have more or less the same level of motivation and interest to participate in such surveys. If this were the case, such variables can be effectively used in weighting adjustment. This requires of course, that adequate information on psychographics is available for the population, based on high response rate random samples.

The reference survey approach also has a disadvantage. Bethlehem (2007) shows that if a reference survey is used, the variance of the post-stratification estimator is for a large part determined by the size of the small reference survey. So, the large number of observations in the online survey does not help to produce accurate estimates. One could say that the reference survey approach reduces the bias of estimates at the cost of a higher variance.

## 7. Sample size and response rate: the wrong paradigm?

The theoretical framework of probability sampling show that large sample sizes and high response rates will have an impact on the quality of estimates. The former will reduce sampling error which implies increased precision and smaller confidence intervals. The latter are expected to reduce nonresponse error which means an increased accuracy and thus smaller bias.

Because of the large numbers of persons in a panel or web survey it will be possible to select (in online panels) or identify (in online surveys) a sufficient number of members of a specific group (single mothers, Cape Verdeans, persons with physical handicaps, non-voters, dog-owners). Because of the large sample sizes, high response rates and presence of small subgroups, combined with shorter turnaround times and lower costs, many people argue that online self-selection surveys are better than, or at least an affordable alternative for, probability surveys that are usually smaller and that also suffer from nonresponse.

However, as mentioned above, selection probabilities in online research are generally unknown. The sampling paradigm in which sample size and response rates affect precision and accuracy of estimates is rooted in probability sampling and cannot be transferred to non-probability sampling, if only because there are many different types and models for non-probability sampling (Groves, 1989, p. 249). One could argue that the importance of sample size and response rates do not belong to the paradigm of online surveys.

## Sample size

A frequent misunderstanding about online research is that large numbers make a sample better. Couper (2000) comments on the claims of a self-selected online survey: ‘“We received more than 50,000 responses – twice the minimum required for scientific validity –”..... while the survey did not yield a random sample and the selection probabilities are unknown, “this does not mean that the survey cannot yield representative social science data” (Emphasis in the original). They claim that the selection probabilities can be ‘estimated’ by comparing the distributions on standard demographic variables to official government statistics and applying weighting. This assertion is based on the assumption that matching two ‘samples’ on a variety of demographic characteristics will ensure that they also match on the survey variables of interest.’ (pp. 480-481). Not surprisingly, despite the large number of respondents, they did not resemble the U.S. population on a number of key indicators.

Dillman and Bowker (2001) express a similar opinion about online surveys: ‘Conductors of such surveys have in effect been seduced by the hope that large numbers, a traditional indicator of a high quality survey (because of low sampling error), will compensate in some undefined way for whatever coverage and nonresponse problems that might exist. Large numbers of volunteer respondents, by themselves, have no meaning. Ignoring the need to define survey populations, select probability samples, and obtain high response rates, together provide a major threat to the validity of web surveys.’

Couper (2001, pp. 173, 184) also pointed to the misguided assumption that large samples necessarily mean more valid responses. Only in the case of probability samples does an increase of sample size to an increase of precision. In non-probability samples, no inference to the underlying population is possible, and larger samples do not necessarily give better estimates than smaller samples.

Large online surveys have the advantage that specific subgroups can be identified. Information about such groups may be difficult to obtain in traditional surveys, because few people belong to these groups, they are hard to identify, or unlikely to participate in surveys. Again, the underlying assumption is that the elderly single women, low educated, ethnic minorities or other usually underrepresented groups who participate in an online survey are similar to people with the same characteristics but who do not participate. In some cases this might be a likely assumption, in others definitely not, and in most cases it will be difficult to test.

An additional caveat is that self-selection in panels and online surveys may require heavy weighting because of vastly varying participation propensities. Because of large weights the effective sample size is likely to be much smaller than the number of participants in a survey (Duffy et al., 2005). The effectiveness of a sample should be corrected by the average of the squared weights would result in a substantial reduction when even some weights are very large.

## Response rate

One of the main selling points of online panels used to be high response rates. Even though response rates in online panels are rapidly decreasing (see below) this claim ignores why high response rates are important, namely because high response rates ensure that everybody who was randomly selected in the sample actually also participated. When initial recruitment is based on self-selection or when initial nonresponse is high, high response rates on a survey



may hide a wide range of other survey errors (see also Bethlehem, 2007). Sudman and Kalton (1986) discussed the use of mail panels as a means of sampling rare populations twenty years ago. Their conclusions are still valid: 'Although 80% to 90% of panel households cooperate on a study, the major problem with mail panels is that the initial cooperation rate of households invited to participate in a panel is often 10% or less. Mail panels are usually balanced by major demographic variables to remove the most obvious selection biases, but other biases still remain. These unknown selection biases may distort the survey results, and the researcher will not be able to assess the possible distortion unless some independent check can be made'.

It is possible in a panel to generate a very high response rate by approaching only those who always participate when they are invited. This is unlikely to improve survey estimates, as this over-eager group might provide highly biased results. This shows again that high response rates in a panel will generally not be able to compensate for low initial cooperation, or an unknown selection bias. Defining response rate as the response propensity of willing respondents or boosting response rates by pre-selecting the most cooperative panel members makes response rates difficult to compare with those of probability samples. For this reason, response rates in non-probability samples do not have the same meaning as those in probability samples.

Nowadays, the response rates in panels are rapidly declining. This was one of the major concerns at the 2006 ESOMAR conference on panel research in Barcelona. A comparative Belgian study, for instance, reported the following response rates: 54% (random walk, face-to-face), 21% (online panel), 15% (random sample, mail) and 12% (random sample, telephone), see Schillewaert et al. (2006). Response rates of 20% do not seem to be unusual at all now. Because of this, two additional advantages of online panels may be losing their value. Firstly, it was assumed that weighting for under-represented groups (due to sampling errors, under-coverage and non-response) would not be necessary in a sample from a panel, because the structure and composition of the panel could be determined in advance (as in quota samples). Doing this is now much more difficult because of decreasing response rates. Control over the final composition can now only be achieved by taking into account unequal response propensities from many different groups, based on information about earlier participation in similar surveys. This again assumes that non-responding panel members in a specific group are similar to respondents.

Another purported advantage of online panels was that the topic of the survey had no impact on participation, because the decision to participate in a panel is general and the response rates on individual surveys was high. This would be an advantage because one important cause of nonresponse bias is the relationship between the topic of a survey and the decision to participate. With declining response rates in online panels, interest in the topic may have become an important determinant of survey participation again. Of course, in online surveys based on self-selection, the topic of the survey is likely to be the most important determinant of participation, possibly resulting in highly biased results.

## 8. Time and money

The turnaround time of online research is generally much shorter than in research based on other modes of data collection. In face-to-face surveys and telephone surveys several attempts have to be made to get in touch with a sample person, and to convert initial refusers. In mail

surveys, questionnaires have to be printed and sent to target respondents, followed by reminders. Fieldwork may take weeks or even months. In online surveys – where the emphasis is more on mass than on response rates – the preparation takes less time, and fieldwork takes far less time. According to Day et al. (2006, p. 268) “... the first 12-24 hours are the most important in any online project with approximately two-thirds of panellists responding in this period”.

This short turnaround time is a very great advantage of online research, contrasting sharply with long periods and many efforts spent on contacting hard-to-contacts sample persons in face-to-face surveys (see for instance Lynn et al., 2002; Stoop, 2005) in order to enhance response rates and reduce nonresponse bias. The difference in turnaround times reflects the different aims of traditional surveys and online surveys. In the first case, ideally, a lot of efforts are done to obtain the participation of every sample member – at the cost of a long survey period and lots of money – in order to come up with correct key estimates. In the second case it is possible to collect information on a current issue within a very short time, where the focus is less on minimizing survey errors and more on the number of respondents, speed and costs. These are two different survey paradigms that are hard to compare because the aims and strategies differ. And finally, one can of course wonder who these people are who complete questionnaires within the first hours after they survey has been launched.

Online surveys and surveys among a sample from an online panel are usually much less expensive than surveys based on other modes of data collection. In the former case, most of the costs are fixed costs, supplemented with some additional costs for remuneration of respondents. There are no costs for interviewers, no costs for printing and mailing, and no costs for data processing. Indeed, the increasing costs of traditional surveys were one of the main reasons for the spectacular growth of online research. As has been shown above, however, the theoretical underpinning and the aims of traditional survey research and online research do seem to differ substantially. Costs are very important, but costs should always be seen against the background of the purpose of a survey. Deming said about this almost half a century ago (1960, p. 31): ‘cost has no measure without a measure of quality, and there is no way to appraise objectively the quality of a (quota) sample as there is with a probability sample’.

## 9. So when to use online panels and online surveys

We have tried to show that traditional surveys research and online survey research came from different survey traditions and have different aims. What is clear is that in official statistics, where key estimates have to be provided that are not likely to be disputed and that are not based on refutable modelling, probability sampling is to be preferred. Again going back in the history of survey research, take account of how Sudman (1966) felt about this ‘To be more explicit, where survey results will receive very sophisticated analysis or when critical decisions will be based on them, it will be worthwhile to pay a substantial cost to achieve high standards of sampling, processing, and control. Thus, the Census Bureau rightly has very high standards on the Current Population Surveys. On the other hand, many exploratory studies do not require such high standards since the analysis may be more limited and the questionnaire may itself be a major source of error. Here quota sampling would be justified.’ Deville (1991) seems to agree ‘Official statisticians, on the other hand, are responsible for data that can be used by the entire society; and that can be used, in particular, in the

arbitration of disputes between various groups, parties, and social classes. .... Official statistics should not tolerate any uncontrollable bias in its products. It should carry out sample surveys using probabilistic methods.'

Probability samples may have a long turnaround time, be very expensive and require substantial funding and great efforts to achieve high response rates. On the other hand, they have a secure foundation in statistical theory which allows inferences to the target population. Acceptable response rates (Stoop, 2005) are still possible. So what to do?

Recently, a number of studies have been published comparing online panels and traditional surveys (Duffy et al., 2005; Schillewaert et al., 2006). One problem is that sometimes there are differences and sometimes there are not, and generalizing across online panels or across topics is very difficult. Web surveys have attracted the interest of a wide range of academic researchers and statisticians (see *Journal of Official Statistics*, Vol.22, No.2, 2006, special issue on web surveys). Although they are attractive, they are generally very reticent when it comes to presenting online panels as a serious alternative for random samples. One promising strain of research is the recent Dutch panel comparison study mentioned above (Vonk, Van Ossenbruggen and Willems, 2006). As usual it is clear that more research, and an open mind, is needed.

When representativeness is not an issue, online panels might be a useful tool for exploratory studies, experiments, tests, and other purposes as long as undisputed results – both point estimates and the size of relationships – are not the aim of the study. When representativeness is an issue, one sensible thing to do when preparing a survey is to take account of what Groves has to say about this at the end of a study on the relationship between nonresponse rates and nonresponse bias: '*Despite low response rates, probability sampling retains the value of unbiased sampling procedures from well-defined sampling frames. Coverage error of well-defined sampling frames can be evaluated relative to a desired target population, prior to the survey being launched. Probability sampling of the frame permits use of auxiliary variables on the frame to improve the estimation from the respondent-based data. Volunteer panels lose these advantages. Low response rate probability sample surveys need to marshal the power of auxiliary variables for post-survey adjustment*' (Groves, 2006, p. 669).

A second sensible thing to do would be to consider building a panel based on random sampling and including the Internet and the non-internet population, see Scherpenzeel (2006). An access panel conforming to strict methodological specifications could function as a host survey for a wide range of cross-sectional and longitudinal studies.

And finally, the final sensible thing to do when using non-probability online panels is to follow Fowler's admonition (2002, p. 56) and be transparent: 'If a researcher decides to use a non-probability sample, however, readers should be told how the sample was drawn, the fact that it likely is biased in the direction of availability and willingness to be interviewed, and that the normal assumptions for calculating sampling errors do not apply. Such warnings to readers are not common. In many cases, non-probability samples are misrepresented seriously, and that constitutes a serious problem for the credibility of social science research.'

## References

- Bethlehem, J.G. (1988), Reduction of the nonresponse bias through regression estimation. *Journal of Official Statistics* 4, pp. 251-260.

- Bethlehem, J.G. (2002), Weighting Nonresponse Adjustments Based on Auxiliary Information. In: Groves, R.M., Dillman, D.A., Eltinge, J.L. and Little, R.J.A. (eds): *Survey Nonresponse*. Wiley, New York.
- Bethlehem, J.G. (2007) *Reducing the bias of web survey based estimates*. Discussion Paper 07001, Statistics Netherlands, Voorburg, The Netherlands
- Bowley A.L. (1906): Address to the Economic Science and Statistics Section of the British Association for the Advancement of Science. *Journal of the Royal Statistical Society* 69, pp. 548-557.
- Converse, J.M. (1987) *Survey Research in the United States. Roots & Emergence 1890 – 1960*. Berkeley, University of California Press.
- Couper, M.P. (2000) Web Surveys. A Review of Issues and Approaches. *Public Opinion Quarterly*, Vol. 64, pp. 464-494.
- Couper, M.P. (2001) The Promises and Perils of Web Surveys. In: Andrew Westlake, Wendy Sykes, Tony Manners and Malcom Riggs (eds.) *The Challenge of the Internet*. Proceedings of the ASC International Conference on Survey Research Methods. Chesham, UK, May 2001.
- Day, D., R. Risk, J. Koo and B. Martin (2006) *Ensuring Data Integrity for Business Decisions. An In-depth Analysis of the Components that Affect Data Quality*. Panel Research 2006. ESOMAR Publication Series, Volume 317, pp. 253-269.
- Deming, W. (1960) *Sample design in business research*. New York: Wiley.
- Deville, J.C. (1991) A Theory of Quota Surveys. *Survey Methodology*. Vol. 17, N° 2. pp. 163-181.
- Dillman, D A. Bowker, D. (2001), The web questionnaire challenge to survey methodologists. In: Reips, U.D. and Bosnjak, M. (eds.), *Dimensions of Internet Science*, Pabst Science Publishers, Lengerich, Germany.
- Duffy, B., Terhanian, G., Bremer, J. and Smith, K. (2005) Comparing data from online and face-to-face surveys. *International Journal of Market Research*, Vol. 47, No. 6, pp.615-639
- Fowler, Floyd J. Jr. (2002) *Survey Research Methods*, 3<sup>rd</sup> Edition. Thousand Oaks, California: Sage Publications.
- Groves, R.M. (1989) *Survey Errors and Survey Costs*. New York: John Wiley & Sons, Inc.
- Groves, R.M. (2006) Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*, Vol. 70, pp. 646-675.
- Kalton, Graham (1987) *Introduction to Survey Sampling*. California: Sage Publications.
- Horvitz, D.G. en D.J. Thompson (1952), A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* 47, pp. 663-685.
- Kalton, G. (1983), *Introduction to Survey sampling*. Sage Publications, Inc, Beverly Hills, California.
- Kish, L. (2003) The Hundred Years' War of Survey Sampling (Reprinted from *Statistics in Transition*, 1995). In: Graham Kalton and Steven Heeringa (eds.) *Leslie Kish, Selected Papers*. New York: Wiley, pp. 5-19.

- Kruskal, William and Frederick Mosteller (1979a) Representative Sampling, I: Non-scientific Literature. *International Statistical Review* 47, pp. 13-24.
- Kruskal, William and Frederick Mosteller (1979b) Representative Sampling, II: Scientific Literature, Excluding Statistics. *International Statistical Review* 47, pp. 111-127.
- Kruskal, William and Frederick Mosteller (1979b) Representative Sampling, III: the Current Statistical Literature. *International Statistical Review* 47, blz. 245-265.
- Lynn, Peter, Paul Clarke, Jean Martin and Patrick Sturgis (2002) The Effects of Extended Interviewer Efforts on Nonresponse Bias. In: Robert M. Groves, Don A. Dillman, John L. Eltinge and Roderick J.A. Little (eds.) *Survey Nonresponse*. New York: Wiley, pp. 135-148.
- Morton-Williams, J. (1993) *Interviewer Approaches*. Aldershot: Dartmouth Publishing.
- Neyman, J. (1934): On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society* 97, pp. 558-606.
- Saris, W.E. (1998), Ten Years of Interviewing Without Interviewers: The Telepanel. In: Couper, M.P., Baker, R.P., Bethlehem, J.G., Clark, C.Z.F., Martin, J., Nicholls II, W.L. and O'Reilly, J.M. (eds.), *Computer Assisted Survey Information Collection*. Wiley, New York.
- Scherpenzeel, Annette (2006) *Een online panel als platform voor multidisciplinair onderzoek*. Paper presented at the DANS-symposium Access panels en online onderzoek, panacee of slangenkuil.
- Schillewaert, Niels, Annelies Verhaeghe, Bert Weijters and Kristof de Wolf (2006) *Social class and life style differences between modes of data collection*. Panel Research 2006. ESOMAR Publication Series, Volume 317, pp. 174-193.
- Schnell, Rainer (1997) *Nonresponse in Bevölkerungsumfragen, Ausmaß, Entwicklung und Ursachen*. Opladen: Leske und Budrich.
- Stoop, I.A.L. (2005) *The Hunt for the Last Respondent*. The Hague: Social and Cultural Planning Office.
- Sudman, Seymour (1966) Probability Sampling with Quotas. *Journal of the American Statistical Association*, Vol. 61, pp. 749-771.
- Sudman, Seymour, and Graham Kalton (1986) New Developments in the Sampling of Special Populations. *Annual Review of Sociology*, Vol. 12, pp. 401-429.
- Thomas, Randall K. (2006), Online Panels: the Research Revolution. *Panel Research 2006, ESOMAR World Research*, ESOMAR Publication Services, Vol. 317, pp. 6-7.
- Vonk, T., Van Ossenbruggen, R. and Willems, P. (2006), The effects of panel recruitment and management on research results, a study among 19 online panels. *Panel Research 2006, ESOMAR World Research*, ESOMAR Publication Services, Vol. 317, pp. 79-99.

## About the Authors

**Jelke Bethlehem** is Senior Methodologist at the Division of Methodology and Quality of Statistics Netherlands. He carries out research in the fields of survey methodology, nonresponse, computer-assisted interviewing and online surveys. He is also parttime professor in Statistical Information Processing at the University of Amsterdam. He can be contacted at Statistics Netherlands, PO Box 4000, 2270 JM Voorburg, The Netherlands, tel. +31 70 337 4995, e-mail [jbtm@cbs.nl](mailto:jbtm@cbs.nl).

**Ineke Stoop** is head of the department of ICT and Data Services at the Netherlands' Social and Cultural Planning Office. She is responsible for the surveys sponsored by the SCP and research on data quality. She has specialised in nonresponse issues. She is also a member of the Central Coordinating Team of the European Social Survey. She can be contacted at SCP, PO Box 16164, 2500 BD, The Hague, The Netherlands, tel. +31 70 3407671, fax +31 70 3407044, e-mail [i.stoop@scp.nl](mailto:i.stoop@scp.nl).



# Privacy, Trust and Self-Disclosure to Web-Based Surveys

Adam N. Joinson, Carina B. Paine, Tom Buchanan, Ulf-Dietrich Reips

## Abstract

Despite increased concern about the privacy threat posed by new technology and the Internet, in many environments there is little evidence that people's privacy concerns lead to privacy-protecting behaviours. In the present paper, two studies are presented in which people's privacy concerns are studied in light of their willingness to disclose personal information to web-based surveys. In Study 1, measures of privacy concern are collected, followed six weeks later by a request to complete a web-based survey containing requests for sensitive personal information, alongside measures of trust in the requestor, and perceived privacy related to the specific request ( $n = 759$ ). Structural equation modelling identified two privacy-related pathways determining self-disclosure: one based on respondents' perceived privacy and trust (a situational pathway), the other respondents' general privacy concerns and privacy-protecting behaviours (a dispositional pathway). In Study 2, privacy and trust were experimentally manipulated, and disclosure to a web-based survey measured. The results indicated that the impact of low privacy on self-disclosure is moderated by trust, such that high trust compensates for low privacy when examining item non-response. Implications for understanding the links between privacy attitudes, design of web-based surveys and item non-response to sensitive items, are discussed.

## Keywords

Privacy, trust, web-based surveys, self-disclosure

## 1. Introduction

Privacy has long been a concern of survey methodologists interested in data integrity and response rates (e.g. Andreasen, 1970). Surveys and research administered via the Internet, rather than using paper methodologies, been associated with reductions in socially desirable responding (Frick, Bächtiger & Reips, 2001; Joinson, 1999), higher levels of self-disclosure (Weisband & Kiesler, 1996) and an increased willingness to answer sensitive questions (see Tourangeau, 2004).

Similarly, survey methodology techniques that reduce human involvement in question administration also tend to increase responses to sensitive personal questions. For instance, compared to other research methods, when data collection is conducted via computer-aided self-interviews people report more health related problems (Epstein, Barker & Kroutil, 2001), more HIV risk behaviours (Des Jarlais, Paone, Milliken, Turner, Miller, Gribble, Shi, Hagan



& Friedman, 1999), more drug use (Lessler, Caspar, Penne & Barker, 2000), and men report less sexual partners, and women more (Tourangeau & Smith, 1996). Medical patients tend to report more symptoms and undesirable behaviours when interviewed by computer rather than FtF (Greist, Klein & VanCura, 1973). Clients at a STD clinic report more sexual partners, more previous visits and more symptoms to a computer than to a doctor (Robinson & West, 1992). Ferriter (1993) found that pre-clinical psychiatric interviews conducted using computer-mediated communication compared to face-to-face yielded more honest, candid answers. Similarly, automated or computerized telephone interviews, compared to other forms of telephone interviewing, lead to higher levels of reporting of sensitive information (see Lau, Tsui & Wang, 2003; Tourangeau, 2004).

There is some evidence that this response pattern might be related to privacy – for instance, Moon (1998) reports that participants are more likely to disclose socially undesirable behaviours to a geographically remote computer compared to one closer to them. Joinson, Woodley and Reips (2007) report that although personalized salutation in e-mail increases response rates to a web-based survey, it also increases item non-response to sensitive questions, implying a privacy cost to personalization (see also Andreasen, 1970 for a mail survey equivalent).

However, the relationship between privacy concerns and actual behaviour is neither straightforward, nor has any link been established incontrovertibly. There is evidence that although many Internet users express privacy-protectionist attitudes, this rarely translates to their actual behaviour (Acquisti & Grossklags, 2003, Jupiter, 2002; Metzger, 2006; Pew Internet and American Life Project, 2000; Spiekermann, Grossklags and Berendt, 2001). Privacy is also a disposition (i.e. an individual's general level of privacy concern) and an interpretation of a specific interaction or situation (i.e. an individual's perceived privacy at any one moment). In the present paper we examine the links between privacy (both dispositional concerns and situational interpretation), trust and self-disclosure to web-based surveys. Uniquely, self-disclosure is behaviourally measured, and we utilise both survey-based and experimental methods. In Study 1, measures of dispositional privacy, perceived privacy and trust are modelled in light of disclosure behaviour to a web-based survey. In Study 2, privacy and trust are experimentally manipulated, and the impact on disclosure to a survey examined.

## 2. Study 1

### Method

#### Participants

Participants were 759 members of an online research panel of Open University (OU) students called 'PRESTO'. The OU is an adult distance learning institution with nearly all students studying part time from home or work. PRESTO members are recruited annually and commit to completing six online surveys over 12 months. The sample is selected using stratified sampling, and is broadly representative of the whole student population (e.g. by age, gender and discipline and geographic location). Of the 759 respondents, 64% (487) were female, 36% (272) were male. The mean age of the sample was 42.58 years, (range=17–84 years,  $SD=11.11$ ).

## Materials

### *Time 1: Privacy dispositions*

A set of 16 privacy attitude items and 12 reported privacy behaviour items developed by Buchanan, *et al.* (2007) was given to participants. For all privacy items, responses were made on a 5-point scale (anchored at 'very concerned' and 'not at all concerned'). The privacy behaviour items consisted of six 'general caution' items (e.g. reading privacy policies, license agreements etc.) and six 'technical protection' items (e.g. removing cookies, clearing internet browser history regularly etc.; both anchored at 'always' and 'never'). The privacy concern items covered a variety of Internet-related privacy concerns (e.g. 'Are you concerned about online organisations not being who they claim they are?' 'Are you concerned that an email you send someone may be inappropriately forwarded to others?'). Participants were also asked about their Internet use (history, breadth of use, and time spent online). This data is not analysed in the present study.

### *Time 2: Self-disclosure and situational aspects of privacy*

Participants completed a 10-item measure of behavioural self-disclosure. In this measure, participants respond to a sensitive item such as 'How many different sexual partners have you had?' using one of three options: they could submit the default option 'please choose'; disclose the information requested; or choose an 'I prefer not to say' option. A non-disclosure score was calculated by summing the number of items where an 'I prefer not to say' option was chosen. A further six items of a non-sensitive nature (e.g. season of birth) were included as filler items.

Following the disclosure measures, participants completed measures of trust and perceived privacy designed to elicit their perception of the survey situation. Both measures were answered using a five-point Likert scale (anchored at 'Strongly Disagree' and 'Strongly Agree'). The trust measure comprised eight items that incorporated the major dimensions of trust (Bhattacharjee, 2002; Jarvenpaa, Knoll & Leiner, 1998): Benevolence (e.g. 'The intentions of this survey are good'; 'The data I have provided will be kept secure and not exploited'); Competence (e.g. 'This survey's authors have the appropriate skills and competence to conduct online surveys'; 'This survey is professional'); Reliability (e.g. 'This survey's authors are a dependable research group'); Integrity ('I do not doubt the honesty of this survey or its authors'; 'The authors of the survey are trustworthy') and General trust (e.g.; 'I felt comfortable giving my personal information'). The reliability for this measure was .91. Four additional filler items related to the design of the survey (e.g. 'The design of the survey was clear') and motivation (e.g. 'I felt motivated to complete this survey').

The perceived privacy measure had two questions relating to anonymity ('I felt anonymous completing this survey') and confidentiality ('I am sure that my responses will remain confidential'), answered using the same scale. The reliability for this measure was .73.

## Procedure

An invitation to complete the study was sent to panel members by e-mail. For Time 1, members were informed that the survey consisted of a series of questions about any privacy concerns they may have when they use the Internet, and their privacy related behavior. At Time 2, participants were told that the survey related to their 'Life experiences and season of birth', and that some of the topics covered in the survey may be sensitive, but that it was important for them to respond. The 'prefer not to say' option was outlined and they were told that the use of it would not imply any particular response.

At both time points, participants were informed that all information provided would remain confidential and that they could withdraw from the survey at any stage. For all items participants were prompted to use the full scale when responding and not only the labelled response options. Participant's responses were submitted at the end of each page of the survey.

The time 1 survey was left open for two weeks. Participants took, on average, 13 minutes to complete this part of the survey. Six weeks after data collection at time 1 was complete, an invitation to complete the time 2 survey was sent out to the same panel of participants. The delay between time 1 and time 2 was introduced to minimize the possible impact of the privacy measures on later disclosure behaviour. The time 2 survey was left open for two weeks. Participants took, on average, 12.3 minutes to complete it.

## Results

To examine the nature of any relationship between situational and dispositional aspects of privacy on self-disclosure to the web-based survey, structural equation modelling (SEM) using the AMOS software program was completed. One advantage of SEM is that it allows for the comparison of entire models in terms of goodness of fit to the data (also called confirmatory SEM), rather than the testing of single pathways. Three possible models derived from the existing literature on privacy, trust and behaviour were tested. The first, proposed by Nickel and Schaumburg (2004), Sultan, Urban, Shankar & Bart (2002) and Malhotra *et al.* (2004) predicts that the impact of perceived privacy, privacy concerns and personality on behaviour is mediated by trust (see Figure 1, Model 1). The second model tested, based on the work of Metzger (2004), proposes that in addition to this mediation, a separate path exists between privacy concern, past behaviour and privacy-related behaviour (Figure 1, Model 3). The final model is based on the results our own preliminary work (Paine, Joinson, Buchanan & Reips, 2006) that suggested two independent pathways, with the situational path comprising perceived privacy mediated by trust, and the dispositional path comprising Internet privacy concern (see Figure 1, Model 3). In all cases, the measurement error (not shown in the models) was weighted at 1.

A comparison of the three models, using goodness of fit indices (GFI), is presented in Table 1. Multiple GFI are used (Hu & Bentler, 1999). Specifically, the Chi-squared value divided by the degrees of freedom ( $\chi^2/DF$ ), the comparative fit index (CFI) and the root mean square error of approximation (RMSEA) were used, alongside the variance explained, to evaluate the models. As a rule of thumb, an adequate model fit to the data would have a CFI of .95 or above, a RMSEA of below .05, and a  $\chi^2/DF$  between 1 and 3.

Table 1: Goodness of Fit (GFI) indices, three models

	Fit indices				
Model	$\chi^2$	df	$\chi^2/DF$	CFI	RMSEA
Model 1 (Trust as mediator)	41.634	6	6.939	.932	.089
Model 2 (Trust as mediator, path through behaviour)	14.834	5	2.967	.982	.051
Model 3 (Separate pathways)	5.429	3	1.810	.995	.033

Using the multiple indices in Table 1, Model 3 represents the best fit to the data, and can be characterised as a good fit. This model also explained marginally more variance in the dependent variable (self-disclosure) than the other two models. This model was further refined using exploratory structural equation modelling, with the best fit presented in Figure 2.

Figure 1: Three potential models of the relationship between privacy, trust and selfdisclosure

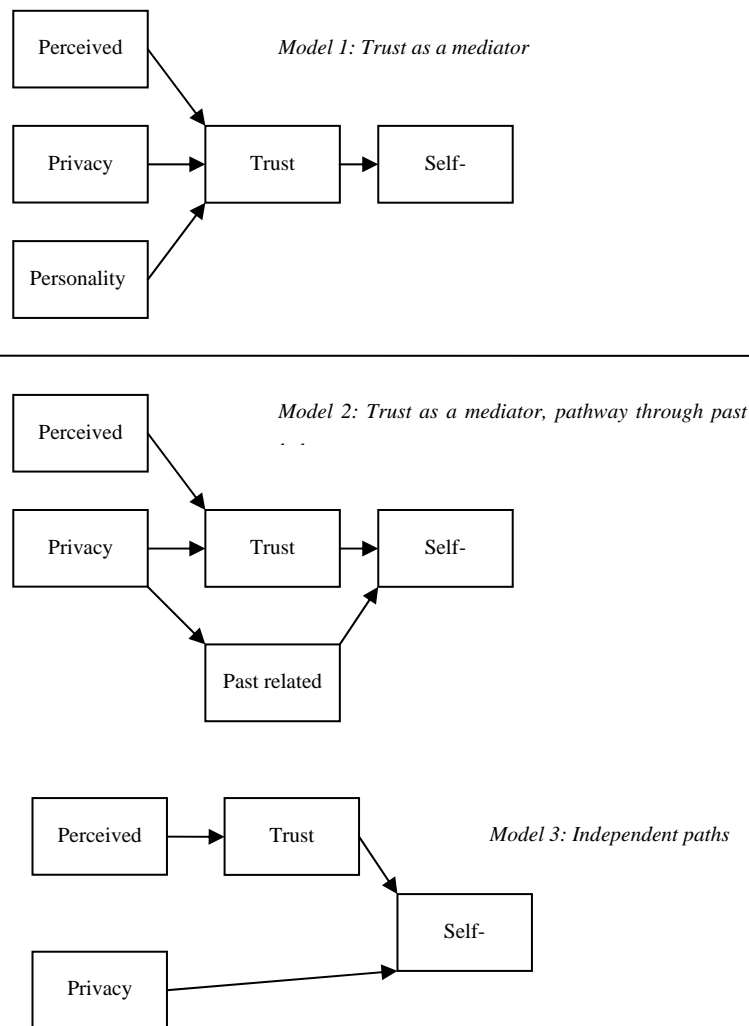
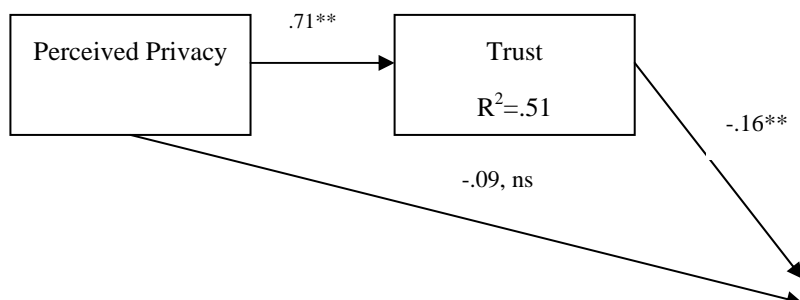
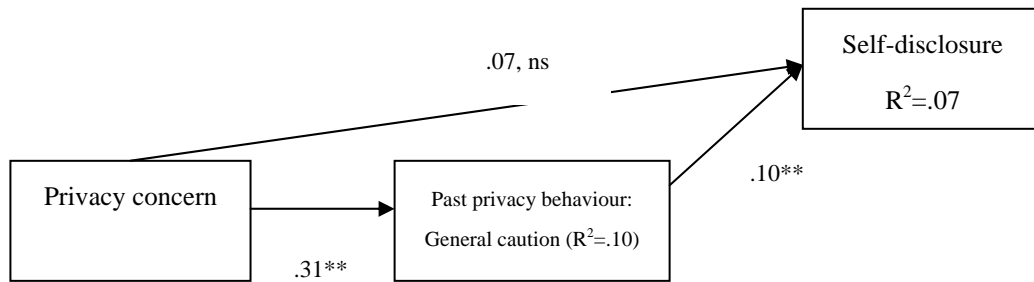


Figure 2: Best fit model, separate situational and dispositional pathways to predicting disclosure to a web-based survey





Note: \*  $p < 0.5$ , \*\*  $p < 0.01$ , \*\*\*  $p < .001$

#### Discussion

People's specific privacy concerns predicted their willingness to disclose personal information to a web-based survey six weeks later. Two situational measures were also collected during the second part of the study; trust and perceived privacy. Both these measures also predicted people's willingness to disclose personal information to the web site. The structural equation model outlining separate pathways associated with situational interpretation and dispositional privacy preferences was the best fit to the data. The pattern of results suggests that, on a situational level, much of the effect of perceived privacy on behaviour can be explained by the impact of perceived privacy on trust, which then impacts on behaviour. The results also suggest that people's past behaviours are critically important in understanding the links between their privacy concerns and disclosure behaviour.

However, the evidence that the impact of perceived privacy on self-disclosure to web-based surveys is mediated by trust should not be unequivocally accepted. The self-report measures used to assess situational trust and perceived privacy may be unduly influenced in Study 1 by the act of disclosure that preceded the completion of the measures. Furthermore, the relationship between privacy and trust at a situational level may be an interaction (i.e. moderation) rather than mediation, something best tested using an experimental methodology (Baron and Kenny, 1986). Study 2 examines the relationship between situational privacy and trust in light of disclosure using an experimental design. Privacy and trust are experimentally manipulated via different web survey designs, and disclosure measured.

### 3. Study 2

#### Method

##### Participants

Participants were 181 Internet users recruited via advertisements on psychology and survey request web-sites. The majority ( $n=144$ , 80%) were female (missing data for 5 people). Almost three quarters (73.9%) were based in the United States, with the remaining from 16 other countries (the majority UK and Canada). The age range spread from under 16 (one person, removed from the analyses) to over 65 years, with the largest proportion aged 20-24 years (31.7 years).

##### Experimental manipulations

For the privacy manipulations, the first page of the web survey contained either a strong or weak privacy policy developed using the guidelines identified by Culnan (1999) and previously used in the field (e.g. Metzger, 2004; Miyazaki & Krishnamurthy, 2002).

Specifically, the strong privacy policy included information on the type of information collected, that it would not be re-used or passed onto others, the security steps taken and provided full contact information for the researchers. The weak privacy condition did not include full disclosure of information collected and did not adequately protect information from re-use or security lapses. Pilot testing of the strong and weak privacy statement ( $n = 57$ ) confirmed that the strong privacy policy was perceived as stronger at protecting privacy than the weak statement ( $p < 0.01$ ).

Trust was manipulated in a number of ways, based on the work of a number of researchers (e.g. Bhattacharjee, 2002, Fogg et al. 2001). In the high trust conditions, the survey was hosted in an educational domain (\*open.ac.uk), while in the low trust condition it was hosted on a domain designed to reduce trust (www.surveylance.net). The high trust condition included an institutional logo, no spelling mistakes and no advertisements. The low trust condition incorporated advertisements for gambling and money transfer services (links deliberately broken) and spelling and coding mistakes. Otherwise, the text within the web-pages was identical. Pilot testing ( $n=20$ ) confirmed that the trustworthy site was rated as significantly more trustworthy compared to the untrustworthy site ( $p < .05$ )

#### Self-disclosure

Disclosure was measured using the same technique outlined in Study 1. Participants completed four sensitive measures, each with an 'I prefer not to say' option. These items were, "How many serious relationships have you had since age 18?", "How many sexual partners have you had?", "Are you a religious person?", and "What is your annual income?". The disclosure measures were followed by a series of demographic questions (Age, Gender, Country) alongside a season of birth to maintain the face validity of the study.

#### Perceived trust and privacy

Perceived trust and privacy was measured using the same questions and response options as outlined in Study 1.

#### Procedure

A link to the study was placed on a series of psychological and survey related web sites (e.g. web experimental lab). The study topic was advertised as 'Life experiences and season of birth'. If participants clicked the link to the study, they were randomly allocated to one of the four conditions using Javascript. A 'no script' option directed them to a separate study. Only one participant was directed to this study using this link, suggesting that Javascript was not an impediment to completion.

The experimental manipulation was embedded in the front page introducing the study. To proceed, participants clicked on a consent button, and were then taken to a seriousness check – participants indicated on this page whether or not their answers should be included in the analyses. Following this, they then proceeded to the disclosure items (arranged on a single page).

#### Results

Given the limited number of self-disclosure items (four) in the present study, participant's responses were dichotomised into those disclosing to all questions (76.1%) and those non-disclosing to at least one question (23.9%). The proportion of non-disclosers and disclosers in each condition is shown in Table 2.

Table 2: Percentage of full disclosure by condition

Trust	Privacy	
	High	Low
	High	Low
High	78.3%	82.1%
Low	85.1%	60.4%

A Chi-square test of association identified a significant association between condition and disclosure, specifically related to the combination of low privacy and low trust ( $\chi^2(1, 95) = 7.28, p < .05$  privacy in low trust,  $p = .19$ , *ns* privacy in high trust). The pattern of results suggests that the impact of privacy on self-disclosure to a web-based survey is moderated by trust.

## Discussion

The results of the present study demonstrate a strong moderator relationship between privacy and trust on self-disclosure to a web-based survey. Self-disclosure was only substantially reduced when a weak privacy policy was combined with cues designed to reduce trust. In the conditions that combined high trust with low privacy, or low trust with high privacy, there was no evidence that self-disclosure was reduced. The results of Study 2 suggest that the relationship between privacy and trust may be significantly more nuanced than one of simple mediation found in Study 1. Specifically, the results suggest that the impact of privacy on self-disclosure to a web-based survey is moderated by trust, but that this moderation is not linear. The results also suggest that to fully understand people's reactions to potential privacy threats posed by web-based survey questions, it is imperative to also measure their trust in the survey organization or sponsor.

## General discussion

The present studies are, to our knowledge, the first to include both situational and dispositional aspects of privacy and trust in the study of online disclosure to web-based surveys, both experimentally and survey-based. Importantly, in Study 1 we also separated the measures by six weeks, reducing the likelihood of any priming effect between the privacy measures and privacy related behaviour, and measured actual behaviour rather than reported actions or intentions.

The results of the two studies present strong evidence that privacy – both dispositional concerns and as designed into the specific survey – influences people's willingness to disclose personal information to a web-based survey. Second, the structural equation modelling in Study 1 suggests unique, independent effects of both dispositional and situational privacy processes that may have important implications for understanding the apparent disjuncture between people's reported privacy concerns and their actual behaviour. If there are no substantial links between people's privacy concerns in general and their interpretation of the situation, then it would be expected that any link between general privacy concerns and people's responses to sensitive items in a web-based survey would be weak or non-existent. Furthermore, in Study 1 the impact of perceived privacy on self-disclosure was mediated by trust.

The results of Study 2 show that trust acts to moderate the impact of reduced privacy on self-disclosure to a web-based survey. The relationship between privacy and trust found in Study 2 goes a long way towards explaining why people may be willing to forgo privacy concerns when faced with a trusted requestor, and why privacy is important when faced with a request from an organization one does not trust. The implications for survey organisations are critical – while much of the time personal or sensitive information may be requested by a survey, in the presence of high trust, this should not pose a problem for item non-response. However, in the absence of trust, we would argue that the same request is considerably more likely to be ignored.

The research literature on trust suggests that survey organisations may be able to boost trust in a number of ways. First, reputation is important in establishing trust – so survey organizations should seek to highlight their reputation not only to potential clients, but also to potential respondents. Trust and reputation can also be encouraged by including references to a ‘real world’ presence (e.g. through photographs of buildings, geographically located telephone numbers, full contact details). Other mechanisms for building trust in online surveying would be through demonstrations of competence – for instance, in the recruitment of respondents, management of panel membership, distribution of points or prizes and actual design and usability of the surveys. In Study 2, one of the key triggers towards lack of trust was the inclusion of adverts, which presumably led respondents to doubt the professionalism and benevolent intentions of the fake survey organization. For survey organisations, we would suggest that steps to protect respondent integrity and privacy would also lead to increases in trust.

The results of the present paper show that survey organizations ignore issues of privacy and trust at their peril. While the importance of respondent privacy in non-response (both item and unit non-response) is relatively well recognized, the potential interactions between privacy and trust in response behaviours identified in the present research highlight the importance of not only protecting respondent privacy, but also of engendering trust when requesting potentially sensitive personal information.

## References

- Acquisti, A., & Grossklags, J. (2003). Losses, Gains, and Hyperbolic Discounting: An Experimental Approach to Information Security Attitudes and Behavior. 2nd Annual Workshop on “Economics and Information Security”.
- Andreasen, A.R. (1970). Personalizing Mail Questionnaire Correspondence. *Public Opinion Quarterly*, 34, 273-277.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.
- Bhattacharjee, A. (2002) Individual trust in online firms: Scale development and initial test. *Journal of Management Information Systems*, 19, p.211-241.
- Buchanan, T., Paine, C., Joinson, A. and Reips, U-D (2006). Development of measures of online privacy concern and protection for use on the Internet. *Journal of the American Society for Information Science and Technology*, 58, 157-165.



- Culnan, M.J. (1999) Georgetown Internet Privacy Policy Study: Privacy Online in 1999: A report to the Federal Trade Commission. Washington DC: Georgetown University
- Des Jarlais, D.C., Paone, D., Milliken, J., Turner, C.F., Miller, H., Gribble, J., Shi, Q., Hagan, H. and Friedman, S. (1999) Audio-computer interviewing to measure risk behaviour for HIV among injecting drug users: A quasi-randomised trial. *The Lancet* 353 (9165): 1657–1661.
- Epstein, J.F., Barker, P.R., and Kroutil, L.A. (2001). Mode effects in self-reported mental health data. *Public Opinion Quarterly*, 65, p.529-550.
- Ferriter, M. (1993). Computer aided interviewing and the psychiatric social history. *Social Work and Social Sciences Review*, 4, p.255-263.
- Fogg, B.J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., Treinen, M. (2001). What makes web sites credible? A report on a large quantitative study. In: *Proceedings of the Conference on Human Factors in Computing Systems CHI 2001*. ACM Press, New York, pp. 61–68.
- Frick, A., Bächtiger, M. T., & Reips, U.-D. (2001): Financial incentives, personal information and drop-out in online studies. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet Science* (pp. 209-219). Lengerich: Pabst.
- Greist, J. H., Klein, M. H. and VanCura, L. J. (1973) A computer interview by psychiatric patient target symptoms. *Archives of General Psychiatry*, 29, p.247-253
- Hu, L., & Bentler, P.M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation Modeling*, 6 (1) p. 1-55.
- Jarvenpaa, S.L., Knoll, K. and Leidner, D.E. (1998) Is anybody out there? Antecedents of trust in global virtual teams. *Journal of Management Information Systems*, 14, p.29-64
- Joinson, A. N. (1999) Anonymity, disinhibition and social desirability on the Internet. *Behaviour Research Methods, Instruments and Computers*, 31, p.433-438.
- Joinson, A.N., Woodley, A., and Reips, U-R. (2007). Personalization, authentication and self-disclosure in self-administered Internet surveys. *Computers in Human Behavior*, 23, 275-285.
- Jupiter Research (2002). Seventy percent of US consumers worry about online privacy, but few take protective action, 2002. [http://www.jmm.com/xp/jmm/press/2002/pr\\_060302.xml](http://www.jmm.com/xp/jmm/press/2002/pr_060302.xml)
- Lau, J.T.F., Tsui, H.Y. & Wang, Q.S. (2003) Effects of two telephone survey methods on the level of reported risk behaviours. *Sexually Transmitted Infections*, 79, p.325-331.
- Lessler, J.T., Caspar, R.A., Penne, M.A. and Barker, P.R. (2000) Developing computer assisted interviewing (CAI) for the National Household Survey on Drug Abuse. *Journal of Drug Issues*, 30, 19–34.
- Malhotra, N. K., Kim, S. S. & Agarwal, J. (2004) Internet users' Information privacy concerns (IUIPC): The construct, the scale and a causal model. *Information Systems Research*, 15, 336-355.
- Metzger, M.J. (2006). Effects of Site, Vendor, and Consumer Characteristics on Web Site Trust and Disclosure. *Communication Research*, 33, 155-179

- Metzger, M.J. (2004) Privacy, trust and disclosure: Exploring barriers to electronic commerce. *Journal of Computer-Mediated Communication* Volume 9, Available online at <http://jcmc.indiana.edu/vol9/issue4/metzger.html>.
- Miyazaki, A.D. & Krishnamurthy, S. (2002) Internet Seals of Approval: Effects on Online Privacy Policies and Consumer Perceptions. *The Journal of Consumer Affairs*, 36, 28-49.
- Moon Y. (1998) Impression management in computer-based interviews: the effects of input modality, output modality, and distance. *Public Opinion Quarterly*, 62, p.610–22
- Nickel, J. and Schaumburg, H. (2004) Electronic privacy, trust and self-disclosure in e-recruitment. Late breaking results paper presented at CHI 2004, 24-29 April, Vienna, Austria.
- Paine, C.B., Joinson, A. N., Buchanan, T. & Reips, U-D. (2006). Privacy and Self-Disclosure Online. Work in Progress paper presented at the Conference on Human Factors in Computing Systems (CHI), Montréal, Canada, April 2006, pp. 1187-1192.
- Pew Internet and American Life Project (2001) Fox, S., Rainie, L. Horrigan, J. Lenhart, A., Spooner, T. and Carter, C. Trust and privacy online: Why Americans want to rewrite the rules. Available at <http://www.pewinternet.org>
- Robinson, R., & West, R. (1992) A comparison of computer and questionnaire methods of history-taking in a genito-urinary clinic. *Psychology and Health*, 6, 77 – 84.
- Spiekermann, S., Grossklags, J. & Berendt, B. (2001) E-privacy in 2nd generation E-Commerce: privacy preferences versus actual behavior, in: *Proceedings of the Third ACM Conference on Electronic Commerce*, Association for Computing Machinery (ACM EC'01), Tampa, Florida, US, pp. 38-47.
- Sultan, F., Urban, G.L., Shankar, V. & Bart, I.Y. (2002) Determinants and role of trust in e-business: A large scale empirical study. MIT Sloan School of Management working paper 4282-02. Available online at <https://dspace.mit.edu/bitstream/1721.1/1826/2/4282-02.pdf>
- Tourangeau, R., & Smith, T.W. (1996). Asking sensitive questions: the impact of data collection mode, question format, and question context. *Public Opinion Quarterly*, 60, 75–304
- Tourangeau, R. (2004). Survey Research and Societal Change. *Annual Review of Psychology*, 55 (1), 775-802.
- Weisband, S. and Kiesler, S. (1996) Self-disclosure on computer forms: Meta-analysis and implications. *Proceedings of CHI96*. Available online at [http://www.acm.org/sigchi/chi96/proceedings/papers/Weisband/sw\\_txt.htm](http://www.acm.org/sigchi/chi96/proceedings/papers/Weisband/sw_txt.htm)

## About the Authors

**Adam Joinson** is senior lecturer in the School of Management at the University of Bath. Previously he was senior lecturer at the Institute of Educational Technology, The Open University where he managed online survey panels. He can be contacted by e-mail [joinson@gmail.com](mailto:joinson@gmail.com), or via his website: <http://www.joinson.com>

**Carina Paine** is a research associate at the Open University, who works on the ESRC 'Privacy and Self-Disclosure Online' project (<http://www.prisd.net>)

**Tom Buchanan** is a Reader in Psychology at the University of Westminster. He has been researching the use of the Internet in personality assessment since the mid 1990s.

**Ulf-Dietrich Reips** holds the position of Assistant Professor in the newly founded department of Social and Business Psychology at the Department of Psychology, University of Zürich. He has been researching the use of the Internet to conduct psychological experiments since the mid 1990s.

## Acknowledgements

The research reported in this paper was funded by the Economic and Social Research Council (ESRC) E-Society Programme (RES-341-25-0011) awarded to Joinson, Buchanan and Reips.

# Using Chat Tools to Perform Evaluation Interviews

Eve-Marie Larsen

## Abstract

To assess the administrative complexity of mobility for non-European students moving throughout Europe on the ERASMUS MUNDUS Network and E-business Centred Computing (NeBCC) program an extensive evaluation study is being undertaken throughout the period of student mobility. This evaluation is undertaken in a mixed mode with students completing a series of online surveys and participating in group interviews at each location they are studying at. To maximize the quantity and quality of the feedback that is given the interviews are strictly confidential and are conducted by a researcher who is non operational in the programme to reduce the problems associated with an unequal power balance. This allows the students to open up to a greater degree and share their feedback and their experiences. Other factors which may also effect how freely a student participates in the interview include the effect of the physical environment and the students own assumptions of that environment.

Suler(2004) identifies the online disinhibition effect where people feel more uninhibited and thereby are willing to speak openly and share information more freely when communicating virtually using online tools, such as chat tools, than they are in a face to face situation. Six factors have been identified as reasons for this effect including the inherent minimization of authority associated with online communication as well as invisibility, asynchronicity, solipsistic introjection, dissociative anonymity and dissociative imagination.

This paper considers the phenomena of online disinhibition and outlines some research being undertaken in order to exploit this phenomena by interviewing students in an online environment in an attempt to improve the results of the quality control measures on the NeBCC program. The suitability of using online chat tools to perform interviews is discussed with reference to the limitations of face to face interviewing techniques and the methodology and early results of this research are highlighted.

## Keywords

Survey methodology, Quality assurance projects, Online disinhibition, Interview techniques, Mobility

## 1. Introduction

The NeBCC degree is coordinated within the ACET Centre at the University of Reading (UoR) and is a joint degree run in conjunction with Aristotle University of Thessaloniki

(AUPh) and University Carlos III Madrid (UC3M). The program is a mobility scheme and students study for a total of eighteen months taking 4 modules at each of the partner universities, and a further 24 week research project based at one of the partner institutions. The students on this programme are solely responsible for arranging their mobility throughout Europe with the coordinators at each location providing support where necessary. Only a limited number of students enrolled on this programme originate from within Europe with the majority of students coming from “third countries” which adds significant complexity to their mobility in terms of administrative requirements which must be addressed if the student is to complete the course. The program is in its second iteration with the first intake of students having just completed their research project. The second intake of students has just completed their studies at the second location and is in the process of mobility to location three. Applications for the third intake of students have just been processed.

In order to raise the visibility of issues associated with the complex nature of the student experience on this program and also the unique problems associated in terms of maintaining academic uniformity and administrative coherence at a program wide level self-evaluation is a mandatory consideration. Self evaluation allows feedback to be given to the managing partners involved in the degree which can be used to drive the evolutionary changes essential for this new form of academic collaboration. The ACET centre has a rigorous Quality Assurance Programme (QAP) coordinated by researcher’s independent to the teaching and day to day activities on the program. The QAP was launched informally shortly after the first period of mobility for the first intake of students and has since been integrated into the formal regulatory processes on the program All aspects of the program and the students experiences are evaluated by an extensive surveying program and in-depth interviews with the students and professors. Evaluation of the students’ experience involves a series of confidential surveys and group interviews at each location. Each survey consists of questions regarding the nature of mobility, the content of the course, the use of technology on the course and the effectiveness of the organizational infrastructure. An important factor to effective interviewing is that of building trust and open communication between the interviewer and interviewee. In order to achieve this goal the interviewer was always the same person, a student and presented as a “voice” for the student body.

## 2. QAP Round One

The first round of the QAP resulted in various initiatives being undertaken within the program relating to improving particularly the visibility of student issues and the student experience. These initiatives include implementing a class representative as well as mentoring schemes and the formalisation of the Quality Assurance Project. Another initiative was the enabling of an online community as a channel of communication between the students raising the visibility of student issues throughout the student body and allowing students to engage in collaborative problem solving. It is also used to raise the visibility of student issues to the administrators and management teams by the use of a “Contact the administrators” board as well as high level analysis of posting activity. Students are invited to join the community as soon as they are accepted on the programme. The technology enabling the community is open source and is an integrated solution based on the phpBB bulletin board software and the phpmychat chat server. Students have access to the community on acceptance to the course and use the chat server for synchronous chat sessions with key faculty members as well as with students from previous intakes as a means to solicit advice and information about the

program. Use of the community continues throughout the student's registration on the course and student access is never revoked.

The interviews undertaken with the first intake of students took place in an office setting set aside specifically for that purpose. On completion of these interviews it was observed that the same information was often being given repeatedly and consideration was given to the technique being used to gain feedback in terms of the cost (both time and financial) the mobility of the interviewer and the quality and quantities of new feedback that was being generated. Of particular concern was the fact that considerable effort was being made to get the feedback and on the whole after interviewing a sample of the class (approximately 50%) little new or relevant information was being given. Analysis of the feedback that was given highlighted the fact that students typically were comfortable answering factual questions, but less comfortable answering speculative or personal questions.

Other problems included the issues associated with the physical environment that the students were being interviewed in and their perceptions of this environment. The simple presence of a Dictaphone and a desk and the fact that the interviews were being conducted within the university buildings by someone that the university has commissioned to conduct the research inherently reduces the camaraderie and the implication that the researcher was on the student "side" striving for information pertaining to the students situation in order to improve their situation. The presence of these physical cues inherently establishes a power imbalance and gives the impression of an authority figure. The fact that the students are called by the course coordinators to be in a certain place at a certain time to partake in an interview creates a formality and authority to the proceedings. All of these factors could be an inhibitor to free speech and the building of trust.

The following techniques were considered and rejected as alternative solutions in order to address these issues:

**Performing the feedback interviews using AV conferencing tools.** AV conferencing, in effect, mimics the interaction that may occur in the face to face environment. Arguably, AV conferencing can be considered a substitute for face to face contact. As the scope of this work is to get an alternative methodology in order to achieve a greater quantity or quality of feedback this option was not considered a relevant solution. Furthermore, one of the major issues highlighted by the students on the program is the lack of reliable internet connectivity at their accommodation and also, in one location the lab. The technical complications of implementing AV conferencing made this option a consideration for a future project.

**Interviewing a sample of the class** Another possibility was that of only interviewing a sample of the class and accepting the feedback given as representative of the class's experiences. While a highly practical alternative it was considered that part of the success of the evaluation program rests on the relationship with the researchers and the fact that the students feel that their voice is listened to and their opinion counts. In fact, in one survey a student suggested the evaluation program be improved by "*reducing the number of guys in each group.....*" in order to allow for more time to be spent giving their feedback. Performing the evaluation interviews on only half of the class undermines the sense that the feedback is valuable and would give the students the impression that a process is being completed, one which is less about their opinion and more about completing some policy requirements. Also, although little or no new feedback is given once the whole class has been interviewed sometimes the occasional interesting piece of new feedback is given and this opportunity

should not be missed. Finally, this methodology does nothing to improve the issues associated with formality and authority.

### 3. Chat tools and the online disinhibition effect

The solution decided upon was to use the chat server on the NeBCC community as a means to ensure that the students are all given the chance to give feedback in all areas and reduce the issues relating to formality and authority. As the community is an existing community it affords the opportunity to exploit the existing technology as well as the community spirit and relationships that already exist. Furthermore, the online disinhibition effect proposed by Suler in 2004 highlights the possibility of students responding to the interviews differently and the potential that more or new feedback could be generated.

The basic premise of the online disinhibition effect is the fact that text chat is a form of social lubricant. People often act in a different manner than they would in a face to face environment with mixed results. In its most positive or benign form people are able to communicate problems or issues they are facing more freely, un-inhibited by the physical nature of presence. In its most negative or toxic form people can feel more inclined to act or react in an inappropriate or un-proportionate manner, responding angrily or violently, often at odds with how they would react in the “real” or physical world. Analysis of the community postings on the online NeBCC community show, to date, no examples of toxic disinhibition in the community and the culture of the community is one of open and honest information exchange. The methodology proposed in this paper attempts to harness benign disinhibition by performing interviews online using the community chat tool in order to obtain a new or different perspective on the students experience on the program.

Suler (2004) identified the following reasons for altered behaviour in the online environment

Dissociative anonymity (you don’t know me) – this characteristic relates to the fact that the chatter feels the disinhibitive effect of anonymity and thereby feels free to respond more openly or out of character. This relates partially to the scenario developed in this paper as although the students know and have a relationship via the community with the researcher conducting the evaluation this relationship was formed in the first instance via the online community when the students did not know the researcher and the online relationship has developed from that point.

Invisibility (You can’t see me) - This attribute is relevant in this experiment as not only are the students physically separate from the possible perceived source of authority, in this case the researcher, they were also in some cases separate from each other.

Asynchronicity (See you later) – this is a particularly relevant factor for this experiment. The fact that the interviewer and the students are not all co-located means this could be a potential source of disinhibition.

Solipsistic Introjection (It’s all in my head) – this phenomena details the similarity between the conversations that one may have in ones own head with the conversations one would have using text chat. Much as with internal dialogue one may assign a character, appearance and personality to perceived persona in an online environment and then act accordingly based on these assumptions. This phenomenon is extremely relevant in terms of this research when considering the relationship the students had prior to arrival on the course with the researcher undertaking the interviews. On the first face to face meeting

with the interviewer four separate students expressed surprise when they realised that the researcher who had been communicating online with them was female, despite the researcher having what is a commonly recognised female name and a female avatar, having taken the assumption that the researcher involved would be male. Despite a prior three month virtual relationship this assumption was somehow never revoked indicating that the process the students were experiencing was internal in nature, formed within their own imaginations. Pre-defined assumptions were made sometimes ignoring obvious external prompts from the real world and virtual world. Much the same as one would with an internal dialogue one feels free to speak in ways one would never consider in the real “physical” world and one becomes emboldened to speak freely.

Dissociative imagination (It’s just a game) – a relevant attribute for this work this characteristic looks at the process where ones online persona and world is distinct from that in the real world. The responsibilities or repercussions of actions online are not given the same consideration as the person would give in the physical world.

Minimizing authority (we are equals) – possibly the most fundamental possible advantage of this method of interviewing is the possibility of the minimization of sense of authority. By removing the physical environment and its cues, the physical presence of the researcher and the formal arrangement of the interview allows for the impression of equality. This is especially true when considering in the first place the relationship between the researcher and the students was fostered in an online environment as one student assisting them then this minimisation of authority is particularly relevant.

#### 4. Methodology applied

A traditional interview script was developed based on the scripts used in previous years and previous projects undertaken within the ACET centre. Questions were divided into topics and the topics categorised based on how much feedback was anticipated. These topics were then allocated a rating hot, neutral or cold based on discussion with the coordinators of the course, analysis of the online community posting and viewing history, and consideration of the interviews that were undertaken in the previous year. The hot topics were the topics that were deemed to be the most controversial or problematic thereby being the most likely to generate the highest amount of feedback. The topics for consideration were:

Administrative Issues - These questions allowed students to give details of their administrative experiences on the course. Of particular relevance were questions pertaining to the student’s accommodation, VISA application process, financial arrangements and the administrative support that they had received to date. Due to the fact that the students are required to satisfy a tough set of requirements to ensure their place on the course and that failure to satisfy these requirements would result in losing their place on the course it was anticipated that this would be one of the hot topics of discussion.

Teaching and Learning - These questions related to the actual academic content and standards of teaching across the course. In general the students registered on the course are highly motivated in their studies and are very keen to discuss their progression throughout the course, their opinion of the professors and academic quality at each location. Due to these facts this topic was allocated also a hot rating.



**Mobility** - Moving from one location to another has its own burdens associated with it from the issues associated with physical mobility, for example the arranging of transportation of self and belongings, to the financial implications of mobility. In previous years this topic was of interest and did receive a reasonable amount of feedback but was not one of the topics the students were particularly passionate about discussing. Based on this mobility was given a neutral rating.

**Culture** - This program is particularly unusual in that the students are exposed to a rich cultural diversity not only within the class room but also within their living environment as they move from country to country. The issues surrounding this topic include the problems associated with the languages and cultural norms of each location as well as their own expectations of perceptions based on their own cultural background and upbringing. While this topic did receive some interesting feedback in the previous intake it was not one of the problematic areas thereby it received a neutral rating.

**Use of technology** - The program is technology managed and this topic addresses the use of the Blackboard LMS on the program from the functionality offered by the tool to how it was used in the different locations. While this topic was of interest to the administrative and management teams this was not a topic of great interest to the students in previous intakes. As a result this topic received cold rating.

**The ERASMUS MUNDUS community** - This topic related not only to the use of the online community but also to the dynamic of the class and bonds that were formed. Given that the students work very closely together on academic projects and also often live together, the pressured environment that they are in is an important factor for consideration and it is essential that the classroom dynamic is a positive and healthy one for the well being of the group as a whole. In the previous intake the students were particularly reluctant to discuss the classroom dynamic and give any anecdotal information about the group dynamic or cohesiveness. Therefore this topic received a cold rating.

The students were divided into 6 groups based on a number of variables. The first variable for consideration was their online posting history on the NeBCC community. Consideration was given to the number of posts on the NeBCC community with the top 6 posters being randomly assigned a position in each of the groups. The bottom 6 posters were allocated a group at random. The reason for this initial categorisation was to minimise the confounding effect of their usual online behaviour and reduce the possibility of having one group of students reluctant to communicate online and another group of particularly keen students. The remaining 15 students were then categorised by the course coordinators based on their face to face experiences with the students. They were categorised into two categories: familiar and non familiar. The reason for this category is that based on previous experiences some students were comfortable making comments and complaints online but not face to face. One particular incidence of this is one of the high profile members of the online community who was one of the most prolific and occasionally over zealous posters but was virtually unknown by administrative staff not involved with the online community. The remaining students were randomly assigned to a group.

The interview scripts were then divided into two scripts with each script containing One hot, two neutral and one cold topic. The two neutral topics were asked to every group to act as a controlling group for the questions as a basis for comparison of how well balanced the groups were. The scripts created were as follows

**Script A :**Administration, Mobility and Culture, The use of technology

**Script B:** Teaching and learning, Mobility and Culture, The EM community

The six groups generated were then classified into Group A or Group B. The students allocated to Group A received interview script A in the formal face to face interview, the students in Group B received interview script B in the formal face to face interview. For the online interviews this scenario was reversed with each group receiving the script of questions they had not been given in the face to face interviews.

#### 4.1 Analysis Methodology

There are two stages which will be undertaken in the analysis of the data generated from this project. This paper will report the early results (those entitled phase one). The results of the second phase of analysis are out of the scope of this paper.

##### 4.1.1 Phase one

This phase involves two stages

High level review of the content and tone of the online interview by using the chat scripts and audio files generated throughout the process to generate a report for the management committee Based on the process of generating the report, which is one of the main objectives of conducting the interviews in the first instance, some initial comments and observations can be made into the feedback generated by both methodologies and the usefulness of the comments generated.

Implementation of a survey to get initial feedback from the students on their feelings about being asked for their feedback face to face and online.

The survey given to the students started with a section relating to the group they were placed in and to obtain an opinion of their perception of the evaluation process in general. The survey then went on to ask some high level closed questions about the nature of the online and face to face interviews in terms of the level of formality, constructiveness, administration of the interview, level of comfort with the format of the interview and the relevance of the questions. Students were then asked to identify their preferred method of interview and some reasons for this selection. Finally students were asked to give some open ended feedback about the face to face interview, the online interview, the feedback process in general and the content of the feedback left by their group. All closed questions were 5 point likert scale questions where students were asked to state, in their opinion, how strongly they agreed with a set of statements with the options of Strongly Agree, Agree, Partially agree, Disagree and Strongly Disagree.

Based on this a comparison of the students perception and experiences of the interviews face to face and online and an initial assessment of the usefulness of the two methodologies can be gained.

##### 4.1.2 Phase two (in progress)

This phase involves a full transcription of the face to face interviews and codifying of the results using a standard CAQDAS tool yet to be decided. Feedback will be categorised by topic and then classified as

Factual – information about concrete facts and events

Speculative – information offered which is not factual but is based on opinion or supposition

Personal – personal information relating solely to the respondent and their circumstances or opinions

An analysis of the number of responses made by each group in the control questions will be made to ensure the validity of the groups and further comparative analysis into the quantity and quality of the feedback that was left using both methodologies will be made. Based on this analysis final conclusions will be able to be drawn as to whether online interviewing can be used as an effective tool and scenario based recommendations can be made for the use of online interviewing techniques for evaluation of academic programs.

## 5. Results

At the highest level of analysis the presence of some form of online disinhibition is evident. The tone of the feedback appeared much more informal and personal and students seemed to be more confident in expressing their personal opinions and making observations about their class mates. Examples of comments that illustrate this:

On discussing the timekeeping of the other members of the group:

*“it is bad the way our guys conduct themselves”*

In the face to face interviews there has only previously ever been one direct criticism of another class member and the issues surrounding this criticism were very serious. Other than that the students effectively closed ranks in the face to face interviews and, while happy to make critical observations about the course and the professors, never made reference to any internal disagreements or misgivings.

On discussing the class representative:

*“i think it was a good idea and we had a switched on guy”*

In the face to face interviews the students were happy to discuss the validity of having a class rep and improvements that could be made in terms of the administration of this initiative and extra support required. Again there was a reluctance to actually discuss how the current elected candidate was getting on personally.

On discussing the amount of administrative support given throughout the program

*“its only that some of our guys want spoon feeding when they can do some things themselves”*

In the face to face interviews there was a sense of collective dissatisfaction about certain administrative issues and the degree of support given over some of the issues regarding obtaining VISA's and accommodation. At no point in the face to face interviews was there any discussion that there were any other issues other than a failure on behalf of the administrators.

On discussing the teaching staff on the course

*“Databases was good...we had a good teacher and a handsome one.....all the girls really paid attention in this class....”*

This comment was completely at odds with the formal tone found in the face to face interviews and highlights the disinhibiting effects identified by Suler.

### 5.1 Results from the survey

In general the results from the survey are mixed. The students expressed a direct preference for the face to face interviews with 73.33% citing face to face interviews as their preferred methodology. Reasons given for this included

*“Because it’s more comfortable to discuss like a conversation”*

*“Because its easier to express what you mean....English is not my native language”*

*“Easier to know whether the interviewer is interested in us or not”*

Reasons given by the respondents who selected a preference for the online interviews included

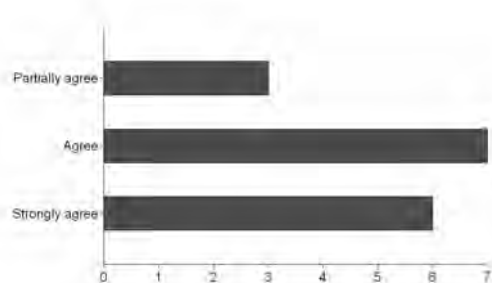
*“I prefer online chatting”*

*“Ease of making your point of view and the ability to refer to what has already been said in order to make further contributions”*

Before considering the effectiveness of the method of evaluation consideration was given to the interviewer/interviewee relationship and the group cohesiveness of the groups the students were assigned to in order to rule out any secondary factors which may have confused the results of the survey. Students were asked some questions to assure that they understood that the interview process was confidential and to establish that they trusted and felt comfortable with the interviewer.



*I felt comfortable with the interviewer confidential*



*I trust that the feedback I give is*

94% of respondents felt comfortable speaking out in groups and 94% also felt comfortable speaking in the groups they were assigned to. In general the group cohesiveness was high with the students usually agreeing with the feedback given by their group and feeling that the feedback was relevant and constructive

Overall the students felt they were given the opportunity to express their own opinions and that their group did not complain unreasonably. No significant issues were raised by any of the groups with regard to the group that they had been put into or the content of the feedback left by their group.

The students were then asked to consider, in separate sections, their face to face interview and their online interview. The results for both methodologies were fairly consistent with students feeling they were able to give all the feedback that they wanted using either

methodology, that the degree of group cohesiveness was unaffected by the methodology of interview, that the questions asked were relevant, interviews were well organised and most students in both cases felt the feedback given would be taken seriously. The three main areas where differences were perceived by the students were in the formality of the interview, how comfortable the students felt with that mode as a way of giving feedback and with how constructive the interviews were.

All students responded that they felt comfortable giving their feedback in the face to face interview but some of the students were less sure of giving feedback in the online environment with approximately 31% of respondents only partially agreeing with the statement "*I feel comfortable giving my feedback in the online environment*". The majority of respondents felt the face to face interviews were formal with only 14% of respondents disagreeing with the statement that "*the face to face interview was formal*", 29% partially agreed with this statement and the remaining respondents either agreed or strongly agreed with this statement. However, for the online interviews only 38% of respondents Agreed or strongly agreed with the statement that "*the online interviews were formal*" with 39% only partially agreeing and the remaining respondents disagreeing with this statement. This is a significant result in terms of defining that the students felt the online interviews were less formal than the face to face interviews. This fact is supported by the evidence that most online interviews ran late due to lateness of the students. In the face to face interviews the students were all in the appointed place at the appointed time. With the lack of formality associated with this mode of interviewing came also a casual approach to the process which caused some problems in terms of administering the interviews but also is a potential reason the students felt more comfortable to give personal or speculative feedback.

All respondents felt that the feedback given in the face to face interviews were constructive with 31% only partially agreeing with the fact that "*the online interviews were constructive*". This discrepancy could be considered due to the perceived lack of formality of the interviews causing a more casual approach to be taken to the online interviews. Another factor that could have influenced how effective the online interview's was the length of time allocated to them. While 28% of students felt that the face to face interviews should have been longer 46% felt that more time should have been given to the online interviews despite the fact that the same length of time was allocated to both face to face and online interviews.

Finally students were asked if they felt online interviewing was a viable alternative to face to face interviewing in terms of collecting feedback about their experiences. Despite the mixed results 54% felt it was a viable alternative, 23% partially agreed it was a viable alternative and the remaining 23% felt it was inappropriate as an alternative.

## 6. Conclusions and further work

While the Online disinhibition effect is evident from the chat transcripts more analysis is required in order to identify statistically the degree of impact of this effect in terms of the quality and quantity of feedback given. A comparative analysis evaluation at a personal level as to the profile of students who appeared to be most uninhibited in the online environment with their real world behaviour should also be considered.

There are many reasons why this disinhibition has occurred as identified by Suler and there is evidence to support that the online environment has produced a less formal, fluid environment

where the students feel more comfortable in expressing personal feelings and opinions. However, there were some issues with the constructiveness and duration of these interviews which raises some questions with regard to using online interviews as an alternative to face to face interviews. Based on the early results of this experiment it is recommended that careful thought be given to the appropriateness of using online interviewing techniques and that these techniques should only be considered when the required feedback is supplementary to formal self evaluation processes or the feedback is required is :

## References

- Personal – the online environment facilitates the required disinhibition and the respondent appears more willing to share personal information. This would be particularly effective in getting students to self evaluate and reflect on experiences.
- In small, focussed areas of interest – this methodology appears to take more time in order to get all the information required. This methodology would be successfully applied to a small list of questions rather than a large list of varying questions
- However, it is considered that in order to have a complete, rounded understanding of student's experiences of mobility online interviewing is a necessary activity as it provides an opportunity to get a more personal understanding of the student's perceptions.
- Buckingham A and Saunders P (2004) *The Survey Methods Workbook* ISBN 0-7456-2255-3
- Fielding N and Lee R (1998). *Computer Analysis and Qualitative Research* ISBN 0-8039-7483-3
- Krippendorff, K. (1975). Some principles of information storage and retrieval in society. *General Systems*,20:15 { 35}.
- Larsen EM, Alexandrov NS and Alexandrov VN (2007) *Defining the User Requirements for Virtual Learning Organisations* awaiting publication
- Suler J (2004) The online disinhibition effect, *CyberPsychology and Behaviour*, 7, 321-326
- Wolcott H(2001) *Qualitative Research* ISBN 978-0761924296

## About the Author

**Eve-Marie Larsen** is a final year PhD student in the Applied Computing and Emerging Technologies (ACET) Centre based within the School of Systems Engineering at the University of Reading. Her research focuses on Requirements Capture & Systems Design for collaborative working in Education. She is responsible for the Quality Assurance Project (QAP) on the ERASMUS MUNDUS Network and E-business Centred Computing program coordinated within the ACET centre. Her involvement with the QAP involves the extensive surveying and interviewing of mobility students, analysis of survey and interview data and the development of collaborative tools and methodologies for collaborative working throughout the program. Email e.larsen@reading.ac.uk



# Separating the Wheat from the Chaff: Ensuring Data Quality in Internet Samples

Reg Baker & Theo Downes-Le Guin

## Abstract

The emergence of online access panels as a sample source has been a major factor in the rapid growth of online surveys. The increased use of non-probability samples in place of probability samples has generated considerable debate and discussion within the research industry about the levels of sampling error and bias inherent in these samples. At the same time, market research buyers, for the most part, have found the obvious advantages in cost and turnaround compelling enough to consciously overlook some of the unattractive aspects of online access panels. But this may begin to change as a new set of non-statistical concerns emerge. These concerns are focused on issues such as respondent qualifications, increased satisficing, and possible professionalization of survey respondents. In response to these concerns, both panel companies and research agencies have begun to look more closely at panel composition and online survey research results. Though this scrutiny happens with varying degrees of formality, the results across organizations have largely converged on three definitions of potentially problematic panel respondents: professionals (i.e., individuals belonging to several panels and completing a large volume of surveys); inattentives or satisficers (i.e., respondents who complete surveys but do so with little cognitive effort); and fraudulents (i.e., respondents who misrepresents themselves in order to qualify, either at the panel registration phase or in individual surveys). This paper reviews the emerging research record as both panel companies and research agencies begin to look more carefully at the behavior of online panelists. It describes the steps that researchers can take to ensure data quality when doing research that uses online access panels.

## Keywords

Access panels, online research, data quality, panel quality

## 1. Introduction

To describe the growth of online research over the last ten years as “dramatic” is to understate the case. Virtually non-existent in 1996, *Inside Research* (2006) reports the global spend in online market research for 2006 to be \$2.6 billion. No surprise to anyone the bulk of this research, at least in the US, relies on online access panels. It is fair to say that online panels have made online research go.

The emergence and increasing acceptance of online panels as a legitimate sample source has not been without controversy. Probability sampling has been a key underpinning of survey



research since the 1930s. In the 1980s telephone surveys using random-digit dialing (RDD) emerged as the most popular methodology for U.S. market research studies, largely due to the availability of a sampling frame that included virtually the entire U.S. population. The most relevant exception to that generalization became mail panels, databases of volunteers, sometimes numbering in the hundreds of thousands, who had signed up through one means or another to do postal surveys. The general acceptance of this methodology throughout much of the industry likely paved the way for acceptance of additional non-probability sampling approaches.

Enter Gordon S. Black, who in the mid-1990s began to assemble a large panel of volunteers to do surveys on the Web. In 1998, he summarized the challenge as follows:

*To do what we are doing we have had to set aside the staple of our industry -- the simple random sample... Random sampling is a very powerful tool in every avenue of science and industry for increasing the accuracy of estimates while decreasing the cost of the process... We are not challenging the validity of random sampling... We are instead investigating whether findings from huge samples of Internet respondents, coupled with sophisticated weighting processes, are as accurate as anything done on the telephone or door-to-door.*

The extent to which Black's key requirement of sophisticated weighting processes has been met is outside the scope here. Suffice it to say that there are arguments both pro (Taylor (2007)) and con (Schonlau (2004)). The focus of this paper is the current challenge facing the panel industry, what we have described elsewhere (Downes-Le Guin et al. (2006)) as "a perfect storm." At its heart this challenge is about the evolving behavior of online panelists and the degree to which surveys done with them produce results that our clients can safely use to make business decisions.

## 2. The New Online Panelist Pathologies

One sure fire way to know when we have a problem is when it is given a name. So it is with online panelists. Panel companies and the researchers who use their services have begun to recognize three important panelist pathologies (See, for example, Smith and Brown, 2005; or Downes-Le Guin, 2005). Each of these pathologies is measurable to some degree, and each can affect individual survey results if present in sufficient magnitude.

### Hyperactives

The first of these has been dubbed "hyperactives," sometimes also referred to as "professional respondents." These are panelists who do lots of surveys and who may also belong to multiple panels. There is some evidence that response patterns on certain types of surveys may differ based on the number of previous surveys the respondent has completed. (See, for example, Coen et al. (2005); and Cartwright and Nancarrow (2006)). The evidence on multiple panel membership is even more compelling. Table 1 combines findings from two studies of multiple panel membership, one by Smith and Brown (2006) using Harris Interactive in the UK and US and the other by Fine et al. (2006) using AMR Interactive in Australia. The data suggest that respondents who belong to more than one panel are more likely to be female and of lower socio-economic status. Fine et al. conclude that "there are major differences between people who are members of only one panel and those who are members of multiple panels... These differences are seen in demographics, attitudes, and behaviour." A

comprehensive study of 19 panels in the Netherlands (Willems et al. (2006)) found similar differences with hyperactives reporting, among other things, to be “... less satisfied with his or her income, less satisfied with how democracy is working, and less healthy.” The obvious implication is that the mix of hyperactives in a given sample may influence survey results.

**Table 1**

	Single Panel		Multiple Panels	
	HI	AMR	HI	AMR
<b>Female</b>	52.2%	61.9%	74.7%	71.3%
<b>18-39</b>	40.8%	41.7%	24.3%	49.0%
<b>Less than HS</b>	19.8%	19.4%	23.0%	22.9%
<b>Employed FT</b>	43.7%		35.6%	
<b>Less than \$25K</b>	20.3%		14.1%	

In our experience the levels of hyperactive respondents can vary substantially, both among panel companies and even within surveys using the same panel company. General population surveys of respondents with few qualifying questions often show the lowest levels of hyperactivity, while surveys targeting lower incidence or frequently surveyed respondents can be substantially higher. For example, in general population surveys we frequently see rates of multi-panel membership (based on self reports) of around 30 percent. By way of comparison, on surveys of physicians we often see rates of mulitpanel membership at 50 percent and higher, depending on specialty.

#### Inattentives

The second type of problematic panelist is the inattentive, that is, a respondent who gives less than full attention to reading the questions and providing well thought out answers. The phenomenon is not confined to online research and access panels. It was first identified by Krosnick (1991) who called it “satisficing.” Three main causes are typically cited: insufficient respondent motivation; reduced respondent cognitive ability; and poor questionnaire design. Satisficing in online research is most often characterized by behaviors such as short completion times, high levels of item nonresponse, non-differentiation or “straightlining” in matrix or grid questions, fewer selections in multiple response questions, and inconsistent or nonsensical responses. Recent studies by Lugtigheid and Rathod (2005) and Gilasic (2006) report that in online studies these behaviors become especially pronounced with surveys of 20 minutes or longer.

As discussed later in this paper, we frequently use a combination of “verification” items in grids and interview length to identify inattentives. Depending on the survey design, the combination of verification item failures and short completion times (e.g., 50 percent of the median interview length) will generally yield between two percent and 10 percent rates for inattentive respondents.

We are unaware of any studies that show whether levels of satisficing are worse with online than with other methodologies or whether online panelists are more likely to satisfice.

Nonetheless, there are a number of features of the online survey experience that would seem to create the necessary conditions for satisficing. These include a wealth of survey opportunities, universal use of incentives, self administration, increasing survey complexity, and, sadly, repetitive and boring survey design (the latter compounded by the fact that online surveys are frequently fielded on very short timelines and without adequate pretesting compared to other modes).

#### Fraudulents

Finally and perhaps most conspicuously in the minds of research buyers, are the “fraudulents,” sometimes called “gamers”. These are individuals who assume false identities or simply misrepresent their qualifications either at the time of panel registration or in the qualifying questions of individual surveys. Their primary motive is incentive maximization. They tend to be seasoned survey takers who can recognize filter questions and answer them in ways that they believe will increase their likelihood of qualifying for the survey; one classic behavior is selection of all options in multiple response qualifying questions; another is overstating purchase authority or span on control in a B2B survey.

In an earlier paper (Downes-Le Guin et al. (2006)) we described a number of first hand experiences with fraudulent panelists. For example, we described a study in which the target respondent was both home and business decision makers to represent potential purchasers of a new model of printer. The study was multinational with a mix of sample sources including a US customer list provided by the client, a US commercial panel, a European commercial panel, and an Asian phone sample that was recruited to do the survey online. One multiple response qualifying question asked about the ownership of ten home technology products. About 14 percent of the US panelists reporting owning all ten products, including the Segway Human Transporter, an expensive device known to have a very low incidence of ownership among consumers. This behavior was virtually nonexistent in the other sample sources.

A second example comes from a study of U.S. interventional cardiologists, a physician specialty that treats heart disease with non-invasive procedures such as catheterization. ICs are not trained nor licensed to perform open surgery procedures, yet in our survey when respondents were asked to indicate which of about 60 procedures they perform six percent claimed to be performing five or more open surgery procedures (including “extremity amputation”) and another six percent claimed to be performing two to four such procedures. Neither we nor the client believed these respondents were who they claimed to be, and so they were dropped from the study.

Fraudulence has arguably garnered disproportionate focus in the market research industry because, at first blush, it seems more obviously wrong and severe a behavior than panelist hyperactivity and inattentiveness. Left unmonitored, it can have a profound effect on both response error and credibility of the online panel as sample source. At the same time, gross examples of fraudulence are increasingly rare as public panel providers have converged on a set of best practices around identity verification and deduplication. While these practices are far from foolproof, it’s useful to remember that “curbstoning” and other forms of interviewer falsification have haunted both in-person and call center research from their beginnings. Hyperactive and inattentive panelists, on the other hand, are far harder to consistently define, detect and treat in a consistent fashion from survey to survey.

### 3. Separating the Wheat from the Chaff

We hope the foregoing discussion makes clear that conducting online research with access panels has multiple pitfalls. Avoiding them requires that researchers develop a comprehensive strategy that includes criteria for panel vendor selection, thoughtful questionnaire design, and careful cleaning of survey response data.

#### Panel Vendor Selection

Deming's (1986) Point 4 seems instructive here: "End the practice of awarding business on the basis of price tag alone." The online panel industry, at least in the U.S., has evolved into something approaching a commodity business. It is difficult to see how that leads to quality research. We favor an approach to vendor selection that starts with careful evaluation of panel recruitment, panelist registration, sample development, and panel management practices.

As in most evaluative exercises knowing the questions to ask is at least half the battle. ESOMAR has been especially helpful in this regard, having developed a set of questions known as "The ESOMAR 25" (2005) that focus on key elements of panel recruitment, panel management, and sample development. The International Standards Organization (ISO) has convened a working group from the same technical committee that developed the ISO 20252 Standard for Market, Opinion, and Social Research with the goal of providing a comprehensive framework for establishment and auditing of access panel standards. Other trade and professional organizations (e.g. the Market Research Association in the US) have developed or are in the process of developing similar guidelines. Of special interest should be questions about:

- The recruitment process and especially steps taken to ferret out multiple registrations.
- Panelist verification, that is, steps taken to verify that a panelist is who s/he says s/he is. This is especially important on B2B panels or panels of professionals such as physicians.
- Limits placed on participation.
- Policies for replacing "bad" respondents on individual surveys.

Panel companies are mostly aware of the ESOMAR 25 and many can easily provide their response to them. Some have even posted those responses on their Web sites. They also are aware of the panelist pathologies described above. They recognize the threat this poses to their business, and many have instituted special programs and procedures aimed at reducing multiple and/or fraudulent registrations. These procedures include scrubbing of email addresses against list of known hyperactives, monitoring of IP address frequency, verification against external sources, and registration and profiling surveys that contain some of the same techniques described below. Panel companies that have embraced these kinds of procedures enthusiastically are the only companies one should seriously consider.

#### Questionnaire Design

While panel vendors can do a great deal more than they currently are doing to ensure the legitimacy and suitability of the sample members they send us the responsibility for valid survey results ultimately rests with the researcher. This is the minimum that our clients expect from us.

As researchers, our interaction with panelists is pretty much confined to the survey experience, or more specifically, the survey questionnaire. The main focus of questionnaire design has rightfully been the substantive issues needed to answer the basic research questions the client has posed. But there are other important foci that are especially important for online research with access panel respondents.

The first is providing the respondent with as engaging an experience as we can. Not all inattentive behavior can be blamed on the panelist; bad questionnaire design can quickly wear down even the most motivated respondent. Classic questionnaire features that encourage inattentive behaviors include screen after screen of grid questions, continuing iterations of conjoint screens on which the changes from screen to screen are almost imperceptible, and questions that are so poorly worded as to not be taken seriously. Many online questionnaires seem to have evolved little beyond the webbification of CATI screens or mail surveys.

Luetters et al. (2007) have proposed the use of innovative answering devices they call “Sniperscales” that bring some video game techniques into surveys. They willingly admit that they are walking what they call “the narrow path between science and entertainment” but claim that the use of such devices reduces inattentive behavior. At the same time, the widespread assumption that respondents must be entertained and even visually stimulated seems to fly in the face of the best practices that have emerged around virtually every form of web interface design and human/computer interaction. If these disciplines serve as an effective analog for good online survey design, our goal should be to create user interfaces that are clean and uncluttered – focusing on succinct and unambiguous wording rather than gratuitous visual elements.

In the end probably the biggest design challenge we face is survey length. With sufficient incentive we usually can cajole the needed number of respondents to complete even the longest online survey, but the quality of the data may not be what it needs to be.

A second focus is detection of the bad respondent behaviour. In our organization we have come to refer to this as “setting traps” to identify respondents who are candidates for deletion from the survey dataset because they exhibit one or more of the panelist pathologies. Techniques include:

- Use of multiple response qualifying questions that include very low incidence behaviors (e.g., the earlier Segway example) or inconsistencies (i.e., you can’t be both a and b)
- Inclusion of consistency checks by asking essentially the same question but in two different ways in separate parts of the questionnaire.
- Inclusion of “verification items” in grid questions that span multiple screens. These might be questions that reverse direction in scales or simply give an instruction such as, “For verification purposes please click the item on the far right.”
- Collection of time stamps to measure how long a respondent took to answer individual questions or complete the entire survey.

#### Editing/Cleaning Response Data

The final element in a comprehensive panel quality assurance strategy is post data collection editing designed to identify problematic respondents who are candidates for deletion and possible replacement. This starts with checking the just-described traps as well as evidence of other problematic behaviors. In this context, the researcher should keep in mind that every

survey regardless of sample quality or quality of questionnaire design will have some level of error. The challenge is identify patterns that reflect bad respondent behavior rather than survey error. And so we advocate a scoring system across multiple items and then only delete those respondents who fail multiple items. Elsewhere (Downes-Le Guin (2005)) we have referred to this as “three strikes and you’re out.” Smith and Brown (2006) characterize this more eloquently and accurately as a “quality index” on which the researcher sets cut-off scores. High scoring respondents are deleted while low scoring respondents are retained.

Among the items that might make up such an index are these:

- Unusually short completion times as compared to the median interview length
- Selection of all items in a multiple response or other obvious cheating behavior in qualifying questions
- Selection of bogus or low probability answers
- Internal inconsistencies
- Low differentiation or “straightlining” in grids
- High levels of item nonresponse
- Failure of verification items in grids
- Gibberish or duplicated responses in open ends

The checks that might be included in such an index are limited only by the researcher’s imagination and the effort required to implement them. Mostly we have found that a modest amount of effort (e.g., four hours or less) in post data collection edit can identify the most serious problems. In many instances we perform our initial checks during data collection so that additional sample can be arranged and the survey timeline more or less preserved. We also have experimented with building these QA checks into the Web questionnaire program so that they are performed on a respondent-by-respondent basis in real-time as individual surveys are completed. A key to success of this type of index, however, is that it be specific to each study. The categories of indexed items may be essentially the same from study to study, but how they are treated and the weight given in the index should be sensitive to the specific panel being used, the subject matter of the study, practical limitations on how many panelists we can afford to eject from the final dataset, and so forth.

#### 4. Conclusion

The online panel industry and the methodology it makes possible are now about ten years old. Both have been enormously successful in a relatively brief span of time, but both now face a major challenge. Whether one calls it growing pains or a major crisis, it is clear that our approach to building and operating panels must change, and so must the ways in which researchers use them. Panel companies for the most part recognize that the quality issues are real, and most have begun to compete on quality as aggressively as they have in the past competed on size, coverage, and price. At the same time, definitions of “quality” vary widely and are inordinately focused on the most easily identified issues around fraudulence that can be managed by the panel providers. Those who specify the sample design and deal with the data – that is, full service research firms and their clients – must take increased responsibility for defining and addressing other aspects of quality. Industry organizations such as ESOMAR must consider whether their obligation should extend beyond offering question and guidelines to some form of auditing. The previously-mentioned ISO Working Group would seem to be moving the industry in that direction. That being said, the industry has awakened to the

problems, and hopefully have begun to recognize the key role they must play. Our clients ultimately rely on us as researchers to deliver valid results and recommendations. It may be gratifying for us to know that our online sample suppliers are taking quality seriously, but they will only meet that responsibility if we continue to insist on it.

One cannot help but wonder whether the very success of online panel research is not now the cause of its potential undoing. By one estimate, after deduplication, there are about 18 million people in the US who have signed up to do online research (Rivers (2006)). That amounts to about nine percent of the adult population. We know little about what portion of that 18 million participates in surveys on a regular basis, but it likely is a still smaller group. (By way of contrast, studies show that, on average, respondents to telephone studies in the US participate at a rate of no more than one or two studies per year.) At the same time the demand for online panelists continues to soar, creating the very conditions that lead to the pathologies we have described. In this context the QA techniques put forth in this paper might be little more than a holding action while we refine or perhaps even redefine the paradigm. The most obvious next step is finding ways to encourage a greater share of the population to join panels and to participate regularly. Less obvious is some alternate means for getting the respondents we need for our online surveys that not only increase the size of the respondent pool but also allow us to reduce the bias that has always been inherent in the volunteer panel design

## References

- Black, G. and Terhanian, G. (1998) "Using the Internet for Election Forecasting." *The Polling Report*. October 28.
- Cartwright, T. and Nancarrow, C. (2006) "The Effect of Conditioning When Re-interviewing a Pan-European Panel." Presented at ESOMAR World Research Conference: Panel Research 2006, Barcelona.
- Coen, T., Lorch, J. and Piekarski, L. (2005) "The Effects of Survey Frequency on Panelists' Responses." Presented at ESOMAR World Research Conference: Panel Research 2005, Budapest.
- Deming, W.E. (1986) *Out of the Crisis*. MIT Press.
- Downes-Le Guin, T. (2005) "Satisficing Behavior in Online Panelists: What's a Research Buyer to Do?" Presented at the Annual MRA Conference, Chicago, IL.
- Downes-Le Guin, T., Mechling, J., and Baker, R. (2006), "Great Results from Ambiguous Sources: Cleaning Internet Panel Data." Presented at ESOMAR World Research Conference: Panel Research 2006, Barcelona.
- ESOMAR (2005) *Conducting Market and Opinion Research Using the Internet*. [http://www.esomar.org/uploads/pdf/ESOMAR\\_Codes&Guideline-Conducting\\_research\\_using\\_Internet.pdf](http://www.esomar.org/uploads/pdf/ESOMAR_Codes&Guideline-Conducting_research_using_Internet.pdf).
- Fine, B., Menictas, C., and Casdas, D. (2006) "Attitudinal Differences: Comparing People Who Belong to Multiple Versus Single Panels." Presented at ESOMAR World Research Conference: Panel Research 2006, Barcelona.

- Gilasic, M. (2006) "The Effects of Questionnaire Length on Quality of Data in a Web Survey." Joint Program in Survey Methodology, University of Maryland.
- Inside Research* (2007) "Special Report: Online MR Continues Strong." January.
- Krosnick, J. (1991) "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology*, 5, pp. 213-236.
- Lugtigheid, A. and Rathod, S. (2005) *Questionnaire Length and Response Quality: Myth or Reality*. Survey Sampling International.
- Luetters, H., Westphal, D., and Heublein, F. (2007) "*SniperScale: Graphical Scaling in Sata Sollection and its effect on the response behaviour of participants in online studies.*" Presented at the General Online Conference, Leipzig.
- Rivers, D. (2007) "Are We Running Out of Respondents? Understanding Survey Participation." Presented at AMA Executive Insights Conference, Phoenix, AZ.
- Schonlau, M., Zapert, K., Simon, L., Sanstad, K., Marcus, S., Adams, J., Spranca, M., Kan, H., Turner, R., and Berry S. (2004) "A Comparison Between Responses from a Propensity-Weighted Web Survey and an Identical RDD Survey." *Social Science Computer Review*. 22: 128-138.
- Smith, R. and Brown, H. (2005) "Assessing the Quality of Data from Online Panels: Moving Forward with Confidence." Harris Interactive.
- Smith, R. and Brown, H. (2006) "Panel and Data Quality: Comparing Metrics and Assessing Claims." Presented at ESOMAR World Research Conference: Panel Research 2006, Barcelona.
- Taylor, H. (2007) "The Case for Publishing (Some) Online Polls." *The Polling Report*. January 15.
- Willems, P., Ossenbruggen, R., and Vonk, T. (2006) "The Effects of Panel Recruitment and Management on Research Results: A Study Across 19 Online Panels." Presented at ESOMAR World Research Conference: Panel Research 2006, Barcelona.

## About the Authors

Reg Baker is Chief Operating Office of Market Strategies International.

Theo Downes-Le Guin is a Senior Vice President of Market Strategies International.





# Increased fieldwork efforts, enhanced response rates, better estimates?

Ineke Stoop, Josine Verhagen, Erik van Ingen

## Abstract

Three studies of the effects of enhanced field efforts in face-to-face surveys show that these efforts can substantially reduce the noncontact rate and result in a considerable number of converted refusers. However, evidence also seems to indicate that the 'additional' respondents belong to groups that were over-represented anyway. As a consequence, the final sample may be even more imbalanced than in a survey with a lower response rate, and final nonrespondents may differ substantially from late contacts and converted refusers.

The paper describes these surveys, their designs and efforts to improve them. It focuses on the measures taken to enhance response rates and the results of these measures. Central are the analyses of the differences between easy and difficult respondents and possibly nonrespondents, and the question whether enhanced response rates reduce nonresponse bias.

## 1. Keywords

Survey design, enhancing response rates, nonresponse bias

## 2. Introduction

Nonresponse rates have been increasing in virtually all types of surveys in almost all countries. Nonresponse can have a negative impact on survey quality, firstly because of a smaller sample size and consequently smaller precision of survey estimates, secondly because the propensity to participate in surveys may differ across groups, and finally because of the possibility of nonresponse bias. This section will give a short overview of these detrimental effects, and of measures that can be taken to overcome them. The next section will describe efforts made by the Dutch Social and Cultural Planning Office (SCP) to improve fieldwork quality and enhance response rates. Section 3 will present and discuss the effect of these measures and the final section will focus on what to do to minimize nonresponse bias. The surveys discussed in this paper are face-to-face household studies.

A high nonresponse rate in a survey will mean a smaller achieved number of interviews than the original sample. This means a smaller precision of estimates. An obvious way to mend this would be to start with a larger sample. This has two drawbacks. Firstly, especially in face-to-face surveys, the costs of nonresponding units are high. Interviewers have to travel and make calls anyway, even if the target respondent cannot be reached or declines to participate. In mail surveys, the costs of nonresponding units will also be substantial, whereas the costs of nonresponse in telephone surveys are relatively small and in web surveys almost negligible. Secondly, a larger sample will not solve the problem of unequal participation pro-

pensities. Take for instance a survey on the interest for and use of technological innovations among 4,000 individuals. At a response rate of 50%, this will result in a final sample of 2,000 respondents. Now assume that the response rate of women is 75% and of men 25%. This implies that the effective sample size is not 2,000 but 1,500. This is caused by the fact that weights are required due to the unequal representation of men and women in the final sample, and that the efficiency of the weighted sample depends on the average of the squared weights. Based on these same 4,000 target respondents, if the response rates of women would be 90% and men 25%, the final sample size would be 2,300 and the effective sample size 1,565, and in the more extreme case of a response rate of 90% among women and 10% among men, the final sample size would be 2,000 and the effective sample size 720. This shows that increasing the sample to counter nonresponse and obtain a larger sample is very inefficient when response propensities are unequally distributed.

One way to solve this is to use booster samples, i.e. larger samples among subgroups that are expected to have lower response rates. A larger selection probability can thus compensate for a smaller response probability. This strategy is often used to mend lower response rates in urban areas. What if, however, in the example above sex has nothing to do with survey participation, but interest in technological gadgets has and – a well-known fact – women are more interested in gadgets. Recruiting additional men would lead to a higher number of male survey participants but will not help to minimize bias as non-tech men do not participate anyway.

A better strategy might be to aim for higher response rates. This strategy too might backfire. Consider the imaginary mail survey presented in table 1, measuring internet use per day. In the regular mail survey (a) the response rate was 45% and estimated internet time 2 hours. An internet follow-up of nonrespondents (b) would have increased response to 57% (estimated internet time 2.95 hours). Alternatively, a more expensive telephone follow-up (c) would have resulted in a response rate of 69% (estimated internet time 1.39 hours). A combination of both follow-ups (d) would have lifted the response rate to 78% (estimated internet time 2.15 hours). Neither follow-up would have decreased bias, however, as the initial mail survey was spot on the population estimate. Separately, the internet follow-up and the telephone follow-up substantially increase bias: the internet survey because it appeals to internet nerds only who never open their mail and are always on line, and the telephone follow-up because it succeeds in obtaining the cooperation of the digitally challenged who still have a landline phone, never use a computer and think that a mail survey on internet use was not meant for them.

Table 1 Response rates and bias (imaginary studies)

	share population	internet use population	mail survey (a)	internet follow- up (b)	telephone follow-up (c)	internet and telephone follow-up (d)	full sample (e)
	%	hours per day	response rates (%)				
ordinary people	60	2.00	75	80	80	80	100
internet nerds	10	8.00	0	90	0	90	100
digitally challenged	30	0.00	0	0	70	70	100
total response rate			45	57	69	78	100
estimate and bias (hours internet per day)							
estimate			2.00	2.95	1.39	2.15	2.00
bias			0	0.95	-0.61	0.15	

These somewhat artificial examples illustrate that the handling nonresponse is far from simple and that the relationship between survey estimates and response behaviour, and between response rates and nonresponse bias can be fairly complicated. In real life, these relationships are hardly less complicated, as shown convincingly by Groves (2006). Also in real life, many organisations try to enhance response rates, hoping that this will increase survey quality. Accidentally, this too is far from simple.

When trying to enhance response rates in a longitudinal survey two questions arise when these field efforts have succeeded. Firstly, were our efforts worthwhile when survey outcomes appear rather similar despite large differences in response rates across time? This may occur either when respondents and nonrespondents are very similar (and nonresponse is not a problem) or when the higher response rate may have been achieved by recruiting a select part of the nonrespondents who are very similar to the respondents (and the final results will still be biased). And secondly, when survey outcomes and response rates do differ across time, is the high response survey necessarily better than the low response one? Similar questions have extensively been studied over the last few years (Curtin et al., 2000; Keeter et al., 2000; Lynn et al., 2002; Merkle and Edelman, 2002; Teitler et al., 2003; Groves, 2006). The following sections will present three studies sponsored by the SCP in which extensive efforts were made to improve survey design and enhance response rates.

### 3. Fieldwork efforts and response rates at the SCP

The Dutch Social and Cultural Planning Office (SCP) conducts research on social and cultural aspects of government policy. Its research findings are mainly intended for the government, the Upper and Lower Houses of Parliament, senior ministry officials, and for professional and academic staff in the public sector and universities. SCP reports cover a wide range of issues: health care, social services, employment, social security, justice and criminal procedure, housing, education, leisure, the media, cultural affairs, attitudes towards government policies, distribution of income and wealth, public sector economics, the position of target groups and local policy issues. All publications are available on the SCP website ([www.scp.nl](http://www.scp.nl)).

The empirical data required for SCP research come from administrative files and public registers, but mainly from general social surveys (see also Stoop, 2002). Major data providers are Statistics Netherlands, government ministries, the Dutch data archive and international organizations. The SCP sponsors, in many cases jointly with other organisations, a number of surveys, among which the Amenities and Services Utilization Survey, the Cultural Changes in the Netherlands Survey and the Dutch Time Use Survey. It is also a partner in the European Social Survey (ESS).

#### Amenities and Social Services Utilization Survey (AVO)

In the Amenities and Services Utilization Survey (AVO in Dutch) the SCP collects information on the use of a great number of social and cultural amenities and social services in connection with household and individual characteristics. Among the topics covered are sporting activities, the use of sports and recreational facilities, visits to theatres, concert halls, cinema, exhibitions and museums, reading, buying books and library membership, media use, PC and internet use, leisure activities, hobbies, the use of educational and medical facilities and social assistance, membership of organisations and labour unions, social security benefits, income and income sources, pension scheme, the use of patient-care services, day-nursery facilities, health-care facilities for children, extent of infirmity, and social networks. It is a

burdensome survey, which was first carried out in 1979 and is repeated every four years. The sampling and field strategy used in the AVO is relatively straightforward. A gross sample of postal addresses is drawn. If available, the names and telephone numbers of residents are added. Shortly before the interview an advance letter and a brochure are sent to the residents. The interviewer visits the target household at home, interviews one person (or makes an appointment) and leaves extensive paper drop-off questionnaires for each member of the household aged six years or over. Children aged 6-15 are given an adapted questionnaire which may be filled in by their parents. Then the interviewer comes back to collect the questionnaires. In the case of multiple household addresses up to three families are interviewed. The net sample consists of 6,500 households, or 15,000 individuals (the number of individuals decreases each time as family size in the Netherlands decreases). The AVO is also used as a screening instrument to identify special groups (people with physical disabilities).

The response rate on the AVO was around 60% in the first three rounds (1979, 1983, 1987). In 1991 it reached an abysmal low of 43%. Extensive field efforts resulted in a response rate of 69% in 1995, mainly due to careful monitoring of interviewers and the fieldwork process, a large number of calls to first contact, an extensive strategy aimed at refusal conversion, and wherever possible anticipating the preferences of the respondents (Stoop, 2004). Response rates dropped a bit afterwards, to 66% in 1999 and to 60% in 2003. The last drop was largely due to more restricted definitions: whereas in previous rounds addresses where no contact could be established even after 8 calls were considered to be ineligible, these were now scored as "non-contact" except when independent information on ineligibility was available.

In addition to enhancing response rates, information on fieldwork progress was coded and keyed, based on detailed contact forms. As of 1999 the results of each individual call at each household was recorded. Also, subsequent to the 1999 fieldwork a follow-up survey among a small sample of persistent refusals was conducted. The response rate of this survey was 80%.

In future rounds the AVO survey will be fielded by Statistics Netherlands. This will allow linkage with public registers and administrative records.

### Cultural Changes in the Netherlands Survey (CV)

The Cultural Changes in the Netherlands Survey (CV in Dutch) dates from the seventies. It aims at approximately 2,000 net respondents. The survey has been repeated every year or two years since then. The resulting data are used by a wide range of academic researchers. The first questionnaire was composed of variables and scales from several earlier surveys. Questions cover preferences, values and opinions. Response rates used to be fairly high for the Netherlands, but may be questionable because of a combination of random walk and unregistered substitution. In the nineties there was a sharp decrease in response rates, possibly due in part to the length of the questionnaire (involving a face-to-face interview lasting almost two hours). As a result the questionnaire was abridged, a new sampling strategy was adopted and the response process was meticulously registered.

The sampling strategy of the CV-survey has varied through the years. The 2002 sampling strategy was fairly complicated. It started with a gross sample of postal addresses. Each province represented a separate stratum (proportionate stratification, according to size). Subsequently, with a probability proportionate to the number of inhabitants of cities or large neighbourhoods, approximately 800 starting addresses were chosen. Each starting address generated three clusters of addresses: five addresses on the left side, five on the right and five

across the street. Within each group of five addresses one interview had to be conducted. The interviewer called in-person at every address until a respondent agreed to cooperate immediately or to set a date and time for the interview. When a telephone number was available, the interviewer could also try to recruit a respondent by phone and make an appointment for an interview. If no interview had been achieved the first time, the interviewer came back a second or third time to call at addresses where nobody was at home on a previous occasion. In the final stage one person was selected at random from each household. In a small number of cases, according to a specific selection scheme, a second person was interviewed. It will be clear that this combination of random sampling of clusters, quota sampling at the last stage and the possibility of telephone recruitment is far from perfect. It makes it impossible to compute inclusion probabilities and response rates, and may have led to an overrepresentation of people who are immediately willing to cooperate, are often at home and those who have a landline phone and are listed in the telephone directory.

In 2004 the sampling procedure was thoroughly revised. Now, a random sample was drawn from postal addresses in two stages: within five levels of urbanization first postcodes were randomly selected and then respondents. Proportionate stratification according to size was used, although the two most urbanised groups were overrepresented to compensate for a lower expected response rate. In the end, the response in these postcodes was fairly high resulting in an overrepresentation of highly urbanized areas. Many efforts were made to reach the selected addresses, such as calling on different times of the day and different days of the week, re-issuing the remaining sample units and a higher incentive at the end of fieldwork.

Verhagen (2007) describes the efforts made at enhancing response rates, the achieved response rate (58%) and analyzes the obtained response in order to find out if differences in survey outcomes between 2002 and 2004 are due to cultural changes in the Netherlands or can be ascribed to changes in fieldwork procedures.

### Dutch Time USE Study (DTUS)

The DTUS is conducted once every five years as a joint venture by a number of different public and private organisations. The SCP has played a leading role since the first survey in 1975. The survey aims at approximately net 2,000 respondents who face a burdensome task as they have to answer a lengthy initial questionnaire, keep a diary in which they register what they are doing every 15 minutes in a fixed week in October, and finally answer another shorter questionnaire. The DTUS is based on a combination of proportionate stratification according to size and probability proportional to size sampling, based on postal address. Phone numbers, if available, were added to the addresses. Potential respondents were approached by telephone or by mail. One person was selected at random from the family members aged 12 and over and asked whether he or she might be willing to participate. To prevent drop-out in the diary phase the researchers emphasised the cumbersome nature of the task and the large amount of time required. Drop-out rates are substantial, however. In 2005, 688 respondents (out of 2,892, or 24%) completed the initial interview but failed to complete the diary. Response rates have been declining since the first round. In 1995 the final response rate was 18% (response here means initial interview, completed diary and final interview).

Before the 2000 data collection an experiment was conducted (NWO grant 400.47.100) to find out how response rates could be increased and whether more detailed information on time use could be obtained (for instance with whom time was spent). The response rate experiment included a) an advance letter from the sponsor or the survey agency, b) a 10 or a

20 Euro incentive, and c) a 2-day or a 7-day diary. Surprisingly, the largest (positive) effect was caused by the advance letter being sent by the sponsor. Extra fieldwork efforts in 2000 to enhance response rates resulted in a response of 21%. The increase compared with 1995 seems small. This is likely due to problems in the implementation of the survey.

In 2005 a different sampling frame was used. With the help of Statistics Netherlands an individual sampling frame could be used based on population registers. The sample was drawn in two stages, first postcode areas and subsequently individuals. This new design made it possible to send an advance letter to the designated respondent personally, rather than to the "residents at ...". This turned out to be fairly successful, as the response rate went up to 37%. Now the improved design was implemented satisfactorily. The recording of the fieldwork process (including the coding and keying of call records) was much more detailed in 2005 than in previous years.

In both 2000 and 2005 a small number of questions was asked to refusers to get some idea of nonresponse bias, and especially to find out if participation in the survey is related to time pressure on respondents (similar to the basic question approach from Kersten and Bethlehem (1984) and the PEDAKSI approach from Lynn (2003)). Van Ingen (2006) has analysed the impact of the extended field efforts in 2005 on nonresponse bias by analyzing the answers refusers gave to the basic questions and by comparing easy and hard to reach respondents.

In 2006 data for a DTUS-new design have been collected. This new design (NWO grant 480.04.008) is based on a host survey on mobility and collects time use data throughout the year.

#### 4. The impact of increased field efforts and nonresponse bias

The previous section has shown that response rates can be increased substantially between rounds (from 43% to 69% in the AVO, and from 21% to 37% in the DTUS) and that moving from a fuzzy survey design with clustering and quota aspects to a proper random sample can result in a quite acceptable 58% response rate. We can conclude that the design and the implementation of surveys sponsored by the SCP have improved substantially over time. What does this mean for survey results? Do survey outcomes differ? Are difficult groups better represented? And can we assume that nonresponse bias has decreased? We will try to answer these questions for each survey separately.

##### Amenities and Social Services Utilization Survey (AVO)

Nonresponse on the AVO survey has been analysed in great detail by Stoop (2004, 2005). Firstly the impact of increasing the response from 43% in 1991 to 69% in 1999. Although there were substantial differences in survey outcomes, these all but disappeared after weighting by post-stratification. The one exception was movie-going which was more common in the high response survey.

Based on the 1999 AVO, easy to reach and hard to reach respondents could be compared, cooperative respondents and initially reluctant respondents, and respondents and final refusers, i.e. those who cooperated in the follow-up survey. The results indicate, not surprisingly, that households in urban areas, persons who pursue activities outside the home and persons who are part of smaller households are more difficult to catch at home. Making evening calls helps to reach these households. Making additional calls to not yet contacted households helps to increase the contact rate and bring in these groups. Increasing the contact

rate improves data quality because being hard to reach cannot be totally corrected for by socio-demographic population characteristics (urbanicity, employment, family size). Those who are hard to reach also participate more in cultural activities, for instance.

In the AVO survey 37% of the target respondents cooperated without further ado. An additional 29% required an appointment, was initially not able to participate or – the majority – refused initially. This means that a large number of respondents have been converted. These ‘soft’ refusals are sometimes treated as a proxy for final refusals. Analyses show that among these ‘soft’ refusals were relatively few men and few singles. Rather than concluding that men and singles are rather cooperative, it turned out that they did cooperate less at the first request for participation and were also less susceptible to refusal conversion attempts by the interviewer. The same is true for the higher educated. This shows that the ‘soft’ refusals are not similar to the final refusals. Combined with the fact that – after correcting for socio-demographics – there were very few differences between cooperative and reluctant respondents, one might conclude that refusal conversion mainly managed to bring in a lot of respondents who were very similar to the ones who were overrepresented anyway.

What about the final nonrespondents who cooperated in the follow-up study? The differences with the regular respondents were relatively small, as can be seen in table 2 which summarizes those variables that showed significant differences. However, they more often partook in popular culture (movies, pop concerts, clubbing), did fewer sports, less often had a PC – but when they had one they used it more for games and chatting on the internet – and less often had a religious affiliation. Although the differences were small and limited to a small number of variables, it appears that in this study enhancing the response rate did not manage to recruit these typical refusals, and thus did not reduce nonresponse bias.

**Table 2. AVO response groups: easy respondents, hard to contact, reluctant and final refusals (in follow-up survey), multinomial regression (N=6360)**

	B			X <sup>2</sup> (df=3)	significance level			total	wald		
	refuser	reluctant	hard to contact		refuser	reluctant	hard to contact		refuser	reluctant	hard to contact
intercept	-2.42	-0.04	-0.38		0.000	0.854	0.088		28.90	0.03	2.92
age (years)	-0.01	0.00	-0.01	9	0.205	0.526	0.004	0.026	1.61	0.40	8.46
age: absolute deviation mean	-0.01	-0.01	-0.01	11	0.313	0.003	0.042	0.014	1.02	8.97	4.15
listed phone number	0.50	-0.10	-0.26	20	0.012	0.222	0.001	0.000	6.35	1.49	10.16
Amsterdam, Rotterdam, The Hague	-0.22	0.69	0.63	88	0.338	0.000	0.000	0.000	0.92	58.66	51.14
good maintenance	-0.21	-0.26	-0.26	18	0.204	0.001	0.001	0.001	1.61	11.22	10.40
detached dwelling	0.00	-0.42	-0.28	21	0.993	0.000	0.010	0.000	0.00	16.58	6.70
male	-0.16	-0.76	-0.05	112	0.264	0.000	0.455	0.000	1.25	101.33	0.56
single	0.14	-0.37	0.29	41	0.428	0.000	0.001	0.000	0.63	16.16	11.90
child at home	-0.30	-0.11	-0.37	19	0.097	0.188	0.000	0.000	2.76	1.73	17.85
employed, student	0.17	-0.13	0.31	22	0.355	0.100	0.000	0.000	0.85	2.70	12.52
higher education	-0.15	-0.51	-0.03	29	0.448	0.000	0.729	0.000	0.58	26.04	0.12
cultural participation (df=9)				27				0.001			
both classical and popular	-0.21	-0.03	0.13		0.312	0.751	0.185		1.02	0.10	1.76
classical	-0.56	0.23	0.17		0.010	0.012	0.088		6.63	6.38	2.91
popular	0.17	-0.07	0.15		0.432	0.493	0.173		0.62	0.47	1.86
no sports activities	0.41	-0.03	0.00	8	0.007	0.681	0.999	0.050	7.19	0.17	0.00
pc in household	-0.87	0.18	0.00	26	0.000	0.037	0.972	0.000	17.60	4.36	0.00



	B		X <sup>2</sup>		significance level				wald		
use of internet, e-mail	0.94	0.07	0.14	21	0.000	0.469	0.133	0.000	19.23	0.53	2.26
no religious affiliation	0.38	-0.11	0.00	11	0.008	0.117	0.960	0.013	7.01	2.46	0.00

Contrast with easy-easy respondent; Chi-square 693 (df=54); Cox and Snell R<sup>2</sup> 0.103 Nagelkerke R<sup>2</sup> 0.115.

Reference categories ( $\beta=0$ ) Other regions, no cultural participation, not employed, student

### Cultural Changes in the Netherlands Survey (CV)

The impact of enhanced field efforts has been analysed by Verhagen (2007). The assumptions in this study were that the 2004 results would differ from the 2002 results because of the improved design and the extra efforts to include hard to reach and initially reluctant respondents. Without the additional 'difficult' respondents, the 2002 and 2004 results were assumed to be more similar.

Before comparing 2002 and 2004, contactability and cooperation in the 2004 study were investigated. As in the AVO survey, converted refusals were rather similar to those who cooperated immediately. It should be noted here that both the sample size and the percentage of respondents who had to be converted were much smaller than in the AVO. Hard to reach were people in paid labour, persons in small families, young people and the higher educated; easy to reach were especially the elderly. With respect to survey topics, differences between the easy and hard to reach were small, especially after correcting for socio-demographics (see table 3). The hard to reach were a bit more satisfied with government, but this would not have led to different survey conclusions. Certain hard to reach groups were thus better represented after the extended efforts, but this did not lead to different results. This indicates that these efforts had no effect on bias.

**Table 3. Cultural Changes response groups: easy respondents, hard to contact and reluctant; multinomial regression (betas) (N=2048)**

	refusers	hard-to-reach	both
Intercept	-1.29	0.55	1.75
Dissatisfaction with government	-0.11	-0.23	-0.49
No trust in other people	-0.38	0.02	-0.10
Disagree with: There should be less rules and more strong leaders	0.04	0.03	-0.18
Disagree with: Criminals have to be changed instead of punished	-0.02	0.06	0.05
Disagree with death penalty	-0.04	0.05	-0.07
Disagree with abortion	-0.05	-0.04	-0.04
Negative opinion about ethnic minorities	0.18	-0.07	-0.22
Work	0.04	-0.18	-0.74
Age	0.01	-0.01	-0.01
Years of education	-0.02	0.02	-0.02
Single	-0.18	0.32	0.33
Apartment dwellers	0.35	0.40	0.84

Contrast with easy-easy respondent; -2 Log Likelihood = 4335.20; Chi-Square (36)=121.67, p=.000; Cox and Snell=0.0, Nagelkerke=0.07

Now what about the difference between the less strict survey design in 2002 and the stricter design in 2004? Verhagen (2007) concludes that the 2004 sample is more similar to the population than the 2002 sample with respect to age and marital status. It seems, however, that the extra efforts in 2004 resulted in too many higher educated persons, which might be an indication of increased efforts leading to overrepresentation of certain (already well-represented) groups.

A comparison between the less strict 2002 sample, the easy respondents in 2004 and the total 2004 sample should answer the question whether the extra efforts in 2004 managed to bring

in difficult respondents, a group less likely to participate in the 2002 survey group that was not approached again in 2002. Although the differences were small, results indicate a pattern contrary to the expectations: the 2002 (not strict) sample and the 2004 total sample were more similar than the 2002 (not strict) sample and the easy respondents in 2004. This means that the processes behind recruiting respondents in 2002 (via clusters and quotas and favouring in practice those who are easy to reach (by telephone)) cannot be equated to the process behind recruiting easy respondents in 2004, and that the extra 'difficult' respondents in 2004 are not the ones who were not reached or who refused in 2002.

We can conclude that the design and implementation of the survey was better in 2004 – if only because now the response rate could be calculated. There is no indication, however, that this resulted in a smaller nonresponse bias.

### Dutch Time USE Study (DTUS)

Time use studies are burdensome surveys. Usually, respondents have to keep detailed diaries for 2 to 7 days, and have to complete these for a designated, fixed period. Despite the low response rates this may cause, diaries provide better estimates of time use than directly asking for an estimate of time use in a regular, less burdensome survey (Kan and Pudney, 2007). One of the recurrent worries of time use researchers is that – especially because of the heavy response burden – respondents with demanding time schedules may be less inclined to partake in the survey, resulting in nonresponse bias. Recent overviews from Stoop (2005) and Abraham et al. (2006) indicate that it is not likely that busy people are underrepresented in TUSes but there are indications that time use on voluntary work is overestimated.

Van Ingen (2006) has analysed the differences between the DTUS 2000 and 2005, and the differences between respondents and those refusers who were willing to answer a small number of basic questions (almost one third). He ascribes the increase in response rates between 2000 (21%) and 2005 (37%) mainly to the impact of an increased number of contact attempts (max 4 in 2000, up to 10 in 2005). The effect of the personalized advance letter is of course difficult to measure. Van Ingen found that through the additional contact attempts relatively many people were reached with paid jobs and a higher education, whereas the proportion of elderly decreased considerably. This makes sense as the former spend less time at home and the latter more. This is not what we would have desired, however, as employed, younger and higher educated people were overrepresented from the beginning (and the elderly underrepresented). Differences in survey outcomes between easy and hard to reach respondents seem to be related to their background variables: as hard to reach respondents are more often in paid work, they spend more time on commuting. In this case too, one may wonder about the efficiency of enhancing response rates as this results in a more unbalanced achieved sample.

Table 4 gives the summary of a comparison of respondents and refusers (those 31% who completed a short questionnaire). The respondents more often read newspapers, watch television and are interested in politics, play sports more frequently, travel more during the week and are engaged in volunteer work more often. Model 1 includes a number of DTUS core variables, model 2 also includes background variables. Significant variables in model 1 generally remain significant in model 2. Contrary to expectations feeling stresses does not distinguish between respondents and refusals. Reading newspapers becomes even more significant when background variables are included. This is most likely due to the effect of

age: the elderly more often read newspapers and less often participate in the DTUS. After controlling for age the effect of age increases.

In sum, these factors add up to a sample with a considerable overrepresentation of active and engaged people. These effects remain significant after correcting for background variables. There is no difference between respondents and refusers in feeling agitated or hurried. Or, both in terms of factual (by demands of work, travelling, household chores, etc.) and experienced time pressure, the final sample showed no underrepresentation of busy people. This confirms earlier findings that busyness is not a cause of nonparticipation in time use surveys.

**Table 4. DTUS response groups: complete respondents and those who answered the basic questionnaire (reference group); logistic regression**

	Model 1	Model 2	Exp(B2)	Std. Dev	Exp(B2*std)
Read newspaper	0.055	0.243*	1.285	0.492	1.131
Practiced sports	0.635*	0.442*	1.575	0.493	1.251
Watching television	0.071*	0.085*	1.083	1.737	1.149
Political interest	0.401*	0.463*	1.563	0.655	1.339
Time spent travelling	0.184*	0.117*	1.124	1.636	1.210
Volunteer work	0.540*	0.637*	2.024	0.447	1.371
Feeling rushed	-0.029	0.027			
Population density		0.143*	1.154	1.392	1.221
Sex (female)		-0.076			
Age		-0.046*	0.955	18.288	0.433
Age <sup>2</sup>		0.000			
Children living at home		-0.074			
Employed		-0.094			
Partner employed		0.128			
Higher education		0.283*	1.327	0.432	1.130
Nagelkerke R <sup>2</sup>	0.102	0.155			

## 5. How to combat nonresponse bias

The results of the three studies presented above are both reassuring and disconcerting. Reassuring for methodologists because strict survey design and implementation is possible and response rates can be enhanced substantially. We do not have information on the exact costs of these improvements, partly because so many things change when a survey is conducted every 4 or 5 years that straightforward price comparisons are difficult. Suffice it to say that in a proposal for the 1995 AVO fieldwork the 60% response variant was less expensive than the 70% response rate variant (because the former comprised a larger sample to realise the same number of interviews and more unsuccessful target respondents that still required a substantial investment in the form of in-person house calls). In later rounds and in other surveys, however, a higher response rate always comes at a cost.

The second reassuring outcome is that substantive researchers can make their longitudinal analyses as changes in design and higher response rate do seem to have a limited effect only on survey outcomes. Substantive researchers do not like changes in survey design nor higher response rates nor anything else that can disturb their trend. As we have seen they can rest assured.

These results are disconcerting because the changes in survey outcomes – after correcting for background variables – are so small. Why then bother and spend all this money? Of major concern is the finding that extra field efforts will result in higher response rates, but these additional respondents are highly similar to the ones that were already well-represented or even over-represented. Moreover, according to the AVO follow-up study, they are not similar to the final refusals. Enhancing response rates may not minimize bias.

The present results suggest that – rather than spending funds on increased fieldwork efforts to indiscriminately haul in additional respondents – it might be preferable to focus on target respondents who are unlikely to participate, and to collect additional information on final nonrespondents. The first strategy means that field efforts will be distributed unevenly among sample persons, and that larger efforts will have to be made to recruit groups that are less likely to participate. Ideally this strategy would be adapted throughout the fieldwork process (Groves and Heeringa, 2006). The second strategy implies collecting auxiliary data from rich sample frames, interviewer observations, call records and reasons for refusal. In addition, high response basic question approaches or follow-up surveys among refusals can show us what we miss or see distorted. High response rates are a **means toward** data quality. Low nonresponse bias is an **indicator** of data quality.

## References

- Abraham, Katharine G., Aaron Maitland, and Suzanne M. Bianchi (2006) Nonresponse in the American Time Use Survey: Who Is Missing from the Data and How Much Does It Matter? *Public Opinion Quarterly*, Vol. 70, pp. 676-703.
- Curtin, Richard, Stanley Presser and Eleanor Singer (2000) The Effects of Response Rate Changes on the Index of Consumer Sentiment. *Public Opinion Quarterly*, Vol. 64, pp. 413-428.
- Groves, Robert M. (2006) Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*, Vol. 70, pp. 646-675.
- Groves, Robert M., Steven Heeringa (2006) Responsive design for household surveys: tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 169, Part 3, pp. 439-457.
- Kan, Man Yee, and Stephen Pudney (2007) *Measurement Error in Stylised and Diary Data on Time Use*, ISER Working Paper 2007-03. Colchester: University of Essex.  
[www.iser.essex.ac.uk/pubs/workpaps/pdf/2007-03.pdf](http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2007-03.pdf)
- Keeter, Scott, Carolyn Miller, Andrew Kohut, Robert M. Groves and Stanley Presser (2000) Consequences of Reducing Nonresponse in a National Telephone Survey. *Public Opinion Quarterly*, Vol. 64, pp. 125-148.
- Kersten, H.M.P., and J.G. Bethlehem (1984) Exploring and reducing the nonresponse bias by asking the basic question. *Statistical Journal of the U.N. Economic Commission for Europe*, 2, pp. 369-380.
- Lynn, Peter (2003) PEDAKSI: Methodology for Collecting Data about Survey Non-Respondents. *Quality and Quantity*, Vol. 37, pp. 239-261.

- Lynn, Peter, Paul Clarke, Jean Martin and Patrick Sturgis (2002) The Effects of Extended Interviewer Efforts on Nonresponse Bias. In: Robert M. Groves, Don A. Dillman, John L. Eltinge and Roderick J.A. Little (eds.) *Survey Nonresponse*. New York: Wiley, pp. 135-148.
- Merkle, Daniel A., and Murray Edelman (2002) Nonresponse in Exit Polls: A Comprehensive Analysis. In: Robert M. Groves, Don A. Dillman, John L. Eltinge and Roderick J.A. Little (eds.) *Survey Nonresponse*. New York: Wiley, pp. 243-258.
- Stoop, Ineke (2000) Covering the Population, Sample Surveys at the Social and Cultural Planning Office. *Kwantitatieve Methoden*, jr. 21, nr. 65, pp. 69-84.  
[www.scp.nl/onderzoek/km\\_stoop.pdf](http://www.scp.nl/onderzoek/km_stoop.pdf)
- Stoop, Ineke A.L. (2004) Surveying Nonrespondents. *Field Methods*, Vol. 16, nr. 1, pp. 23-54.
- Stoop, Ineke A.L. (2005) *The Hunt for the Last Respondent*. The Hague: SCP.
- Teitler, Julien O., Nancy E. Reichman and Susan Sprachman (2003) Costs and Benefits of Improving Response Rates for a Hard-to-reach Population. *Public Opinion Quarterly*, Vol. 67, pp. 126-138.
- Van Ingen, Erik (2006) *Nonrespons in het tijdsbestedingsonderzoek van 2005*. The Hague: SCP.
- Verhagen, Josine (2007) *Robuuste meningen? Het effect van responsverhogende strategieën bij het onderzoek Culturele Veranderingen in Nederland*. The Hague: SCP.

## About the Authors

**Ineke Stoop** is head of the department of ICT and Data Services at the Netherlands' Social and Cultural Planning Office (SCP). She is responsible for the surveys sponsored by the SCP and research on data quality. She has specialised in nonresponse issues. She is also a member of the Central Coordinating Team of the European Social Survey. She can be contacted at SCP, PO Box 16164, 2500 BD, The Hague, The Netherlands, tel. +31 70 3407671, fax +31 70 3407044, e-mail [i.stoop@scp.nl](mailto:i.stoop@scp.nl).

**Josine Verhagen** is a researcher at the department of ICT and Data Services at the SCP, and a master student at Leiden University. She works as a statistical consultant and studies nonresponse issues. Tel. +31 70 3407677, e-mail [j.verhagen@scp.nl](mailto:j.verhagen@scp.nl).

**Erik van Ingen** is a researcher at the department of Time Use, Media and Culture at the SCP. He is also a PhD student at Tilburg University and studies involvement in voluntary associations. Tel. +31 70 3407157, e-mail [e.van.ingen@scp.nl](mailto:e.van.ingen@scp.nl).

# Problems with surveys among ethnic minorities in the Netherlands

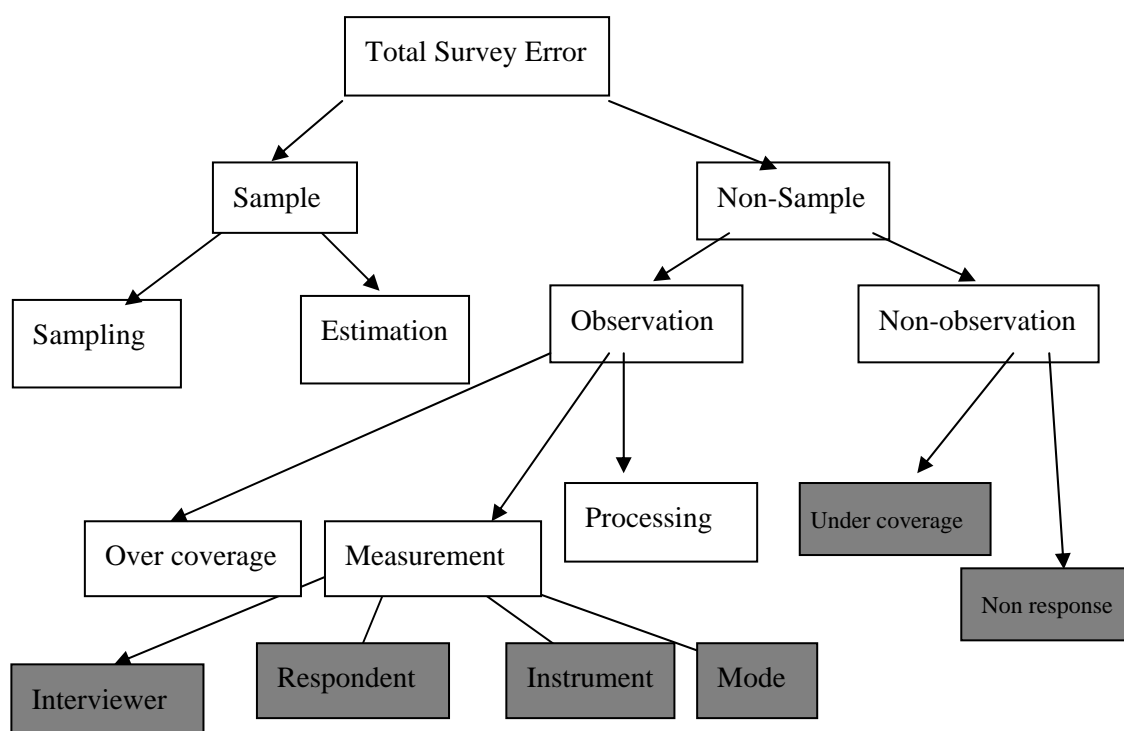
JWS Kappelhof

## 1. Introduction

In Dutch research the most used definition of a non-native is “a person who is resident in the Netherlands and of whom at least one of the parents is born in a foreign country”(CBS, 2001). According to this definition almost 3.2 million of the Dutch inhabitants are considered non-natives, which makes up 19% of the total population in the Netherlands. This group of non-natives consists of approximately 45% western and 55% non-western non-natives (CBS-statline, 2007). People that came from Turkey, Africa, Latin America and Asia (with the exception of Japan and Indonesia) are considered non-western non-natives. Conducting large-scale survey among non-western non-natives is considered difficult in the Netherlands (Kemper, 1998; Martens, 1999; Reep, 2003, Schmeets, Reep & Snijkers, 2003, 2005; Schothorst, 1999, 2002; Stoop, 2005; Van 't Land, 2002). They are seen as difficult to (observe and measure) survey due to cultural differences, language problems, (social)-demographic characteristics and a higher mobility (CBS, 2005; Dagevos & Schellingerhout, 2003; Groeneveld & Weijers-Martens, 2003; Kemper, 1998; Reep, 2003; Van 't Land, 2000). For instance, non-western non-natives tend to be city dwellers (CBS, 2005; Stoop, 2005; Schmeets, 2005) and it is harder to make contact with city dwellers and city dwellers are also less inclined to cooperate with surveys (Couper & Groves, 1996; Groves & Couper, 1998; Stoop, 2005; CBS, 2005). Furthermore, they tend to move more often thereby making them harder to reach (CBS, 2005). The response rates among non-western non-natives in large-scale surveys have also been dropping and this increases the possibility of selective response (CBS, 2003; Stoop, 2005; Voogt, 2004). It becomes clear from aforementioned that there are many possible reasons for distortion of the survey measurements among the non-western non-natives population. In recent years the Social and Cultural Planning office (SCP) conducted two large-scale surveys among non-western non-natives population. A large-scale survey was conducted about the living conditions of non-western non-natives city dwellers (LAS) in 2004 and a large-scale survey about the integration of ethnic minorities (SIM) in 2006. In the design and the implementation phase of both surveys the SCP used much of the earlier research findings (Meloan en Veenman, 2002; CBS, 2005; Van 't land, 2000; Kemper, 1998) that focused on causes for measurement error in surveying non-western non-natives. These difficulties with survey research among non-western non-natives are described using a form of the total survey error model (table 1) whereby the areas that need specific attention are tinted grey. This article describes how the SCP dealt with the difficulties and the trade-offs that they are accompanied by while conducting these surveys. The lay out of the article is as follows. First there will be a brief description of the history of survey research among non-western non-natives in the Netherlands and the choices and problems with defining the target population. The second part will give a brief description of the problems associated with the sample-side while conducting survey research among non-western non-natives in the

Netherlands. The next section deals with the non-sample side and the final part discusses the role of the organisation of the fieldwork and how it effects the quality measurement. Conclusions will be drawn in section 5.

**Table 1: Total Survey Error**



### 1.1 Background

The Dutch government has a minority policy mainly aimed at non-western non-natives because there is societal backlog (Reep, 2003). For this policy the research focuses mainly on four non-western non-native groups, which are Antilleans<sup>1</sup>, Surinamese, Moroccans and Turks. The reason for aiming the research for the minority policy on those groups is both their lower social-economic position in general (CBS, 2002; Dagevos & Schellingerhout, 2003, Stoop, 2005) and the fact that these four groups make up 2/3 of the total non-western non-native population in the Netherlands (table 2). In 2003 a survey was conducted among ‘new’ ethnic minorities, but the focus of the Dutch government is aimed at the four largest minorities.

**Table 2: Overview of the largest four non-western non-native groups in the Netherlands.**

	Total	Moroccans	Turks	Antilleans	Surinamese	Other
N	1739365	329634	368718	129590	333478	578215
%	100	19	21,2	7,4	19,2	33,2

cbs statline 2007

<sup>1</sup> Including Aruba

In the Netherlands government and other institutes have conducted a number of large-scale surveys specifically aimed at the non-western non-native groups in the last 3 decades. See table 3 for a short overview.

**Table 3 Short overview of survey research done by (government) institutes among non-western non-native in the last 3 decades**

<i>Institute(s)</i>	<i>Target population</i>	<i>Survey topic</i>	<i>Year</i>
Statistics Netherlands	Turks, Moroccans	Living conditions	1984
Statistics Netherlands	Antilleans, Surinamese	Living conditions	1985
Institute for Social-Economic Research (ISEO)	Moroccans, Antilleans, Surinamese, Turks	Living conditions (SPVA)	1988,1991, 1994
Statistics Netherlands	Turks	Health	1989
Institute for Social-Economic Research (ISEO) Social Cultural Planning Office (SCP)	Moroccans, Antilleans, Surinamese, Turks	Living conditions (SPVA)	1998, 2002
Veldkamp TNS NIPO	Moroccans, Antilleans, Surinamese, Turks	General	1999,2000
Statistics Netherlands Veldkamp	Moroccans, Antilleans, Surinamese, Turks	Budget	2000
Social Cultural Planning Office (SCP)	Moroccans, Antilleans, Surinamese, Turks	Living conditions (GWAO)	2003
Social Cultural Planning Office (SCP)	Somalis, Iranians, Iraqi, Afghans, Yugoslavs	Living conditions (SPVN)	2003
Netherlands Interdisciplinary Demographic Institute (NIDI)	Moroccans, Antilleans, Surinamese, Turks	Kinship Panel Study (KNPS)	2003
National Institute for Expenditure Information (NIBUD)	Moroccans, Antilleans, Surinamese, Turks	Expenditure	2003
Statistics Netherlands	Turks, Moroccans	Family planning (OGJA)	2004
The Netherlands Institute for Healthcare Research (NIVEL)	Moroccans, Antilleans, Surinamese, Turks	Health	2004
Social Cultural Planning Office (SCP)	Moroccans, Antilleans, Surinamese, Turks	Living conditions (LAS)	2004



<i>Institute(s)</i>	<i>Target population</i>	<i>Survey topic</i>	<i>Year</i>
Social Cultural Planning Office (SCP)	Moroccans, Antilleans, Surinamese, Turks	Integration (SIM)	2006

Besides the large-scale surveys specifically aimed at the ethnic minorities in the Netherlands there obviously have been many other surveys conducted in the Netherlands by the same and other institutes where non-western non-natives were part of the sample and also data regarding these minorities was published. Also from these surveys much knowledge has been gathered regarding the methodology with research among non-western non-natives.

## 1.2 Who is what? Differences and problems with defining, marking out and the coverage level of the non-western non-native target population.

### Defining and marking out

Marking out a population that is difficult to observe is tricky. First there has to be a clear definition of the target population whereby for each potential element it can be unquestionably determined whether it belongs to the target population (CBS, 2004). That definition should both be defensible from a theoretical viewpoint and measurable to come to a careful operationalisation (CBS, 2005). Except an unequivocal definition of who belongs to the target population, the timeframe should also be clearly established (Groves, 1989; CBS, 2004).

Nationality of a person has lost its use for large-scale surveys among non-western non-natives since many members of the target population have Dutch or duo citizenships. Also the overlap between 'non-western non-native' and 'person of an ethnic minority' is not complete. The term 'non-western non-native' implies that a person is originating from elsewhere and living in the Netherlands, but the term 'person of an ethnic minority' also implies that the minority on average has a lower social-economic position in the Dutch society and larger cultural differences (CBS, 2002).

In the remainder of this article the precise definition of what constitutes a non-native or member of an ethnic minority is left out of consideration. The focus is on the research methodology used for the target populations that are the focal point of the Dutch government's minority policy; Moroccans, Antilleans, Surinamese and Turks.

Changes due to time have also forced the most common definition of non-native, "a person who is resident in the Netherlands and of whom at least one of the parents is born in a foreign country (CBS, 2001)", to be a more operational definition that decides to which generation persons are considered non-native and to which ethnic minority a person belongs when they are born out of interethnic parents (CBS, 2002).

In the Netherlands a distinction is made between first and second-generation ethnic minority. First generation ethnic minority are persons of an ethnic minority where both they and also at least one of their parents has been born abroad (e.g. Morocco). Second-generation ethnic minority are persons born in the Netherlands, but at least one of the parents is born abroad. This excludes for instance ambassadors' children but includes children of immigrants. The country of birth of a person is decisive in determining whether someone belongs to a western or a non-western first generation ethnic minority and for determining a second-generation

ethnic member the mother's country of birth is decisive. In case the mother is born in the Netherlands, the father's country of birth is decisive in determining the ethnic minority (De Beer, 2001). Although sometimes for pragmatic reasons the distinction is not completely followed (Van 't Land, 2000) in general the consensus on the definition of what constitutes a non-western non-native, a member of an ethnic minority and a first and second generation ethnic minority in the Netherlands is high.

## 2. Coverage level

Since the start of conducting surveys among Moroccans, Antilleans, Surinamese and Turks in the Netherlands many of the researched populations have varied greatly in coverage level. For the subsequent living condition surveys (SPVA) conducted by ISEO and SCP a household sample was drawn from a limited number of municipalities (10-13) were contacted. These surveys varied greatly in the coverage level for each of the total target populations<sup>2</sup> (Groeneveld & Weijers-Martens, 2003; Zeijleman, Braat & Martens, 1989). The argument behind the choice for these few municipalities<sup>3</sup> was the desire to have a good geographical distribution, but also to have enough potential respondents living in those municipalities for sampling purposes (Zeijleman, Braat & Martens, 1989). The family planning survey (OGJA) conducted by Statistics Netherlands was done among a sample of Turks and Moroccans in the age of 18 till 27 living in the Netherlands<sup>4</sup> (CBS, 2005). The LAS2004 Survey was conducted among a sample of the urban population of Turks, Antilleans, Moroccans, Surinamese and Dutch<sup>5</sup> (age 15-65) of the 50 largest cities in the Netherlands<sup>6</sup> (Schothorst, 2005). The SIM2006-survey was conducted among a sample of the Turks, Antilleans, Moroccans, Surinamese and Dutch population (15+) living in the Netherlands. The differences between the researched populations have consequences for the possibility to generalize to the whole target population of interest. The possibility to generalize to the whole target population of interest can only be done by a clearly defined target population (on age, region, ethnicity, etc) where all the potential elements have a chance of inclusion in the sample. Due to the initial set up<sup>7</sup> of the SPVA-surveys it was recognized that about certain topics of interest there were no clear, undistorted data from partners and children in the households (Zeijleman, Braat & Martens, 1989).

When potential elements of the target population of interest have no chance of being in the sample coverage level is important, but not all. The preferred choice in the trade-off here would be a clear possibility to generalize to a *smaller* target population where the target

<sup>2</sup> The level of coverage of the total target populations for these municipalities in 1988 varied from 35% of the Antilleans, 44% of the Turks, 52% of the Moroccans to 63% of the Surinamese. In 2002 the level of coverage of the total target populations for these municipalities 44% of the Antilleans, 46% of the Turks, 54% of the Moroccans and 64% of the Surinamese. Except the 5% increase among Antilleans the coverage level in 2002 was the same as in 1988.

<sup>3</sup> For reasons of comparability the same design was used for determining the targetpopulation in the subsequent SPVA-surveys. The only variation was the choice of municipalities, they varied from survey to survey (except for Amsterdam, Rotterdam, Utrecht, Den Haag, Eindhoven and Enschede) based on coverage level of the targetpopulations. In 1988 the other municipalities were Groningen, Apeldoorn, Gorinchem and Veghel. In the 1998, 2002 and 2003 editions of the SPVA-survey the other municipalities were Almere, Alphen a/d rijn, Bergen op Zoom, Hoogezand-Sappemeer, Delft, Dordrecht and Tiel.

<sup>4</sup> Not including residents of institutions.

<sup>5</sup> Dutch could be anything but Turks, Antilleans, Morrocans and Surinamese.

<sup>6</sup> According to the Nyfer-classification (Marlet & Woerkens, 2003).

<sup>7</sup> The male and also the Surinamese and Antilleans female head of households (18+) were interviewed first as it was thought that the way into a household for an interviewer was to interview the head of the household first and afterwards the rest of the householdmembers.

population is unequivocally defined and each potential element of that target population has a known chance larger than 0 to be included in the sample.

### 3. Sample

This section deals with sample errors. Sample errors arise because not the whole target population is surveyed, but only a sample (Bethlehem, 2004). Surveys conducted among ethnic minorities could suffer from sample errors because of sampling errors and estimation errors. Sampling errors spring from incomplete sampling frames and errors in the sampling procedure when the actual and the anticipated probability of inclusion in the sample for a potential element differ. Estimation errors happen with all sorts of research and surveys among ethnic minorities don't seem to be more susceptible to it. Steps to limit the estimation error such as sufficient sample size and a correct sample design should be done carefully. For both the LAS and the SIM survey sample size calculations were preformed to limit the estimation error and have sufficient sample size for different analysis. The sample design for the LAS-survey was a 2-stage sample design for Turks, Moroccans, Antilleans, Surinamese plus a Dutch control group living in the 50 largest cities. For each separate minority group first municipality were randomly drawn and in the second stage respondents were randomly drawn from the sampling frame (the selected municipal personal records database of the 50 largest cities in The Netherlands). The method used in the SIM2006 study was a stratified 2-stage sampling method with proportional to size allocation done separately for each group. The Netherlands was divided into three strata (regions). Region1 were the cities with population >200.000, region2 50.000 to 200.000 and region3 < 50.000. Within region 1 the cities (PSU) were self-selecting and the number of respondents were randomly drawn proportional to the size of the strata. Within region 2 and 3 the cities (PSU) were not-self selecting and in region 2 the number of desired respondents was 10 per PSU and in region 3 it was 5 per PSU. The sampling frame were the municipal personal records database of all selected municipalities in the Netherlands. One reason for incomplete sampling frames comes from the fact that municipal personal records database are not up-to-date due to moving, invalid addresses, etc. These sampling frame errors happen more often for ethnic minorities and will be discussed in the section about non-response (3.2).

### 4. Non-Sample

Non-sample errors are errors that could also happen with census surveys. A distinction is made between two types of non-sample error (observation and non-observation error). The observation error is the part of the non-sample error that happens due to the incorrect obtaining, recording and processing of data and could only arise because observations are made (Bethlehem, 2004). The non-observations error only happens when it is impossible to make observations.

#### Observation error

Observation error has three sources: over coverage, measurement error and processing error. It is not uncommon for survey research among ethnic minorities in the Netherlands to have respondent in the gross sample who do not belong to the target population, but there is no evidence that survey research among ethnic minorities is more susceptible to it then survey research among other populations. For this reason error due to over coverage will be set-aside

in the remainder of this article. In addition over coverage can easily be discovered during fieldwork by simply verifying if a respondent satisfies the target population criteria. One related interesting finding about over coverage in research among ethnic minority is that it sometimes happens that respondents do not know they belong to the target population. They have never known that at least one of their parents is born in a country that makes them a target population (ref).

The same brief remark can be made about processing errors that arise because of a misrepresentation during data processing. They happen in all sorts of surveys and are not specific for surveys among ethnic minorities. A more general warning about data processing applies for household surveys. The data from each member of household in a household survey should not be processed as independent data since at least some of the characteristics of household members are correlated (Lynn, 2002).

Measurement error will be discussed in more detail here since survey research among ethnic minorities in the Netherlands clearly suffers from it more than other survey research. The researcher can be unaware of specific cultural influences and or language problems in the survey data gathered among ethnic minorities and these data thereby do not necessarily match the reality. The interviewer, the respondent, the measurement instrument and the measurement mode can cause bias in surveys among non-western non-natives.

## 5. The interviewer

When conducting surveys among ethnic minorities in the Netherlands several choices can be made regarding the interviewers. Since a considerable proportion of the ethnic minorities do not speak Dutch one could choose for bilingual interviewers. The bilingual interviewers share the same ethnic background as the target population and speak both Dutch and the language of the target population (for instance Turkish, Arabic or Berber). An additional advantage of these bilingual interviewers is their familiarity with both the culture and customs and therefore a shorter social distance with the target population (Veenman, 2002). There are also downsides to using a bilingual interviewer. Schothorst (2002) found that bilingual interviewers need more training and instructions than Dutch interviewers. They need more training in the use of the questionnaire and while giving these training instructions it is better to give them orally instead of written down, because written instructions are not always read correctly. Also other researchers have comments on the use of bilingual interviewers (Kemper, 1998; Land van 't, 2000). They both find that although their knowledge of language and culture is an advantage, there is also an advantage for using Dutch interviewers because of their perceived neutrality regarding regional, political and religious divisions within the ethnic community. Research has shown (Land van 't, 2000) that ethnic origin of the interviewer and the language spoken during the interview has a clear influence on ethnic-related subjective questions. According to Kemper (1998) a Dutch interviewer who knows both language and the cultural etiquette would be best, however these interviewers are scarcely available. He also suggests interviewer duos made up of ethnic and non-ethnic interviewers to oppose the influence of language and cultural differences. To summarize it can be said that when conducting surveys among ethnic minorities the choice and training of interviewers is an important step where both ethnic as Dutch interviewers have the advantages and disadvantages. The use of bilingual interviewers leads to more interviewer mistakes and the use of Dutch interviewers leads to more non-response due to language problems. The ethnicity of the interviewer influences the answers in multiple ways with no clear 'best

solution'. In both the LAS and the SIM survey the choice was made to use both bilingual (Turkish-Dutch, Arabic-Dutch and Berber-Dutch) and Dutch interviewers<sup>8</sup>. The interviewers got instructions according to the guidelines given by Schothorst (2002). During the fieldwork phase of the SIM survey the drop out rate among bilingual interviewers was quite high and the recruitment of bilingual interviewers was difficult. It was so difficult that an additional research company, who specialises in the use of bilingual interviewers among ethnic minorities, was contacted to supply us with additional bilingual interviewers. The bilingual interviewers were used during regular fieldwork and also among initial refusals where the Dutch interviewer determined it was because of language problems. (Numbers are forthcoming). A further analysis of fieldwork data and survey data has to determine if the choice for both interviewers did a better justice to the diversity of the cultural and language influences among ethnic minorities.

## 6. The respondent

The non-western non-native respondent differs from the Dutch respondent. In order to get reliable answers from a non-western non-native respondent one has to take different question phrasing and methods of approach into consideration. Non-western non-native respondents like to talk more and during the interview time should be made available for a talk. In general the length of an interview takes one and a half to two times as long compared to Dutch respondents (Schothorst, 2002). The often used precoded answering categories function less well because of non-attitudes among non-western non-native respondents. There is a difference between the "us"-culture of non-western non-native respondents and the Dutch research culture (Kemper, 1998). Question about aspects that are considered important in the Dutch culture are not always considered equally important to Moroccan and Dutch respondents. In order to adequately measure the same construct among Moroccan respondents questions based on Moroccan aspects are essential (Land van 't, 2000). Social desirability in combination with cultural norms and values could have a major influence on the answers. In some cultures it is considered impolite to say 'no'. Other role-dependent respondent characteristics such unfamiliarity with surveys, scepticism and distrust occur more often when interviewing non-western non-native respondents (Veenman, 2002). Besides all the aforementioned measurement problems while conducting surveys among non-western non-native respondents there are other factors that should be taken into account. Most of these factors concern the interview setting and the interviewer-respondent relation. In these cases other rules apply for non-western non-native respondents compared to Dutch respondents. For instance, with Turkish and Moroccan respondents the influence of a mismatch between the interviewer and the respondent on sex, religion, education, age and or ethnic origin should be taken into consideration (Veenman, 2002). Even within an ethnic minority group these differences could cause an effect (Veenman, 2002). Examples of divergent interviewing settings are the difference in balance of power within many families of ethnic minorities; more often there are other family members in the direct vicinity while conducting the interview and the dominant presence of a television (Veenman, 2002). An option to stop some of the disrupting influences is to do the interview somewhere else instead of the house, but a different interview setting there would be other disrupting influences such as peer influences.

---

<sup>8</sup> Surinamese and Antilleans speak among other languages native Dutch so bilingual interviewers were not used.

The message seems to be that with sensitive topics or opinions the actual answers could differ from the reality due to all sorts of disrupting effects. Some of these disrupting effects could be taken into consideration in advance such as matching of respondents and interviewers, simpler language, extended interview length, more 'ethnic' topics, avoiding 'sensitive' topics, improved navigational path to single out non-attitudes, with other effects one can only be aware of their disruptive influence and one should be more careful with claims based on these answers. In both the LAS and the SIM survey we made use of many of the recommendations concerning the disrupting effects and we also attempted to map the illusive other effects by means of checklists. The interviewer was asked to fill out a small checklist after the interview with questions regarding the quality of the interview, such as if the respondent had understood the topics in the survey, if there was someone else present during the interview and if yes; who was it? (Forthcoming: the usefulness of the checklist is still being analysed)

### Measurement instrument

In survey research also the ethnicity of the respondent plays a part in the choice of the measurement instrument. A structured questionnaire with short and concrete questions is preferred. The questionnaire should not be too lengthy and when dealing with a sensitive topic one should make use of longer introductions, indirect or open questions or gradually increase the sensitivity of the questions via routing (Kemper, 1998). Questions or beliefs regarding the future should be avoided when interviewing the respondent of an ethnic minority and the question order can be different in comparison with interviewing a Dutch respondent (Schothorst, 2002; Veenman, 2002). Questions about income do not necessarily have to come at the end of the questionnaire and extra attention should be paid to the difficulty of the questions and answering categories. The wording of the questionnaire should relate to the level of the respondent and translated as much as possible (Schothorst, 2002). The LAS and the SIM survey both made use of different language versions (Dutch, Turkish and Arabic) of the questionnaire. Questions or beliefs regarding the future were avoided and the wording was adapted. A company specialised in language problems among ethnic minority (Bureau Taal) checked the wording of the questionnaire and made the wording more accessible when needed. A specific language problem among Moroccan respondents is the fact that some predominantly speak Berber, which is not a written language. In those cases a questionnaire cannot be translated in advance and consensus needed to be made with bilingual interviewers about which words they would use in Berber to clarify potentially difficult words and or topics.

### Measurement mode

The ethnicity of the respondent also influences the choice of the measurement mode in survey research. The (not translated) written questionnaire is not a useful mode due to language problems and a higher illiterate rate among (older) members of ethnic minorities. This mode would increase the probability of selective response and this would also be the case with internet surveys. In addition to the aforementioned problems internet access and percentage of ethnic minorities, specifically Moroccans and Turks, owning and using a computer is also lower (Ingen van, Haan de & Duimel, 2007, forthcoming). Among younger members of ethnic minorities, especially the ones that had their education in the Netherlands, the written questionnaire is an option (Schothorst, 2002). The most suitable mode is the oral interview. Of all possible types of oral interview computer assisted personal interview (CAPI) is the best and computer telephonic interview (CATI) should be avoided (Schouten, 2005). Among ethnic minorities there are far less landlines and it is seen as highly unusual to do an interview

via CATI (Schothorst, 2002). To assess whether CAPI mode effects potentially ‘sensitive’ topics, Statistics Netherlands experimented with computer assisted self-interviewing (CASI) among young members of ethnic minorities in 2003. There did not appear to be an effect on the answers. Except the choice of the measurement mode the interview needs to be preceded by a letter introducing the interview (Schothorst, 2002).

In both the LAS and the SIM survey the CAPI mode was used and a letter introducing the interview preceded the interview.

### Non-observation

Non-observation errors are due to the inability to do observations among the target population (Bethlehem, 2004). This is caused by under coverage and non-response. Both sources deserve extra attention when conducting surveys among ethnic minorities.

### Under-coverage

Under-coverage happens when not all elements of target population can be found in the sampling frame. These errors in the sampling frame occur when a sampling frame is not up to date. In the Netherlands (semi-) governmental and scientific institutes mainly use the postal data service or municipal personal records database as a sampling frame when conducting research among (the largest four) ethnic minorities. Both frames suffer from sampling frame errors due to moving of respondents, no known address of respondents, death of respondents, or slow registration of new respondents. From the LAS survey and other surveys (Schmeets, 2005). It turns out that these sampling frame errors (in the Netherlands known as false non response) occur more often among ethnic minorities (table 3). The quality of the data and estimators can be affected by high drop out rates because the net sample size would get too small and the survey data would be under used or mistrusted.

**Table 4: overview false non-response in LAS2004**

Ethnicity	Dutch		Turks	Moroccans		Antilleans	Surinamese
	N	%	N %	N	%	N	%
false non response	46	4	153 8	142 7	189 10	153	7

Source: LAS2004 veldwerkverslag

In conducting surveys among ethnic minorities in the Netherlands both household samples based on an address (SPVA) and persons samples (OGJA, LAS, SIM) have been used. If one chooses a persons sample in the design phase the sampling frame error due to the moving of a respondent could be reduced by ‘tracing’ the respondent to their new address, whereas the choice for a household samples based on an address excludes this possibility and there will be a substantial probability the new residents of the address are not eligible. Disadvantages of ‘tracing’ a respondent are the extra costs and time of the fieldwork. Another option of reducing the number of sampling frame errors is to sample more than once during the fieldwork phase. There are several drawbacks with this option. At the subsequent sample draws will be other potential respondents in the sampling frame that were not eligible at the initial sample draw. The second disadvantage with this method is the less than optimal

approach strategy by the interview corps. They may have to make several trips to the same area.

Another bias related to household samples based on an address in the Netherlands is the occurrence of multiple households on a single address (Schmeets, 2005). This occurs more often among ethnic minorities and there could be additional indistinctness about who needs to be interviewed. At the moment there seems no clear solution for sampling frame errors. As long as the accuracy of sampling frames is not up-to-date due to processing time it will limit the possibilities for the researchers to reduce it. One recommendation is to allow for sampling frame errors in the design phase by oversampling certain strata (for instance age- or ethnic groups or regions) by making use of detailed response reports of other surveys conducted among the same target population. Both LAS and SIM had their samples drawn only once and both made use of the (selected) municipal personal records database. In the design phase of the SIM survey this was done by using the detailed response data collected during the LAS survey fieldwork. This led to an oversampling in the 4 major cities of the Netherlands (Amsterdam, Rotterdam, Utrecht and Den Haag).

### Non-response

In surveys the ethnic minorities in the Netherlands generally have lower response rates. There are multiple reasons for their higher non-response rates and a number of them have been already discussed. Language problems continue to be an important reason for a lower response rate among (older) Turks and Moroccans. Sometimes members of ethnic minorities can be persuaded to participate after an initial refusal if the interview can be conducted in the native language (Kemper, 1998).

Another reason for a higher non-response rate among ethnic minorities is caused by several socio-demographic characteristics. Big city dwellers and persons with lower social-economic positions have a lower response rate and ethnic minorities are overrepresented in those categories (Stoop, 2005). Statistics Netherlands compared Dutch and respondents of ethnic minorities with similar socio-demographic characteristics and they found only slight differences in the willingness to participate in survey research (CBS, 2005).

A third reason for a lower response rates among ethnic minorities is the fact that they are hard to contact (Stoop, 2005). Members of ethnic minorities are harder to contact because they are out of their home more often (Kemper, 1998). To make an appointment to conduct an interview is also not a good strategy when surveying ethnic minorities, they seem to honour those appointments less than the Dutch respondents (Kemper, 1998).

Increasing the number of contact attempts, variations in time of contact attempts as well as variation in the location for contact attempt can reduce the non-response. Especially Sundays and evenings are very suited for surveying ethnic minorities (Kemper, 1998; Schothorst, 2002). Sundays are more suited for members of ethnic minorities who are Muslim since Friday is their day of worship and Sunday is more of a regular day for them. Schools seem a good location to conduct an interview with young members of ethnic minorities (Schothorst, 2002).

As mentioned before in both the LAS and SIM survey translated questionnaires and bilingual interviewers were used to reduce language related non-response. The number of contact attempts to at least 4 was another step to reduce the non response for the LAS and SIM survey as was the re-approaching of both respondents with language problems and initial refusals by bilingual interviewers were the initial interviewer thought they refused because of language



problems. Furthermore for the SIM survey incentives such as sending stamps and money were used to increase response rates. Also a special phone number was employed where potential respondents received money if they left a phone number where they could be reached by the interviewer to set up an appointment for an interview. (Response numbers about all efforts are forthcoming).

### Fieldwork organisation

An aspect of survey research that can sometimes be what neglected, but does have a significant impact on the quality of the survey measurements is the fieldwork organisation. Especially during lengthy fieldwork periods, the fieldwork organisation affects the quality of the survey measurements because of selective response caused by (the use of) the interviewer but also the level of under-coverage and non-response are affected.

During the organisation of the fieldwork a trade-off should be made whether the focus is on the respondent or the interviewer. When the focus is on the respondent the fieldwork should be organised in such way that either the gross sample is evenly utilised based on several relevant (for the survey) background variables or respondents are drawn a select from the gross sample. Otherwise the quality of measurements of variables were timeliness in combination with background variables such as the level of urbanicity and age, affect each other could be comprised. In this way bias due to change in time or recent events during the fieldwork period that would cause a different opinion among certain groups can be avoided. Another option when focusing the fieldwork organisation on the respondents is to start the fieldwork with traditionally difficult or hard-to- contact groups because they are most time consuming. Part of the non-response due to 'element of gross sample not used' or 'no contact' that could be selective can be reduced if the focus of the fieldwork organisation is not the easy to contact or cooperative respondents but on the difficult or hard to contact groups (in the Netherlands for example young, Antillean males living in big cities). An additional advantage is the reduction in total survey error due to under-coverage because there is less chance that certain difficult or hard to contact groups with greater tendency to move have left.

When the focus is on the interview during the organisation of the fieldwork the goal will be to make efficient use of the interviewers corps by reducing travel time of the interviewer (and thereby the costs). Selecting respondents from the gross sample who live close together can do this. A common practice is to select the respondents from the gross sample that live close to the home address of the interviewer first. Another common practice is to select both the closely related and the easy respondents first to get an interviewer interested in the survey and thereby creating a smaller chance for an interviewer to quit the survey.

The choice for the interviewer is a choice for time and money and a choice for the respondent is a choice for increasing the quality of the measurements.

Other, related trade-offs that need to be made during the organisation of the fieldwork are the amount of 'freedom' given to the interviewers. When given number respondents to which degree are interviewers free in planning the order or route to approach the respondents? The maximum of interviews that one interviewer is allowed to do is another choice that should be given some thought during the fieldwork organisation. How much interviewer-effect is acceptable when the interviewer is good and motivated? Where and when to use bilingual interviewers is the next choice during the organisation of fieldwork when conducting this type of survey. Also the amount of training and check ups on interviewers needs to be given some thought during the organisation of the fieldwork.

In the LAS survey the focus was on the interviewer and in the SIM survey a more mixed mode was adopted due to careful monitoring of the fieldwork. When potential difficult groups were identified the focus shifted to the respondents by selecting and approaching them first, also the number of contact attempts was increased. Both surveys tried to make as much use of bilingual interviewers especially for elderly respondents of ethnic origin. Interviewers were given quite a bit of freedom to plan their own order and route in approaching respondents although they had to visit the respondent on both different times and days when no contact was made so far. Regarding interviewer effects there was no maximum number of interviews set for the interviewer. However, in combination with regular random checks on interviewers, respondents who had done an interview conducted by an interviewer with a very high response were sometimes contacted again to ask what they found of the interview.

## 7. Conclusion

While conducting surveys among ethnic minorities many things have to be taken into consideration to ensure a high data quality. Obviously the design and all related sample issues need to be given the proper amount of attention, but the main difficulty with conducting research among ethnic minorities lies with measurement error and non-response. To ensure higher data quality both response rate and measurement can be improved. It helps to increase the number of contact attempts, alternate time and days of contacts (evenings and Sundays for Muslims) and to re-approach respondents with (initial refusal due) language problems. (Conclusion about the use of incentives is forthcoming). More than one sample during the fieldwork period is not recommended. Selective use of bilingual interviewers is very useful. Especially with re-approaching (initial refusals) respondents with language problems and older members of ethnic minorities where there is an increased probability of language problems and/or illiteracy. CAPI is the preferred way to conduct the survey. Translated questionnaires work well and can be conducted by Dutch interviewers. The translated questionnaire is given to the respondent so they can look up the questions they don't understand in Dutch. Interviews do take a longer time among ethnic minorities and make sure the length of the questionnaire is not too long. The use of simple language is recommended and the use of both questions about the future and sensitive topics should be avoided when surveying ethnic minorities. Finally, more time and attention should be given to both the organisation and the monitoring of the fieldwork.

## References

- Beer, de J. (2001), De demografische prognoses van het CBS. CBS intranet, maart 2001.
- Bethlehem, J. (2004) Technieken voor survey-onderzoek. CBS-publicatie. (Centraal Bureau voor Statistiek, Voorburg/Heerlen)
- Couper, M.K. & Groves, R.M (1996) social environmental impacts on survey cooperation. *Quality and Quantity*, vol. 30. pp 173-188.
- Dagevos, J. & Schellingerhout, R. (2003) Sociaal-culturele integratie: contacten, cultuur en oriëntatie op de eigen groep. Rapportage Minderheden 2003. SCP-publicatie.
- Feskens, R.C.W., Hox, J., Lensveldt-Mulders, G. & Schmeets, H. Multivariate analyses of non-response among ethnic minorities. Paper presented at Statistics Canada Symposium 2004, Innovative methods for surveying difficult-to-reach-populations.

- Groeneveld, S. & Weijers-Martens, Y. (2003) *Minderheden in beeld*. SPVA2002. ISEO
- Groves, R.M (1989) *Survey errors and survey costs*. (Wiley, New York)
- Groves, R.M & Couper, M.P (1998) *Non response in household interview surveys* (Wiley, New York)
- Hilhorst, M. (2007) *Veldwerkverslag Survey Intergratie Minderheden*.
- Ingen, van E., Haan, de J.& Duimel, M. (2007) *Conceptrapport E-vaardigheden* (forthcoming). SCP-publicatie.
- Kemper, F. (1998) *Gezocht: Marokkanen. Methodische problemen bij het werven en interviewen van allochtone respondenten*. *Migrantenstudies*, 1, p. 43-57.
- Land, van 't H. (2000) *Similar Questions; Different Meanings*. Academisch proefschrift. Amsterdam:Vrije Universiteit te Amsterdam
- Lynn, P. (2002) *Sampling, Non-response and Weighting*. Workbook University of Essex.
- Martens, E.P. (1999) *Minderheden in beeld*. De SPVS-98 (NIWI, Amsterdam).
- Marlet, G. & Woerkens, C. van (2003) *Atlas voor Gemeenten 2003*
- ONS (2003) *Ethnic group statistics: A guide for the collection and classification of ethnicity data*. (Office for National Statistics, Newport).
- Reep, C. (2003) *Moeilijk waarneembare groepen: een inventarisatie*. CBS-rapport H1568-03-SOO (Centraal Bureau voor Statistiek, Voorburg/Heerlen)
- Schmeets, H., Dominguez, S. & Veenman, J. (2002) *Jaarboek minderheden*. Bohn, Stafleu en Van Loghum (Houten/Diegem 2002)
- Schmeets, H & Bie, van der. R (2005) *Enqueteonderzoek onder allochtonen* CBS. CBS-publicatie (Centraal Bureau voor Statistiek, Voorburg/Heerlen)
- Schothorst, Y. (2005) *Veldwerkverslag Leefsituatie Allochtone Stedelingen*.
- Schothorst, Y. (2002) *Onderzoek onder allochtonen: wat mag, wat moet en wat kan?* (In: Houtkoop- Steenstra H., Veenman, J. *Interviewen in de multiculturele samenleving. Problemen en oplossingen*. Koninklijke van Gorgum 2002.)
- Stoop, I.A.L. (2005) *The Hunt for the last respondent*. Academisch proefschrift. Amsterdam:Universiteit van Utrecht
- Veenman, J. (2002) *Interviewen in multicultureel Nederland* (In: Houtkoop- Steenstra H., Veenman, J. *Interviewen in de multiculturele samenleving. Problemen en oplossingen*. Koninklijke van Gorgum 2002.)
- Voogt, R. (2004) *"I'm not interested". Non response bias, response bias and stimulus effects in election research*. Academisch proefschrift. Amsterdam:Universiteit van Amsterdam
- Zijleman, E.L.M., Braat, H.J.M & Martens, E.P (1989) *Kerncijfers S.P.V.A ( Onderzoek Sociale Positie en Voorzieningsgebruik Allochtonen)*

## Choosing Web Surveys: mode choices among Youth Cohort Study respondents

Heather Wardle & Chloe Robinson

### Abstract

Web surveys are increasingly being offered as an alternative mode of completion within many large scale surveys, with the view that this will increase rates of co-operation. The Youth Cohort Study (YCS), has offered respondents the option of completing the questionnaire on-line as an alternative to the standard postal questionnaire at all four sweeps of the 11<sup>th</sup> cohort. Beginning in 2002, the proportion of respondents choosing to complete the questionnaire on-line has increased from 1% in sweep 1 to 21% in sweep 4. Analysis of who chooses to complete the web survey shows that respondents do not consistently use the same mode of completion at subsequent sweeps of the study. By sweep 4, 65% of respondents who completed the questionnaire on-line had not chosen this mode previously and were new “web responders”. The characteristics of these new web responders were substantially different from those who returned the postal questionnaire. Those who completed the questionnaire on-line were more likely to still be in full time education than those who completed the postal questionnaire. Likewise, web respondents had significantly higher levels of academic qualifications than postal respondents. In short, the profile of web-respondents evident at sweep 4 is that of respondents who are already known to be the most likely to co-operate in the study. Therefore, in the case of YCS, offering a web survey option as an alternate mode of completion, an increasingly popular practice among survey researchers, is unlikely to help address non-response bias.

### Keywords

Web surveys, attrition, non-response, mode choices

### 1. Introduction

The Youth Cohort Studies (YCS) have been running since 1985 and are a series of surveys among young people aged 16 and upwards, which monitor their decisions and behaviour in making the transition from compulsory education to further or higher education, employment or another activity. The main aim of the research is to identify and explain the major factors influencing these transitions after school, such as levels of educational attainment, training opportunities and school experiences. The YCS is funded and managed by the Department for Education and Skills (DfES). In this paper findings are described from the 11<sup>th</sup> cohort of the series which ran for four sweeps between 2002 and 2005. The sample members for the first

sweep of the survey were drawn from a sampling frame of 35,000 pupils taken from the 2000/01 year 11 school registers.

This paper will discuss questionnaire mode choices among respondents to this 11<sup>th</sup> cohort. Respondents were offered the option of completing a paper questionnaire and returning it by post or electronically submitting an equivalent, on-line, version of the questionnaire. Attempts were also made to interview cohort members at each sweep who had not responded by a certain date through the method of computer assisted telephone interviewing (CATI). CATI respondents have been excluded from the analysis in this paper because their characteristics (as initial non-responders) differ to those respondents who proactively returned a postal questionnaire or completed a questionnaire on-line. Only two mode choices were advertised to respondents, postal or web, and we are primarily interested in who chose each method. As well as both the postal and internet modes being proactive mode choices for the respondents, there are also minimal mode effects between them because both are administered in a visual format and neither involve an interviewer.<sup>2</sup> This allows for a reasonable comparison of the two modes..The focus of this paper is not to compare responses between the two modes but to compare the actual respondents to the two modes.

This paper will present the proportion of respondents who chose to complete the questionnaire on-line at each sweep of the 11<sup>th</sup> cohort of YCS and explore whether web-responders consistently chose this mode at each sweep as an indication of how preferable the internet is as a means for responding to the survey. An on-line version of the survey offers a number of advantages including being able to automatically monitor questionnaire returns and reduced financial and environmental costs and therefore there is considerable benefit in knowing whether people do opt to respond on-line. This paper will also look at who chooses to complete the survey on-line and consider whether an on-line option as an alternative, potentially more flexible and manageable, mode encourages those who may otherwise have not taken part in the survey.

## 2. Mode choices by YCS11 respondents

The first sweep of the Youth Cohort Study (YCS), 11<sup>th</sup> cohort, began in 2002. This was the first year in which respondents were presented with two options of how to complete the questionnaire. The first was the standard postal questionnaire, used in previous YCS cohorts. The second was a specifically developed web questionnaire, which replicated the postal survey questions. All selected respondents were presented with both options and given a unique web survey password with which to access the web questionnaire. Mode of completion was left entirely to the discretion and preference of the respondent. Additional interviews were conducted using CATI with a subset of the sample who, approximately 20 days from initial contact, had not responded by either the postal or web methods. For purposes of comparison, these initial “non-responders” have been excluded from the figures presented in table 1, so to focus on proactive mode choices made by YCS11 respondents.

<sup>12</sup> Kaminska, O., Bautista, R., Serrano, E. *Best Combination of Modes* The American Association for Public Opinion Research (AAPOR) 62th Annual Conference, 2007

Table 1 shows the proportion of respondents in each sweep between 2002 and 2005 who chose each method of completion.

Table 1. Mode of completion among YCS respondents, by sweep				
Data collection mode	Sweep 1 (2002)	Sweep 2 (2003)	Sweep 3 (2004)	Sweep 4 (2005)
	%	%	%	%
Web survey	1	9	14	21
Postal survey	99	91	86	79
<i>Base</i>	<i>13719</i>	<i>5493</i>	<i>3729</i>	<i>3687</i>

*Base: All postal and web responders at each sweep*

The proportion of respondents who chose to complete the questionnaire on-line rose from 1% in 2002 to 21% by 2005. This represents a significant increase in the percentage of respondents choosing to complete the questionnaires on-line within each sweep. Part of this increase, particularly between sweep 1 and sweep 2, can be attributed to practical improvements in the study administration. For example, the internet option in sweep 2 was advertised in a pre-notification letter to cohort members and also contact was made with cohort members by e-mail after the pre-notification letter which contained a link to the web survey allowing cohort members to access the questionnaire more easily. E-mail addresses were systematically reviewed at sweeps 2 to 4, for example, any spaces were removed and words such as 'hotmail' were checked for spelling errors. Furthermore, for sweep 2 the following URL was purchased [www.pathways2003.com](http://www.pathways2003.com), as it was deemed that the URL used previously was too long and not very memorable. The increase in respondents choosing to complete the questionnaire on-line between sweep 1 and sweep 2 was viewed by researchers to be largely related to these improvements.

These improvements were replicated for all subsequent sweeps of the study. As such, not all of the observed increase in proportion of respondents choosing the web survey as a mode of completion can be attributable to these administrative improvements. Furthermore, analysis of cohort members who chose to complete the web survey in sweep 2 shows that cohort members do not consistently use the same mode of completion for subsequent phases, but rather use different modes in different sweeps.

Table 2. Choices in mode of completion in Sweeps 2, 3 and 4, by sex

Data collection mode	Young Men %	Young Women %	<i><b>Total %</b></i>
Web survey respondent all sweeps	5	3	<b>3</b>
Mainly web survey respondent (chose web survey in 2 out of 3 sweeps)	7	7	<b>7</b>
Mainly postal survey respondent (chose postal questionnaire in 2 out of 3 sweeps)	11	15	<b>14</b>
Postal respondents all sweeps	77	75	<b>76</b>
<i><b>Base</b></i>	<i>629</i>	<i>1376</i>	<i>2005</i>

*Base: All postal and web responders at sweeps 2, 3 and 4 (excludes CATI responders at any wave)*

Table 2 shows that only 3% of cohort members who participated in sweeps 2, 3 and 4 chose to complete the questionnaire on-line in all sweeps.<sup>3</sup> A further 7% of cohort members chose to complete the web survey in two sweeps and the paper questionnaire in one sweep, whilst an additional 14% of respondents completed the questionnaire by returning the paper booklet for two sweeps but also completed the questionnaire on-line for a least one sweep. The net result of these varying choices was that by sweep 4, 65% of cohort members who chose to complete the questionnaire on-line at this data collection phase had not chosen this mode of completion in earlier sweeps. These respondents were therefore “new” web respondents who were now choosing to mix their mode of completion compared with previous phases.

### 3. Web response and non-response profile

#### Profile of web respondents

As all respondents were offered the option to complete the questionnaire on-line, and increasing numbers of respondents chose to do so, even if they had not used this method in previous sweeps, insight into the profile of cohort members who chose this mode of completion can be gained.

Isolating co-operating cohort members from sweep 4, table 3 compares some key statistics between those respondents who completed the questionnaire on-line and those who completed the postal questionnaire.

<sup>3</sup> This analysis focuses on those respondents who completed either the paper questionnaire or the web survey at all sweeps. Therefore any respondents for whom information was collected by the CATI questionnaire are excluded.

Table 3. Key statistics by mode of completion at sweep 4

Key statistics	Postal survey	Web survey
	%	%
Proportion with 5 or more GCSES, grades A*- C	76	90
Proportion reporting at sweep 1 all/most peers planning on going into higher education	72	80
Proportion who had been expelled or suspended	3	1
Proportion who ever truanted	27	20
<b>Base</b>	<b>2915</b>	<b>772</b>

*Base: All postal and web responders at sweep 4 (excludes CATI responders at sweep 4 only)*

A significantly higher proportion of respondents who chose to complete the web survey in sweep 4 had 5 or more GSCE's, grades A\* - C than those who completed the postal questionnaire, 90% and 76% respectively. In addition to achieving higher levels of educational attainment, those completing the web survey were less likely to have been expelled or truanted whilst in compulsory education than their postal respondent counterparts, though some element of co-linearity between these factors and educational attainment may be expected. Multivariate logistic regression has been used to examine the independent factors associated with choosing to complete YCS11 questionnaire on-line in sweep 4. The regression technique adjusts for several explanatory variables simultaneously. Key variables of interest were entered, including a number of socio-demographic factors such as sex, National Statistics Socio-Economic Classification (NS-SEC) of the cohort member, educational attainment, and education/employment/training status. Only variables that were significant in the final model have been presented in table 4.



Table 4 Estimated odds ratios for completing the YCS11 questionnaire on-line in sweep

	N	Odds ratio	95% C.I. <sup>a</sup>
<b>Ethnic group (p&gt;0.05)</b>			
White	5434	1	
Black/Black British	98	0.22	(0.07, 0.70)
Asian/Asian British	401	0.74	(0.53, 1.03)
Other	220	0.98	(0.65, 1.46)
Not answered	27	0.67	(0.16, 2.93)
<b>Whether living with parents (p=0.000)</b>			
Lives with both parents	2423	1	
Lives with one parent	630	1.14	(0.81, 1.61)
Does not live with either parent	2509	1.16	(0.93, 1.44)
Not answered	618	2.47	(1.99, 3.06)
<b>NS-SEC of cohort member (p=0.000)</b>			
Managerial/professional	463	1	
Intermediate	722	1.03	(0.67, 1.59)
Routine/semi-routine	2336	0.74	(0.50, 1.08)
Not answered	2659	1.46	(0.99, 2.15)
<b>Highest level of educational attainment (p=0.000)</b>			
2 or more A/AS levels	3522	1	
1-1.5 A/AS levels	393	0.70	(0.48, 1.02)
5 or more GCSES A*- C	915	0.69	(0.51, 0.92)
Less than 5+ GCSES A*- C	1350	0.44	(0.32, 0.61)
<b>Whether in employment, education or training</b>			
Full time education	3495	1	
Employment with training	771	0.46	(0.30, 0.68)
Employment without training	1048	0.50	(0.36, 0.70)
Other education and training	185	0.43	(0.22, 0.84)
Government Supported Training	292	0.44	(0.24, 0.81)
Not in Education, Employment or Training (NEET)	389	0.50	(0.32, 0.76)

The odds of completing the questionnaire on-line were significantly associated with educational attainment, with the odds becoming lower as level of maximum educational attainment decreased. For example, odds of completing the questionnaire on-line were some 0.44 times lower among those whose maximum educational attainment was less than 5 or more GCSEs, grades A\*-C than those who had achieved 2 or more A/AS levels by sweep 4. Likewise, odds of completing the questionnaire on-line were also lower among those respondents who were in employment (either with or without training) than those who were in full time education and were lowest among those who reported being in other education or training (0.43). Other variables were significant in the final model, including NS-SEC status of the respondent, where the odds of completing the questionnaire on-line were lower among those in routine and semi-routine occupations (0.74) and whether the cohort member lived with any of their parents or not. Ethnic group was also significant, with the odds of Black or

Black British cohort members choosing to complete the questionnaire on line being 0.22 times lower than White cohort members. Gender, experience of truancy and/or suspension or expulsion were included in preliminary analysis, but these variables were not significant in the final model.

Analysis presented in tables 3 and 4 show that there are some key differences between cohort members who chose to complete the questionnaire on-line and those who did not. On-line responders are more likely to have a higher level of academic achievement, still be in full time education at the age of 19-20 and, when asked in sweep 1, more likely to have reported that most peers were continuing to full time education. These differences may not in themselves be surprising, easier access to the internet for those in full time education, for example, may be an important influencing factor. However, it is important to assess how these findings might affect overall survey precision, especially as data shows completing the questionnaires on-line was an increasingly popular choice among cohort members.

Gaining high response rates is considered an important quality marker of a study. However, in parallel to this, researchers must also consider the possibility that bias can be introduced into a study if the profile of those who do not respond to the study is vastly different to those who do. Cohort studies are effected by initial non-response, resulting from who refused to co-operate at the very first sweep, and also non-response between sweeps due to attrition in the sample. Earlier work on non-response to YCS has suggested that those who are less likely to co-operate are those who have lower levels of academic achievement, those who truanted at school, men, those of black ethnic origin, the unemployed, those looking after a family and those whose father is unemployed.<sup>4</sup> The potential for bias to be introduced in the sample due to attrition is considered more fully below.

#### Attrition among cohort members

Cohort studies can be subject to bias due to attrition caused by cohort members “dropping” out in between data collection phases as these people may have different characteristics to those who remain in the sample. Analysis of YCS11 cohort members who responded at sweep 1 but had dropped out by sweep 4 shows that, like other cohort studies, this potential for bias caused by attrition between phases was evident. Table 5 shows that there are very real differences between YCS11 cohort members who participated in all four sweeps (by any mode) and those who had only partial co-operated in some sweeps and by sweep 4 were non-responders. Significantly more fully co-operating cohort members had 5 or more GCSES, grades A\*- C than those who had dropped out of the sample by sweep 4; 75% compared with 59% respectively. Likewise, fully co-operating cohort members were also more likely to report that all or most of their friends planned on continuing into higher education (76%), whilst partially co-operating cohort members were somewhat less likely to report this (69%). Significantly more cohort members who did not participate in all four sweeps were more likely to report ever being suspended or expelled, or playing truant than their fully co-operating counter parts.

<sup>4</sup> Taylor, S and Lynn, P (1996) England and Wales Youth Cohort Study (YCS). The effect of time between contacts, questionnaire length, personalisation and other factors on response to the YCS. DfEE Research Series.

Table 5 Key characteristics of co-operating and non-co-operating cohort members by sweep 4

Key statistics	Cohort members who responded at all four sweeps (fully co-operating)	Cohort members who did not respond to all four sweeps (partially co-operating)
	%	%
Proportion with 5 or more GCSES,	75	59
Proportion reporting at sweep 1 whether most/all friends were planning on going into higher education	76	69
Proportion who had been expelled or	4	8
Proportion who ever truanted	29	37
<b>Sex:</b>		
Male	40	48
Female	60	52
<b>Base</b>	6180	10527

*Base: All respondents at sweep one using any mode*

At sweep 1, 45% of the original cohort were male and 55% were female. Population estimates for 2002, showed that approximately 51% of the English population aged 16-17 were male and 49% female. Therefore, at sweep 1, women were already overrepresented and men underrepresented in the sample population. As table 5 shows, by sweep 4 (where respondents were aged 19-20), 60% of cohort members were female and 40% were male, indicating that attrition among cohort members was most acute among males. The net result was that the sex profile of fully co-operating respondents was further removed from the sex profile of the population as a whole. (Population estimates show that in 2005, 52% of 19-20 year olds were male and 48% female). This, along with differences in educational attainment and behavioural differences such as expulsion from school, indicate areas of potential bias introduced into YCS through attrition between sweeps.<sup>5</sup>

#### 4. Discussion

Web surveys are increasingly being offered as an alternative mode of completion within many large scale surveys, often with the view that this will increase rates of co-operation. A web-based mode of completion was introduced into the YCS11 as it was felt that internet completion would appeal to the younger age group from whom information was being collected. Furthermore, it was also felt that mode effects would be minimal between postal and internet completion as both involve a visual channel of communication and neither involves an interviewer. The purpose of this paper was not to consider mode effects between each mode, but rather to assess any systematic differences between those respondents who chose to complete the questionnaire on-line and those completing the postal questionnaire. The increasing proportion of cohort members choosing to complete the questionnaire on-line across all four sweeps certainly shows that this was becoming an increasingly popular method

<sup>5</sup> Final YCS data are weighted to address non-response.

of completion. By sweep 4 in 2005, one fifth of proactive, co-operating respondents chose on-line completion and it was thought that “increased access to the internet, particularly for young people aged 19/20 could have contributed to this considerable increase in the internet response”.<sup>6</sup>

Consideration of the profile of those who chose on-line completion in sweep 4 supports this assertion. Cohort members remaining in full time education, with a higher level of academic achievement, were most likely to complete the questionnaire on-line and possibly likely to have greater access to the internet through their respective academic institutions. It was thought by YCS researchers that offering flexibility through different mode options and incorporating an on-line option would appeal to the age group represented by the YCS11 cohort and have potential importance in maintaining the goodwill of cohort members. The fact that cohort members frequently swapped their mode of completion between different sweeps suggests that cohort members utilised a mode of completion at each sweep that was convenient for them at that time. By sweep 4, 65% of those who completed the questionnaire on-line had not do so in previous sweeps and were “new web responders”.

Offering the internet as a mode of completion gave cohort members flexibility which they increasingly utilised. However, analysis shows that the on-line option appealed to a particular subset of the cohort sample. Those who chose the on-line option had significantly higher levels of educational attainment, were less likely to be from routine/semi routine occupations, more likely to be still be in full time education at the age of 19/20 and reported significantly fewer behavioural problems, such as expulsion or suspension when in compulsory education. Although, notably there were no significant differences evident in the choices made between male and female respondents. When these characteristics are compared with those who by sweep 4 had become “non respondents”, the differences between the two groups are acute. The profile of those who “dropped” out between cohort sweeps showed that significantly greater numbers were male, those with a lower level of educational attainment and those who reported great behavioural problems when in compulsory education. Therefore, whilst offering the internet as mode of completion may be useful for respondents as it gives them greater flexibility and allows researchers to utilise a number of different contact methods with cohort members (for example, via e-mail), this analysis shows that those who are most likely to choose the on-line questionnaire are those people who are already known to be most likely to co-operate in all phases of the study. This reinforces the viewpoint that whilst, in the case of YCS, offering a web survey option may have greater appeal to some cohort members and some cost saving efficiencies, overall it is unlikely to help address non-response bias and any evaluation of this methodology should take this into account.

## References

For ONS mid-population estimates see:

<http://www.statistics.gov.uk/statbase/Expodata/Spreadsheets/D9390.xls>

Finch, S (1996) England and Wales Youth Cohort Study: An investigation of the feasibility of using different data collection methods for the YCS, DfEE Research Series, 38, 1996.

<sup>6</sup> Phillips M, Johnson M, Robinson C, Tipping S (2005) Youth Cohort Study: Survey of 18 and 19 year olds (Cohort 11 Sweeps 3 and 4) Technical Report. Prepared for Department for Education and Skills. National Centre for Social Research 2005.

Jarvis L, Exley S, Tipping S (2006) *Youth Cohort Study: Survey of 17 Year Olds (Cohort 11 Sweep 2)*

*Technical Report* Prepared for Department for Education and Skills. National Centre for Social Research 2005.

Knight I, “Measuring Change in Labour Force Panel Studies” *Social Survey Methodology Bulletin*, 51, 2003.

Phillips M, Johnson M, Robinson C, Tipping S (2005) *Youth Cohort Study: Survey of 18 and 19 year olds (Cohort 11 Sweeps 3 and 4) Technical Report*. Prepared for Department for Education and Skills. National Centre for Social Research 2005.

Taylor, S and Lynn, P (1996) *England and Wales Youth Cohort Study (YCS): The effect of time between contacts, questionnaire length, personalisation and other factors on response to the YCS*, DfEE Research Series.

## About the Authors

Heather Wardle and Chloe Robinson are both Senior Researchers with the National Centre for Social Research, 35 Northampton Sq, London, EC1V0AX. The YCS study was funded by Department for Education and Skills (DfES). The views expressed in this paper are the authors’ and do not necessarily reflect those of DfES.

# What's the time? Relations between interview periods and output periods in surveys

Paul Smith, Charles Lound, Pete Brodie

## Abstract

Two aspects of quality in survey statistics are their timeliness and relevance. A considerable amount of discussion has taken place in the European Union over the timeliness of economic statistics in comparison with the USA. Part of this difference stems from the different approaches to the definition of the period being measured, and the difference of approach has an impact, often hidden, on the relevance (the degree to which the survey measures the required concept). In this paper we review the different ways in which periods are defined in surveys and their properties

## Keywords

Timeliness, relevance, quality trade-off

## 1. Introduction

Two aspects of quality in survey statistics according to the European quality components (Eurostat 2003 – quality components will be highlighted in bold italic where they are used in this paper, to highlight the trade-offs) are their ***timeliness*** and ***relevance***. A considerable amount of discussion has taken place in the European Union over the ***timeliness*** of economic statistics in comparison with the USA (see for example Öberg & Nanopoulos 2001). Part of this difference stems from the different approaches to the definition of the period being measured, and this difference of approach has an impact, often hidden, on the ***relevance*** (the degree to which the survey measures the required concept).

In this paper we review the different ways in which periods are defined in surveys and their properties. In section 2 we consider the definition of the survey reference period, how it can be modified by the survey procedures, and ways to adjust where the modification affects the relevance; in section 3 we consider the differences between surveys with continuous measurement and point-in-time surveys; in section 4 we examine moving averages and other filters, and in sections 5 & 6 we discuss the varying impacts on quality and draw some conclusions.

## 2. Reference period

The period to which a survey relates is sometimes an approximation to what the output is called. Monthly surveys are quite naturally called January, February, ... but in practice quite a number of series are based on approximations to calendar months based on weeks – the UK's Retail Sales Index (RSI) is one such output, an index of the average weekly sales in 4-week, 4-week and 5-week periods, which together make up a calendar quarter. The 4-4-5 week pattern creates some difficulties for seasonal adjustment, because holidays can sometimes move between periods, and particularly for RSI the growth in sales leading up to Christmas can be more in November in some years than in others. Every 4-5 years, it is necessary to insert an extra 5-week month (in RSI this is always a 5-week January) to keep the weeks and calendar months approximately aligned. The UK LFS (more detail later) takes this extra week as a "leap week" when no interviewing is done. There is a move to introduce a regulation to standardise the periods for Labour Force Surveys within the European Union, and to specify how the weeks will form months and when a year will start.

### Determining the reference period 'in the field'

In the UK where household surveys are run continuously, the questions asked typically refer to specific time periods. There are two ways in which these periods can be defined, and both approaches are in use. The Labour Force Survey (LFS) is specific about the time period at the time of sampling, so when a household is selected, it is selected for a particular week. (In fact the LFS has a pattern of 13 interview areas ("stints"), with a new area being covered each week, so that the areas are repeated quarterly on a consistent pattern). So if a household is selected for interviewing in week 2, the interviewees would normally be asked for information on their employment status in the previous week (week 1). Sometimes, however, for example because the household is difficult to contact, no interview is possible until week 3. In this case, the interview will still have questions which relate to the interviewees' situation in week 1. In fact this is a requirement in the harmonisation of LFSs in the European Union (Council regulation (EC) No 577/98). This means that the reference period for a particular interview is fixed (it is not altered in the field), and the number of interviews capturing information about week 1 is therefore also fixed (except for variations in non-response).

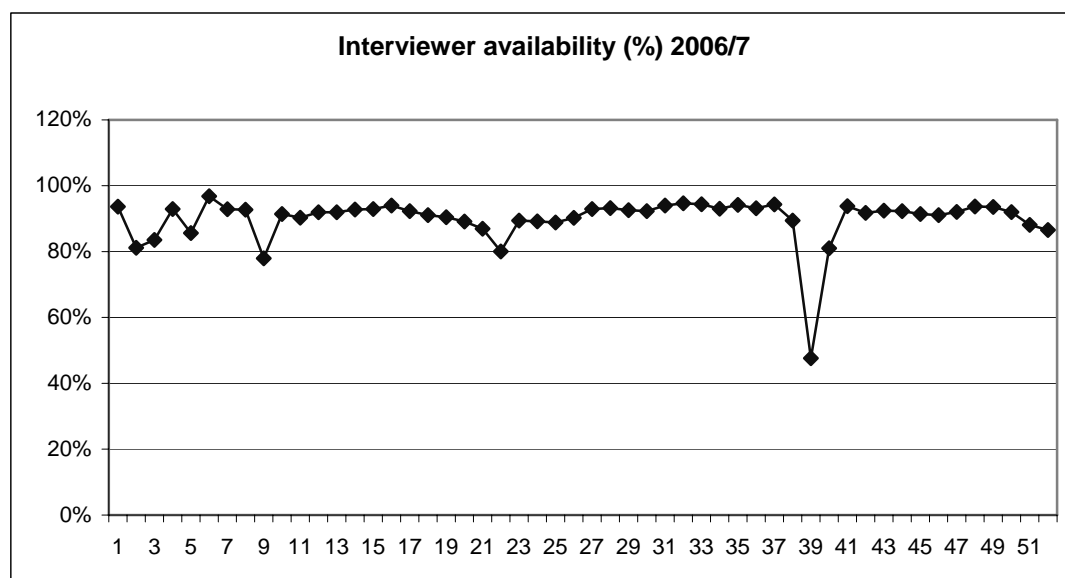


Fig. 1: Approximate availability of interviewers by week for social surveys in 2006-7. Lower values corresponding to weeks 2 & 3 are Easter, 5, 9 & 22 are weeks with public holidays, and 39 & 40 are Christmas.

For the other main household surveys in the UK the reference week for the interview is not fixed at the time of sampling. Continuous survey sample designs do ensure an even and balanced distribution of the set sample across months, but most allow interviewers to plan how they work through their monthly workload of sampled addresses, so that it can be done in the most efficient way and without further constraints. The Expenditure and Food Survey (EFS) is used to estimate average weekly household expenditure through an interview survey and a two-week expenditure diary, left with the respondents, with recording commencing on the day of the interview or the next day. On the EFS, interviewers are asked to aim for an even distribution of interviews and, therefore, diaries across the weeks within the month (Portanti 2007).

In spite of this aim of an even distribution, the interviewer will prefer to interview at any time during the month rather than lose a responding household. Therefore the number of responding households will vary according to the ease with which people can be contacted and the availability of the interviewer. The general availability of interviewers is certainly lower during holiday periods (Fig. 1) and the same will be true for those they are attempting to contact.

The effect of this on the achieved interviews in the EFS for April 2005 to March 2006 is shown in Fig. 2 – the average number of interviews here is 130 per week. There is a very obvious drop in week 39, which is for interviewing in Christmas week, and the weeks on either side are also very low. But there are also other low weeks and a 4, 4, 5 week pattern corresponding to the way in which interviewers work. They are encouraged (in their guidance documentation) to undertake as many interviews as possible in the first two weeks of the survey period, and this gives rise to the periodic series.

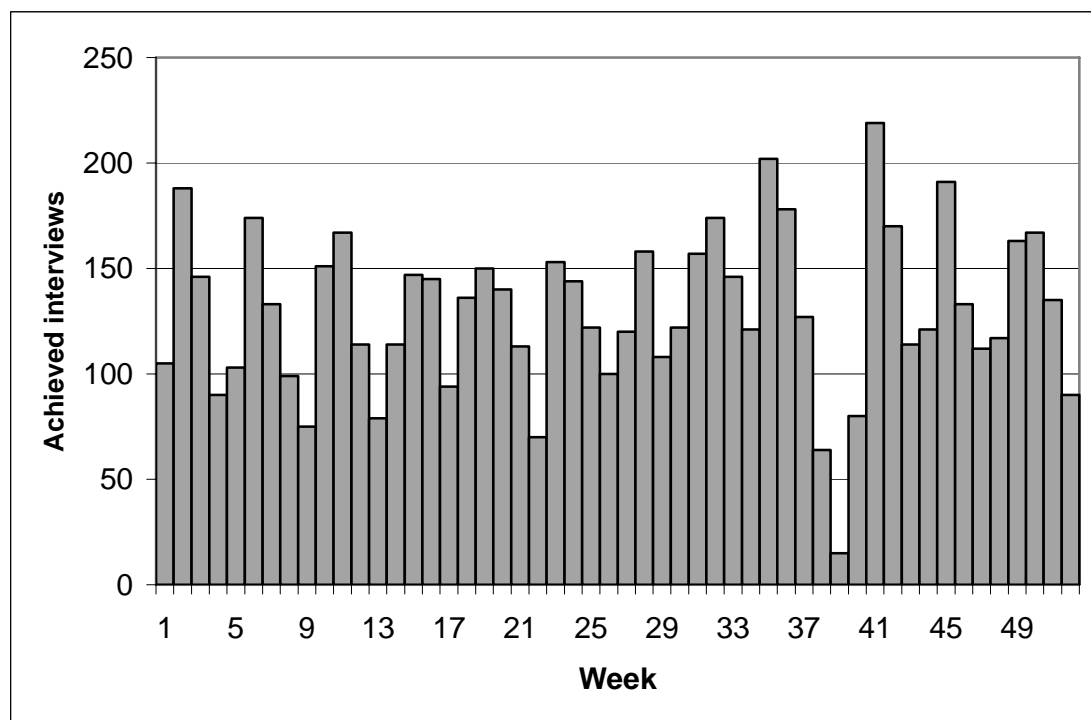




Fig. 2: Number of achieved interviews by week to which questions refer in EFS interviews for April 2005 – March 2006.

### Respondent-determined reference periods

For business surveys where the questions are not being posed by an interviewer, and where compliance with the survey's requirements is statutory, there has been pressure for many years to make it as easy as possible for businesses to complete the questionnaires (ONS 2005, Smith & Penneck 2007). One way in which this is done is to accept responses for any period which corresponds with the required one within certain limits (normally an earliest start date and a latest finishing date). The starting and finishing dates are recorded by the respondent on the questionnaire, and used to adjust the response values where necessary. This is equivalent to determining the reference period in the field, although different from the household survey case in that the reference period is determined directly by the respondent rather than indirectly according to when contact is made by the interviewer.

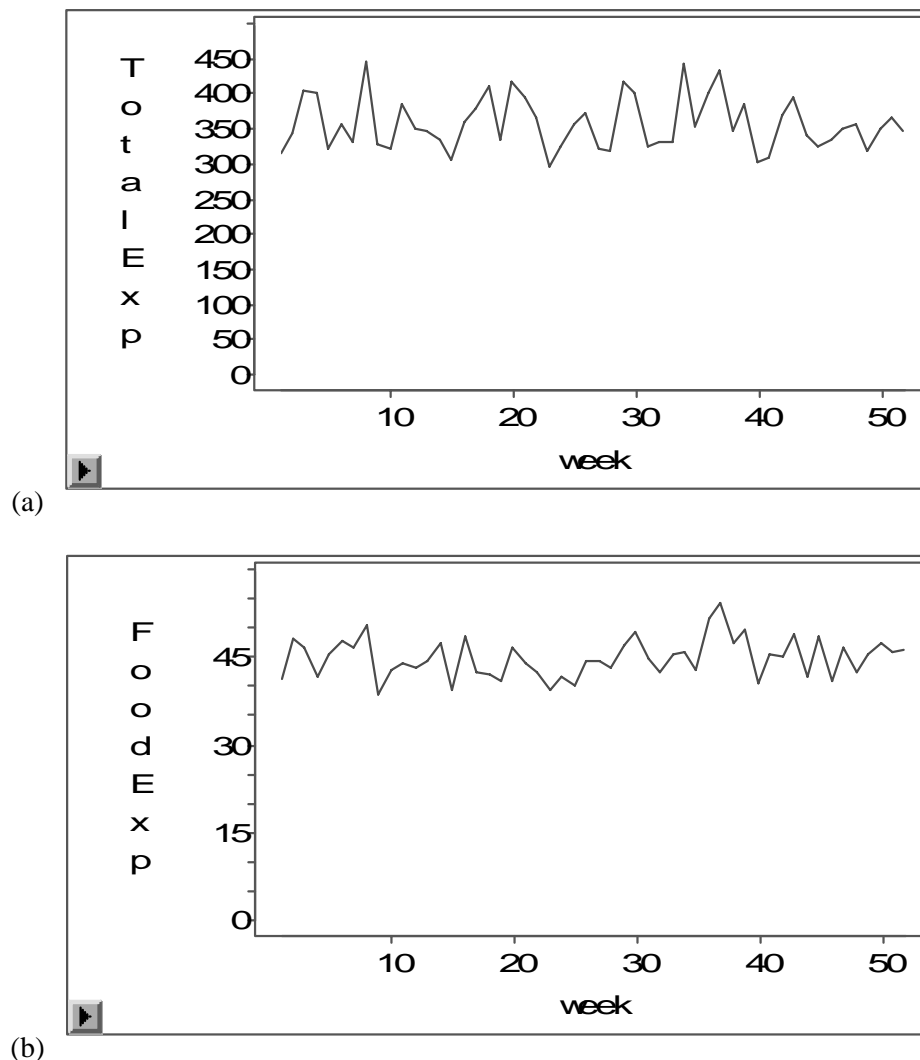
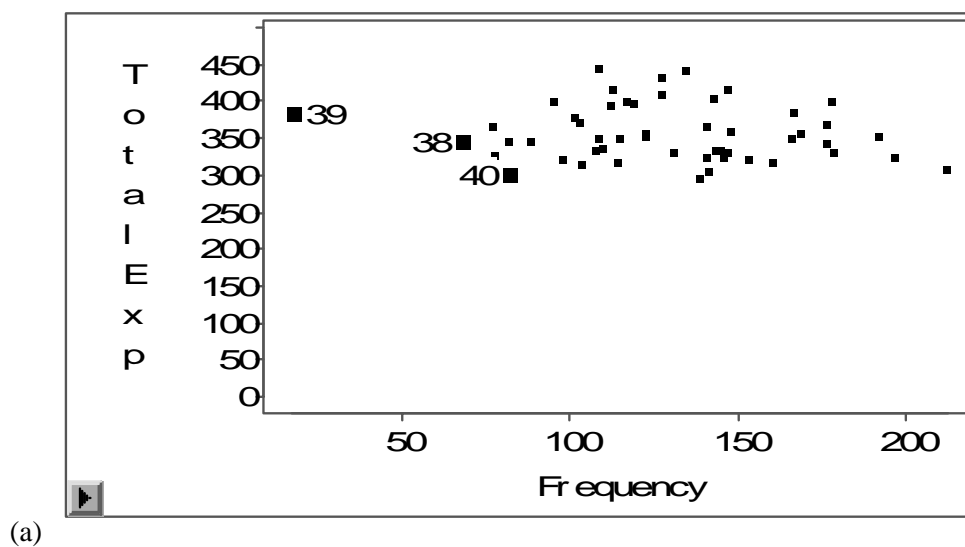


Fig. 3: Pattern of (a) average total expenditure and (b) average food expenditure by week of diary start date for the EFS in 2005-6. The data are unweighted.

### Adjustments for variable reference periods

The current practice on household surveys such as the EFS is to ignore the variability in the weeks and to use the data as if they had equal coverage of the weeks in the year. As long as there is no large week to week variation in the variables being measured, this is unlikely to introduce a significant bias and affect the accuracy (non-sampling error). However in a measure of household expenditure, it might be expected that there are extreme differences around Christmas, particularly for foodstuffs, with substantial expenditure in weeks leading to Christmas, falling to almost nothing in the week of Christmas itself (expenditures on other items may well be counterbalanced by the 'January sales'). We can examine this by looking at the pattern of average expenditure by diary start week (Fig. 3). Although diaries started in the Christmas weeks are characterised by low sales, they are within the range of normal variation over the rest of the year.

We can also examine a scatterplot of the number of achieved diary cases against average expenditure (Fig. 4), which shows a reasonable scatter. Although the weeks near Christmas (highlighted) have low numbers of cases, they do not seem to be atypical in terms of expenditure. Therefore we can be reasonably confident that there is no systematic bias from imbalance in cases by week. Equally there is no relationship between number of cases and expenditure more generally ( $r^2 = 0.02$ ,  $p = 0.29$  for total expenditure,  $r^2 = 0.01$ ,  $p = 0.58$  for food expenditure).



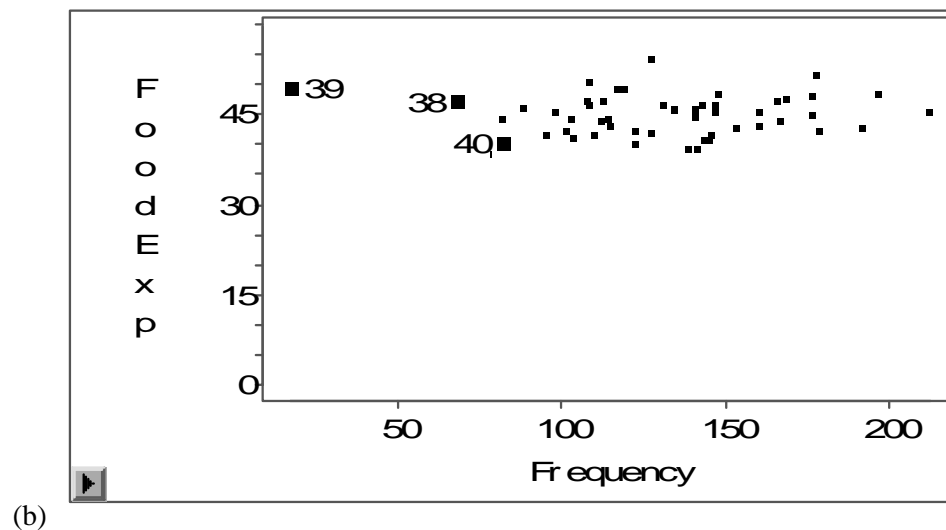


Fig. 4: Plot of (a) average total expenditure and (b) average food expenditure against total number of cases with a start date in each week for the EFS in 2005-6. The data are unweighted. Weeks 38, 39 and 40, which approximately correspond to the Christmas/New Year period, are highlighted.

If a difference had been detected, one solution might be to introduce an additional weight during estimation, to give each week equal weight in the estimates. However this has a disadvantage in that it would increase the variability of the weights and hence the sampling variability of the estimates in order to correct for bias; specific numeric research would be needed to work out which approach gave the best accuracy for a particular survey.

There are several possibilities for adjustment (or not) of the respondent-determined periods. In the Annual Business Inquiry (ABI), businesses are allowed to respond for any 12-month period beginning between 6 April in year  $n$  and 5 April in year  $n+1$ . As long as the response covers a twelve-month period no adjustment is made to the response, and the various different definitions of a year are used as if they are consistent. Several exercises to calculate the average year-end of the responses have shown that it is very close to the calendar year (for example Williams & Richardson 1999); many businesses report for a calendar year, and there are relatively more reporting with year-ends at the end of the financial year which balances out the mismatch between the three months available after the calendar year end and the nine months available before it (Table 1). In some forerunners of the Annual Business Inquiry, information on capital expenditure was collected for both the financial year and the calendar year to give more accurate calendar year estimates (Smith & Penneck 2007), because capital expenditure is often influenced by a few large payments which makes the conversion from one definition to other more difficult (at worst it risks either capturing the same expenditure in two successive years or missing it altogether).

Month	Percentage of businesses
April 1996	4
May	3
June	5
July	3
August	3

September	8
October	4
November	3
December	43
January 1997	4
February	2
March	19

Table 1: Percentage of production and construction businesses returning Annual Business Inquiry questionnaires with year-ends in the months of 1996-7 (from Williams & Richardson 1999).

For monthly turnover surveys an adjustment is usually made to ensure that the length of the period reported corresponds with the required period. So if a survey such as the Monthly Production Inquiry uses calendar months, but a respondent furnishes a four-week period as an approximation, then the value for the four-week period is increased in proportion to the number of working days which should be included. In most cases the days are taken as interchangeable. However, in the case of retail sales there is a very strong within-week seasonality (more sales on Fridays and weekends), and in these cases account is taken of the days of the week which are missing or additionally included relative to the reference period. It would be possible to use this approach in any of the surveys which show seasonality (including purely monthly seasonality), but in most cases the additional *relevance* is outweighed by the decreased *accuracy* (from modelling error) in making an adjustment

### 3. Continuous or discontinuous?

The UK (and in fact many European countries) take the approach that a survey should measure its variables as a sum or an average across the reference period. In the case of sums there is little to debate – total sales in a period would be poorly approximated by sales in some sub-period. However, for stocks (such as number of employees, inventories) or states (such as employed/unemployed/inactive) there is a choice between taking the value at one point in the period and taking the average over the whole period. For example, in the US the Current Population Survey measures employment status in the calendar week (Sunday to Saturday) which includes the 12th day of the month<sup>1</sup>. In the UK, the LFS has continuous interviewing throughout the four (or five) weeks which represent that month. As long as the time chosen for a point-in-time estimate is central in the month, then the two series are centred in approximately the same place. In fact in times of stability of the variable being measured (level, or stable increase or decrease) the results will be very similar. However, just at the point where most interest is focussed, the turning point, there is a risk that the point-in-time estimate will delay the earliest signs of that change.

The choice of a point-in-time or a continuous reference period has an impact on the perceived timeliness of the survey. A point-in-time survey collecting information for the 12<sup>th</sup> of the month will have collected its responses approximately two and a half weeks earlier than a continuous survey covering the same period (all other things being equal). Therefore the

<sup>1</sup> [http://www.bls.gov/cps/cps\\_over.htm#reference](http://www.bls.gov/cps/cps_over.htm#reference)

January results will be available roughly two weeks earlier too. This is important as most measures of timeliness count from the nominal period, that is from the end of January in this example, and not from the end of the reference period.

This is one of the reasons why many US series seem to be produced more quickly than their European counterparts (Öberg and Nanopoulos 2001). In the best examples of both types of survey it is not a property of the efficiency of the National Statistical Institute (NSI) in producing numbers, just a difference in the end of the reference period in relation to the start of counting for the measure of timeliness.

One particularly interesting contrast between point-in-time and continuous collection is between the Producer Price Index (PPI) and the Consumer Price Index (CPI) in the UK. The PPI asks for the average price during a month, whereas the CPI prices are collected directly from retail outlets on a Tuesday near the middle of the month (with some collections on the preceding or following day for a few items whose prices do not fluctuate daily). As well as the differences described above over the timing, there is likely to be some measurement error in the PPI; many businesses will respond with the price of the requested item at the time the questionnaire arrives, and will not consider whether a price change has happened near this date which should be factored in. Nevertheless the effect on the final series should be negligible.

#### 4. Moving averages and rolling estimates

It is becoming commoner in surveys to use information from multiple periods to increase the sample size of surveys, in order for particular accuracy requirements to be met for small populations, or with smaller sample sizes than would otherwise be practical. A particular case of this is the UK Labour Force Survey, which has continuous interviewing which repeats quarterly (with a quarterly reinterview pattern). This is arranged so that over any quarter the LFS sample is an unclustered sample representative of the whole of Great Britain (GB) (with some minor exceptions which will not be discussed here). For periods not composed as the sum of whole quarters, however, the survey is clustered; specifically, for periods shorter than a quarter the survey is clustered and does not cover each part of GB.

This process is translated into a monthly output representative of the whole of GB by using “rolling estimates”, composed of the latest three months of interview data, issued monthly. So the estimate for January is based on data for November, December and January, and so on. This gives an example of labelling akin to the difference between point-in-time and average surveys – the centre of the January LFS estimate actually lies in mid-December (considering it as a moving average). Indeed there is a difference of interpretation between the ONS and Eurostat as to whether the January LFS estimate really corresponds to January. This is an aspect of relevance – whether the estimate corresponds accurately with what it is purported to measure. However, relative to a design which would produce monthly estimates from only data collected in that month, the cost is much smaller (Steel 1996 estimated that a true monthly design would cost 2.3 times as much). Note that cost is not one of the European quality measures, although it is clearly an item of considerable importance to NSIs.

The same situation occurs in other surveys where several years data are pooled to ensure that there is adequate information for making small domain estimates (for example the General

Household Survey, detailed Scottish estimates from the Expenditure and Food Survey, and some tables in the Family Resources Survey).

## 5. Discussion

### Timeliness – relevance trade-offs

One aspect of *relevance* is to ensure that the period which is being measured is the one which is needed. The examples given above demonstrate that in some cases there is a mismatch – where the survey measures a period which is offset from the one required. This is not often explicitly accounted for when making decisions based on survey information, or when comparing statistical outputs in different countries (see Öberg & Nanopoulos 2001). Instead the *timeliness*, the number of days after the end of the nominal reference period to which an estimate relates, is accorded much greater importance. Unfortunately timeliness is very easily quantified, but relevance is very difficult to quantify, so surveys which are very timely are often seen to be better than those which are less timely but might have high relevance. For example it would be possible to make a survey apparently very timely by defining February to begin on 5 January and end on 4 February, thereby producing results on, say 1 March, the day after the end of the reference period! Except that the reference period would not in reality be ‘February’ in this case, and that the relevance for timely economic decision-making would be reduced. If some intervention were introduced on 1 March in response to the ‘February’ estimates, it would have practically no opportunity to influence the ‘March’ estimate as presumably only 2-4 March would be affected by the intervention, leaving 5 February to 1 March unaffected.

Therefore keeping the relevance high avoids the difficulty of conceptual misunderstandings and is to be recommended. Sometimes, however, there are very good resource-based reasons for using variations in the way periods are defined, and in these cases the impact on relevance should be well documented in the survey results.

### Accuracy

Although we have focussed on timeliness and relevance, the different subcomponents of accuracy are interrelated with them, and there are trade-offs between accuracy and relevance, because adjustments made to improve relevance often involve assumptions which decrease accuracy. There is a similar trade-off between *timeliness* (measured from the actual reference period, not the nominal one) and *accuracy* because the response increases gradually as additional time is allowed for follow-up, and therefore reliance on models for imputation, estimation and so on is reduced.

## 6. Conclusions

It is important to make an explicit assessment of the quality trade-offs in a survey, because a gradual evolution of methods (a typical development approach in most surveys) is unlikely to lead to an optimum solution. Consistency can be very helpful for users, but in general improving the quality balance should be more important than perpetuating consistency of a poor quality series. The quality aspects of a survey should be measured and documented, and

users should take account of this information when using the survey results to make decisions and formulate policies.

## References

- Eurostat (2003) Methodological documents - definition of quality in statistics (Working group "Assessment of quality in Statistics" document Eurostat/A4/Quality/03/General/Definition). Eurostat, Luxemburg.
- Öberg S and Nanopoulos P (2001) Report of the task force on benchmarking in infra-annual economic statistics to the SPC (CPS 2001/42/8/EN). Eurostat, Luxemburg.
- ONS (2005) Office for National Statistics compliance plan 2005-6 to 2007-8. HMSO, London.
- Portanti M (2007) Expenditure and Food Survey, Technical Report for the Year: April 2005 to March 2006. ONS, Newport.
- Smith P and Penneck S (2007) 100 years of the Census of Production in the UK. ONS, Newport  
([statistics.gov.uk/about/methodology\\_by\\_theme/downloads/CoP100yearsinthelUK.pdf](http://statistics.gov.uk/about/methodology_by_theme/downloads/CoP100yearsinthelUK.pdf)).
- Steel D (1996) Producing estimates of unemployment and employment according to the International Labour Office definition (with discussion). *Journal of the Royal Statistical Society, Series A*, vol. 160, 5-46.
- Williams M and Richardson R (1999) Model quality report of the 1996 UK annual production and construction inquiries. Pp 31-53 in Davies, P. and Smith, P. (eds) *Model Quality Reports in Business Statistics*, vol III. ONS, UK.

## About the Authors

Paul Smith, Charles Lound and Pete Brodie all work in the Survey Methodology & Quality Division of the Methodology Directorate of the Office for National Statistics. Between them they have extensive experience of both household and business survey methods. They all think survey methodology is fun (most of the time)!

# The Difficulty of Understanding Social Survey Questionnaires from the Published Documentation

N. Graham Hughes

## Abstract

The use of computer interviewing programs has contributed to a significant increase in the complexity of survey instruments. Research has shown that survey data can sometimes be affected by related matters arising during interviews. Current methods of documenting the interviewing programs show a lack of standards or conventions, making it difficult for secondary analysts to identify potential interview context effects. The solutions proposed in recent ambitious documentation projects have yet to be implemented widely. In the meantime this paper suggests some simple presentation conventions to be applied whenever a text version of a questionnaire is being prepared, in the hope that these might enable more readily understandable documentation to be published.

## Keywords

Questionnaire documentation; routing instructions; control checks; presentation; conventions.

## 1. Why the documentation Matters

It is a fundamental assumption of quantitative social research that the answers provided by respondents in relation to each particular variable are equivalent to each other: without this there would be no meaning attachable to the mean, no standard to explain the standard deviation. Social scientists need to be able to conclude that if, for example, 72.6% of respondents agreed with an attitude suggested in a question then that explains something about the population from which those respondents were sampled. It has long been known that this is not possible if each respondent is asked a different question and so in the past great efforts were made to ensure that the same question was asked in the same way of every respondent being interviewed (Groves et al. 2004). This paper looks at the consequences of asking any one question in a wide variety of circumstances on the homogeneity of the data thus collected.

This concern has arisen through the author's work as the Content Manager at the ESRC-funded Question Bank (Qb). The Qb contains a great many survey questionnaires and makes them readily available to other researchers over the internet. Research using the questionnaires in the Qb has shown that no two questionnaires use the same stylistic conventions in their presentation of the various elements that they contain. This makes it more difficult than necessary to interpret and compare extracts from different surveys. Further research has shown that human error, in the form of inconsistencies, is apparent within the



texts. This puts into question the role of a published survey questionnaire as a “truthful” document.

Why does this matter? Well, in the years before Computer Assisted Interviewing (CAI) became widely available a survey instrument frequently took the form of a printed document given to the interviewer who, after careful training, would be expected to follow the fixed instructions contained therein. As Kent & Willenborg put it “... the questionnaire form was the documentation” (1997), so anyone else subsequently could examine the printed document and have a reasonable expectation of understanding what should have happened during the interviews. There was an effective restriction on the extent of complexity in such documents because of the need for interviewers to apply them consistently in the interview situation. The advent of CAI has removed some of those constraints and now, because the interviewer only needs to see one question at a time on the screen, it is possible to include far more complex routing instructions, which the computer can be expected to apply reliably. However at the same time, because the instrument is encoded in the software, there is no document simply available for anyone else to examine in order to learn what happened in the interview. The questionnaires that are published in technical reports and with datasets are produced through manual editing of the CAI program (Kelly, 2000). So as the instruments have become more complex they have also acquired the need for additional interpretation before they can be understood.

The situation now exists where there can be so many different possible routes through a single questionnaire that it may be possible for some respondents to have effectively unique interviews whereby nobody else answers precisely the same set of questions. In these circumstances, can the analyst still be sure that the answer data collected for any single question are sufficiently equivalent, in other words sufficiently homogeneous, to be combined and used as a simple data set? The work of Schuman and Presser (1996) and Tourangeau, Rips and Rasinski (2000) gives cause for concern over context effects in some circumstances. If there are doubts about this, is it possible for analysts to establish clearly which other questions asked earlier in the interview could have made a difference to some respondents’ interpretation of a later question? It is for these reasons that secondary analysts need clear documentation of what has occurred.

The TADEQ or Tool for the Analysis and Documentation of Electronic Questionnaires (Bethlehem and Hundepool, 2004) project attempted to develop software which would allow reporting of CAI programs in standard formats. Initially TADEQ focussed on BLAISE as the source CAI program but there was an intention that it would eventually be compatible with other major programs. For reasons that are not clear, the TADEQ project does not seem to have been realised and analysts are still seeing the use of single linear text documents as the primary explanation of CAI questionnaires.

## 2. Elements Typically Included in Questionnaire Documentation

### Common elements

In order to examine these issues this research has attempted to identify the set of basic elements that make up the current documentation of a CAI questionnaire. The following (brief) list uses unambiguous terminology to highlight their apparent uses in examples of

documentation drawn from the Qb. The examples can be seen by following the web links listed in the Appendix to this paper.

At the heart of the issue is the question text which is rarely made prominent in the documentation but is usually identifiable from its content and language. Example 1 shows a presentation where the question text is difficult to find. The question is almost invariably followed by the set of response categories pre-coded for analysis. Example 2 shows a clearer presentation of question text and response set, but the response codes to the right are not easily aligned with their meanings. Example 3 shows another way of presenting the response set. The question text is usually preceded by the variable name, a mixture of code and abbreviation, as in Example 4 where a lengthy presentation of variable names, section number and code, and question number and name, uses a mixture of capitals and lower case. Example 5 is rather simpler and just shows the variable name in bold face.

Frequently there will be a clue to the type of variable or question somewhere in the vicinity; for example some questions will only accept a single answer, some will accept any number of answers at the discretion of the respondent, some will be open questions with a space constraint on the number of characters of free text, and others will be calculated by the program and not actually asked aloud. These types (single, multi, open, computed) have to be inferred by the reader from instructions apparently given to the interviewer in the text. In Example 6 note “CODE ALL THAT APPLY – Multicoded (Maximum of 11 codes)”. There are other sorts of instruction for the interviewers as well, for example telling him/her to use a particular show-card, advising him/her when to prompt the respondent for more detail, highlighting some questions as very important for the sake of correct routing later, and sometimes telling him/her when a ‘Don’t know’ or ‘refusal’ response is not acceptable. Example 7 includes several prompts finishing with “EXCLUDE BBC WEBSITE”).

### Routing, control checks and text-fills

Whilst the elements described so far seem to arrive in the published documentation as direct representations of text used in the program itself, the treatment of routing instructions and control checks is rather different. With these items the amount of interpretation and editing by those who prepare the documentation is much greater.

For routing (or, as it is sometimes called, “skip patterns”) several important decisions have to be taken as the documentation is prepared. These are

- i) whether to use “Go To” or “If...Then...Else” logical instructions,
- ii) how to represent nested conditions,
- iii) whether to show the conditional logic in plain language or in the algebra of variable names, values and mathematical symbols, and
- iv) how to represent loops, tables, parallel modules and sub-samples.

Examples 8 to 12 illustrate different approaches to some of these presentation decisions.

Control checks may be defined as “hard”, “soft” or “range” checks. They are used to improve the accuracy of the data collection process by identifying possible errors or inconsistencies in the recorded responses and either forcing or facilitating corrections to be made. These present the documentation editors with similar problems to those of routing over how to show the logical conditions that trigger these interruptions to the interview process. Example 13 shows one presentation of control checks. These may not always be fully reported in published documentation.

Another common feature of CAPI interviewing programs which also utilise conditional logic are text-fills or data-fills. These bits of programming cleverness can appear within the text of questions, interviewer instructions, or error messages arising from control checks. Their purpose is to make the interview flow more smoothly as a conversation by using information from earlier questions to personalise the language of subsequent items. Example 14 shows how recording the logical conditions controlling the text-fills within a question text can make the resulting document very difficult to comprehend.

The advantages for the data collection process of using sophisticated routing and control checks relate to shortening the interview by eliminating irrelevant questions and improving the accuracy of the data by reducing incorrect responses. The disadvantages are less apparent in terms of the introduction of variations in the experiences of the different respondents. For the reader of the questionnaire documentation, these elements bring additional challenges. Is it possible to work out what happened in a specific interview, or in the generality of interviews? Can context effects be identified that may have affected some respondents as they answered questions critical to the data being studied? Can it be deduced why some respondents appear to have been asked a particular question while others were not? Can analysts decide whether data collected in two different surveys, but using the same question, is really comparable or not? To deal with these challenges, clear and accurate documentation is needed.

Hopefully the range of examples used to illustrate these elements has begun to create an impression of the confusing variety of styles and conventions that have been used in these questionnaire documents in recent years. This summary has been offered, not as a criticism of the survey agencies that have prepared these materials but as an illustration of what is to be found when their forms rather than their contents are studied.

### 3. Proposal for Standard Conventions

Whilst the ambition of projects like TADEQ and DDI is admirable, it does not seem likely that they are going to be implemented quickly or universally. In the meantime, until they are widely implemented, it will almost certainly be necessary to continue to use manually edited linear documents. The modest proposal here is to suggest a standardised set of conventions that could be used by almost all questionnaire documentation teams whenever they are asked to prepare a text version of a questionnaire, without requiring any new technology or software. The simple goal is to see common ground across a whole range of survey questionnaires so that users can devote their energy to understanding the events of whatever survey they are studying rather than having to first figure out how to read the document. Such a step might mark significant progress towards future automation of the process by which survey metadata is retrieved from these documents.

A wish-list for such conventions would be as follows:

1. Use a fixed order or sequence for the basic elements within each question 'event' in the documentation. (E.g. Question number, variable name, variable type, routing applicable, question text, calculation rule, response set, show-card used, interviewer permitted prompts, range checks, hard checks, soft checks, explanatory material.)

2. Use standard font characteristics to help identify each element above. (E.g. shape of brackets, italics, capitals, bold-face and inverted commas.) This need not restrict the actual font style so that surveys could maintain some individuality of appearance.
3. Use a common structure for routing instructions. Probably the ideal from a user's perspective would be the "Ask if" format with algebraic style code of both previous variables and relevant responses followed by full plain English text, placed in every question event.
4. Avoid using short-cuts such as tables and common response sets referenced by asterisk. Instead try to show the full questions in standard form.
5. Show the conditional logic for all checks in algebraic style code form only (for brevity) with the relevant error message in full text.
6. If computations, conditional logic or routing instructions use data from a source outside the current questionnaire (such as a previous interview or wave of the survey) list all these variables and their associated response sets in a clearly identified section near the start of the document.

### Illustrations and examples

In order to add some substance to these proposals, here is a suggestion for a basic set of stylistic conventions which might be useful as the starting point for a discussion.

Suggested questionnaire style example:

Question # **Variable Name** { Variable type: single/multicode/text/calculated etc }

ASK IF: [Variable1 = # AND Variable2 <> #], [Plain language statement of routing logic]

**"Question text, including (data-fill alternative1/datafill alternative2) and ending in a question mark?"** (Data-fills determined by Variable#)

Or put calculation rules here if it is a derived variable.

1. First response
2. Second response
3. Third response etc.

*SHOW-CARD # / READ OUT RESPONSES LIST*

*Interviewer instructions, in italics, "with any text to be spoken by the interviewer (including data-fills) in bold and between inverted commas".* (Data-fills determined by Variable# & Variable#).

CONTROL CHECKS:

Type: Hard/Soft/Range

[Logical condition as algebraic expression only]

*Instructions for interviewer, including "any text to be spoken"*

Explanatory notes.

In this suggestion bold font is reserved for the variable name and spoken text only, the use of capital letters is kept to a minimum, logical expressions are enclosed within square brackets, text-fills and data-fills are enclosed within curved brackets, and instructions for interviewers are shown in italics. The routing is shown in both algebraic form and plain language, and is

placed after the variable name but before the question text, in order to be as unambiguous as possible. Not shown, but also suggested, is a separating line between each question to indicate when a fresh screen would have been shown to the interviewer.

Below are two illustrations of this style suggestion as applied to existing questionnaire material. The first is from the British Social Attitudes Survey 2005, shown as Example 6. Here the proposed presentation is shown first with the current version as taken from the Qb materials beneath it. Because this example is of a multicoded variable it is suggested that the variable name associated with each response is placed to the left of the response code. The explanatory notes are derived from information referenced by the \$ <sup>1</sup> in the original (an introductory note and footnotes).

#### **Reworked Example 6 – British Social Attitudes Survey 2005 – Main Questionnaire**

Q565-575    **WDisFW**                      {Multicoded – maximum 11 codes}

ASK IF: [Version = A, C or D], [All of sub-samples A, C and D]

**“People have different ideas about what it means to be disabled. Which of the people on this card would you think of as a disabled person?”**

<b>WDisNone</b>	0	None of these
<b>WDisArth</b>	1	A person with severe arthritis
<b>WDisAIDS</b>	2	A person who has HIV/AIDS
<b>WDisSchi</b>	3	A person who has a diagnosis of schizophrenia
<b>WDisDepr</b>	4	A person who has a diagnosis of severe depression
<b>WDisDown</b>	5	A person who has Down’s Syndrome
<b>WDisCanc</b>	6	A person who has cancer
<b>WDisOldH</b>	7	An older person who cannot hear without a hearing aid
<b>WDisBlin</b>	8	A blind person
<b>WDisWhlc</b>	9	A person who uses a wheelchair most of the time
<b>WDisBrok</b>	10	A person with a broken leg, using crutches while it heals
<b>WDisFacD</b>	11	A person with a severe facial disfigurement
	97	All of these

*SHOWCARD E1*

*Interviewer: code all that apply and probe: “Which others?”*

Explanation: Variable WDisFW does not appear in SPSS file, the 12 variables derived from it, and shown above, do appear with values 0=not mentioned, 1=mentioned.

**Example 6 - British Social Attitudes Survey 2005 – Main Questionnaire – Disability**

VERSIONS A, C AND D: ASK ALL

Q565- [WDisFW] \$ -  
 Q575 CARD E1

People have different ideas about what it means to be disabled.  
 Which of the people on this card would you think of as a disabled person?  
 PROBE: Which others?  
 CODE ALL THAT APPLY  
 Multicoded (Maximum of 11 codes)

0	(None of these)	[WDisNone]
1	A person with severe arthritis	[WDisArth]
2	A person who has HIV/AIDS	[WDisAIDS]
3	A person who has a diagnosis of schizophrenia	[WDisSchiz]
4	A person who has a diagnosis of severe depression	[WDisDepr]
5	A person who has Down's Syndrome	[WDisDown]
6	A person who has cancer	[WDisCanc]
7	An older person who cannot hear without a hearing aid	[WDisOldH]
8	A blind person	[WDisBlin]
9	A person who uses a wheelchair most of the time	[WDisWhlc]
10	A person with a broken leg, using crutches while it heals	[WDisBrok]
11	A person with a severe facial disfigurement	[WDisFacD]
97	(All of these)	

Next is shown a reworking of Example 14 to illustrate how routing and text-fills might be clarified.

**Reworked Example 14 – Offending, Crime and Justice Survey 2004 – CAPI Questionnaire**

**V1vehS** {single code}

ASK IF: [V1veh = 1], [Someone living here has owned or regularly used a car, van, motorbike or other motor vehicle in last 12 months]

**“Since the first of (month) 2003, (has anyone who lives here had their / have you had your / have you or anyone who lives here had their) motor vehicle stolen or driven away without permission, even if (they / you) later got it back?”** (Data fills determined by month of interview, respondent aged 16 or more, number of people in household aged 16 or more).

1. Yes
2. No
3. Don't know
4. Refused

**Example 14 - Offending, Crime and Justice Survey 2004 – CAPI Questionnaire – Victimization**

**V1vehS** [ASK if V1veh=1]  
 Since the first of [MONTH] 2003, [IF L1age<16: has anyone who lives here had their/ IF L1age>15 AND ONLY ONE PERSON 16+ IN HOUSEHOLD: have you had your/ IF L1age>15 AND 2 OR MORE PERSONS 16+ IN HOUSEHOLD: have you or anyone who lives here had their] motor vehicle STOLEN OR DRIVEN AWAY WITHOUT PERMISSION, even if [they/ IF L1age>15 AND ONLY 1 PERSON 16+ IN HOUSEHOLD: you] later got it back?

1. Yes
2. No
3. Don't Know
4. Refused

Documents using these conventions could still be prepared using standard word-processing software and then be converted to PDF for publication on the internet. However it would be quite possible subsequently to convert such documents to XML and apply standard content tags in order to make them more flexibly useful to users with appropriate software. This can be seen as a half-way step towards the DDI model. In the meantime the use of strong conventions in this way should help analysts and researchers to identify more clearly just how the data they are using was originally collected.

## Appendix: Examples – Web-Links and Screen Shots

Examples format: - Survey name & year – Questionnaire PDF name – Bookmarked link. (The web-link will take you to the start of the PDF document and you should then click on the relevant bookmark, or use the Adobe page number system, to locate the specific page referred to in this paper.)

1. National Travel Survey 2004 – Individual Questionnaire – Journey to work  
<http://qb.soc.surrey.ac.uk/surveys/nts/04individual.pdf> (p20/42)
2. Health Education Population Survey 2005 – Main Questionnaire – Alcohol  
<http://qb.soc.surrey.ac.uk/surveys/heps/05mainqheps.pdf> (p25/55)
3. British Crime Survey 2004/5 – Main Questionnaire – Main questionnaire  
<http://qb.soc.surrey.ac.uk/surveys/bcs/04mainqbcs.pdf> (p9/118)
4. Family Expenditure Survey 1999 – Household Questionnaire Part5 – Purchase of Furniture  
<http://qb.soc.surrey.ac.uk/surveys/fes/fes99hque5.pdf> (p24/30)
5. Health Survey for England 2004 – Household Questionnaire – Accommodation & tenure  
<http://qb.soc.surrey.ac.uk/surveys/hse/04hqhse.pdf> (9/16)
6. British Social Attitudes Survey 2005 – Main Questionnaire – Disability  
<http://qb.soc.surrey.ac.uk/surveys/bsa/05mainqbsa.pdf> (p57/174)
7. British Social Attitudes Survey 2005 – Main Questionnaire – Newspaper readership  
<http://qb.soc.surrey.ac.uk/surveys/bsa/05mainqbsa.pdf> (p17/174)
8. Continuous Household Survey (NI) 2004 – Household Questionnaire – Tenure  
<http://qb.soc.surrey.ac.uk/surveys/chs/04housechs.pdf> (p7/19)
9. ONS Omnibus Survey 2003/4 – Classificatory Questionnaire – Paid work  
<http://qb.soc.surrey.ac.uk/surveys/omnibus/OMNClass0304.pdf> (p15/26)
10. People Families & Communities Survey 2005 – Questionnaire – Illness or Disability  
<http://qb.soc.surrey.ac.uk/surveys/citizenship/05questcs.pdf> (p81/106)
11. English Longitudinal Study of Ageing 2004 – Private Questionnaire – Expectations  
<http://qb.soc.surrey.ac.uk/surveys/elsa/04w2privatelsa.pdf> (p8/40)
12. Scottish Health Survey 1998 – Individual Questionnaire Part A – General Health module  
[http://qb.soc.surrey.ac.uk/surveys/ShealthS/SHS98\\_IQUIREa.pdf](http://qb.soc.surrey.ac.uk/surveys/ShealthS/SHS98_IQUIREa.pdf) (p3/29)
13. Families and Children Study 2004 (Wave 6) – Main Questionnaire – 7. Housing  
<http://qb.soc.surrey.ac.uk/surveys/facs/04mainquest.pdf> (p46/174)
14. Offending, Crime and Justice Survey 2004 – CAPI Questionnaire – Victimisation  
<http://qb.soc.surrey.ac.uk/surveys/cjs/04capicjs.pdf> (p48/66)



## References

- Bethlehem J and Hundepool A** (2004) TADEQ: A Tool for the Documentation and Analysis of Electronic Questionnaires. *Journal of Official Statistics*, vol.20, No.2, 233-264
- Groves RM, Fowler FJ, Couper M, Lepkowski JM, Singer E and Tourangeau R** (2004) *Survey Methodology*. Wiley, New Jersey.
- Kelly M** (2000) What users want from a tool for analysing and documenting electronic questionnaires: the user requirements for the TADEQ project. *Blaise Users Group conference paper*.
- Kent J-P and Willenborg L** (1997) Documenting questionnaires. *Research Paper No. 9708*, Voorburg: Department of Statistical Methods, Statistics Netherlands.
- Schuman H and Presser S** (1996) *Questions and Answers in Attitude Surveys*. Sage, Thousand Oaks CA.
- Tourangeau R, Rips LJ and Rasinski K** (2000) *The Psychology of Survey Response*. Cambridge University Press, Cambridge.

## About the Author

**Graham Hughes** is the content manager of the ESRC-funded Question Bank, based at the University of Surrey. Since taking up the post in October 2005 he has added a considerable volume of questionnaires and related materials to that freely accessible web-site, all in PDF files. Prior to this appointment he completed an MSc degree at Surrey in Social Research Methods as a career change after working as a Chartered Accountant for the previous 23 years. He has worked with computers since 1984, making the principle of 'one PC per desk' a fundamental basis of the accountancy practice that he co-founded in 1987, and exploiting a variety of opportunities for installation of and training on software for his clients. He brings a combination of maturity and freshness to his current role. He can be contacted at Question Bank, Department of Sociology, University of Surrey, Guildford, Surrey, GU2 7XH; tel. 01483 682794; fax. 01483 689551; e-mail n.hughes@surrey.ac.uk.

# Whither Statistical Metadata?

Andrew Westlake

## Abstract

The use of metadata as description of statistical datasets has been with us for some time, and the value of this approach is demonstrated by the success of initiatives for metadata standards, such as triple-s, SPSS Dimensions and DDI. The need to support discovery of resources and services over the internet has widened interest in and use of the term metadata, focussed by initiatives such as the Dublin Core, the (derived) UK Government Metadata Standard (e-GMS) and ISO-11179.

These standards and structures are now being influenced by more formal ideas from IT systems design. The recent proposals from SDMX and (version 3 of) DDI are strongly influenced by object-oriented design principles, using UML as the specification medium. This extends the specification beyond the structure of the information represented to include the formal representation of the semantics and functionality that are needed in order to use the metadata.

These proposals cover information about the data capture process (including sample and instrument design), about the structure and location of datasets (including coding), and the exchange of data (with metadata) in micro or macro form. The Opus Project has attempted to go further than this and represent the statistical models used in the analysis of data, and thus to provide a link back from conclusions and results based on statistical information (evidence) to the assumptions and data used.

## Keywords

Statistical meta-data, modelling, UML, system design, standards

## 1. Whither Statistical Metadata?

### Applications with Structure

The future of Statistical Metadata (and metadata in general) does not lie in further amateur attempts to design XML data structures for some particular problem. Rather, it lies in a professional approach to system design and software development, producing metadata-aware applications that can use, capture, present, manipulate and re-use metadata and the associated resources which it describes.

We have had innumerable proposals for metadata systems and standards over the past decade, but few have produced anything usable or lasting. The few that have been successful have

either been extremely simple in structure and concept – such as triple-s[1] for data and label exchange, or the Dublin Core[2] for resource discovery – or have involved very significant investment – such as the proprietary Dimensions system from SPSS, complete with an API and its own version of Visual Basic.

Other initiatives have produced valuable insights into the requirements for metadata in statistical applications, but little in the way of usable functionality. For example, in the field of questionnaire and instrument design, the IQML project[4] and the QEDML system[5] (both presented, along with several others, at previous ASC conferences) provide many useful ideas about the representation of questionnaires and the data capture process, but neither has produced a direct legacy – IQML came to a natural end, so that its influence was always going to be indirect, and the QEDML web site has not been updated since 2004, in spite of very considerable enthusiasm about its potential.

This is not the right point for an exhaustive discussion of such initiatives, but there are many that could be mentioned, often with considerable merit but with only indirect influence on actual metadata practice. A few exceptions are discussed later.

### Quality and Use

The purpose of metadata is to support the use of actual statistical data, or some other resource. For metadata to be used it must be both useful and useable. To be useful it must be pertinent to the task for which it is to be used, and must also be of adequate quality for this task. Where metadata (or any other resource) is not used there is no incentive for those collecting it to maintain quality, which in turn makes it less likely that it will be used, a vicious circle. We need to break that circle.

But usefulness is not enough to ensure usability. Reading extensive documentation sometimes provides some information but is rarely the best way to use metadata to support some task. This is particularly true when the metadata is presented as a raw XML document. Instead we need metadata-enabled applications which make use of metadata to simplify or enrich the functionality of the primary task of the user.

In the same way, metadata capture should not in general be a manual task. Metadata is rarely (perhaps never) new information, it is just ancillary information that an adjunct to some other task. So the process of performing the initial task should automatically capture the related metadata. For example, the design of a questionnaire (with all its wording, conditions, filters, skips, fill-ins, etc.), in a suitable design program, already contains all the metadata about the questionnaire structure that might be needed to inform later analysis of data collected from that questionnaire. We just need to expose that information in a way that can be used in other contexts. And the exposure makes it possible to enhance the design process as well, for example by importing questionnaire fragments from a Library or another questionnaire (perhaps designed in a different application).

Similarly, sample design specifications, or information about data transformations or adjustments can be captured automatically for subsequent use. Quality is maintained through such automatic capture, and, where additional manual input is needed, the quantity (the additional workload) is minimised, relevance should be clear, so quality should be high.

## Design by Structure

The idea of metadata aware software is not new. Virtually all survey and statistical packages from SPSS onwards include a specific data dictionary (or codebook) which is used internally, and can sometimes be accessed externally. In the early 1980's the World Fertility Survey was using processing facilities built around a data dictionary system (as were others), and a comprehensive description of such an integrated metadata-aware statistical processing system was presented by Jean-Pierre Kent[6] in an SSDBM paper of 1997.

Many designers of statistical and related systems have internal components which involve explicit representation of metadata, and its use within the system to enhance and facilitate the user experience. There is no direct incentive for developers to invest effort in the design of interfaces to expose this information to other applications, whether complementary or competitive. Such pressure has come from users or independent developers (often academics), and the simplest solution has been seen as the design of an exchange file format, for which import and export facilities can easily be appended to existing software.

The problem with this approach is that different developers use different concepts within their systems, or interpret and implement the same concepts in different ways and with different details and assumptions. The designers of the triple-s interchange standard recognised this problem and chose to take a highest common factor (HCF) approach. This results in a minimalist standard that is very limited in scope, but is usable in full by all the (initial) participating systems. This achieves the objective of moving datasets between supporting systems, but means that the user may have lost some information that was in the source system, because it is not available or is represented differently in the target one.

The independent approach has been to try to define a separate 'standard' as a super-set of concepts to which all systems can subscribe. Where this is done as a union of facilities from an existing set of systems there is a danger of too much detail. This can lead to 'gothic castles', which contain huge amounts of intricate (and sometimes interesting) detail, but no obvious overall structure. The more favoured approach (which was supported by a number of European 4<sup>th</sup> and 5<sup>th</sup> Framework projects through Eurostat) is to perform a more generic analysis and to identify concepts and principles that need to be represented, then to develop a metadata design based on these, using object concepts of inheritance and specialisation to handle details.

Most initiatives in this area (and not just within statistics) seem to concentrate on defining the data structure needed to store instances of metadata, usually as an XML document – hence the early derogatory remark in this paper. While defining a suitable structure is an essential part of the design process, it is only a step on the way to producing metadata-aware applications, and while XML is an essential tool for the exchange of metadata instances between independent applications, an XML Schema is not an adequate medium in which to specify processing functionality. Many proposals for standards do include some ideas for actual applications, but this is usually seen as a separate phase following after the structural definitions, and in consequence very little gets done.

## Design for Use

The purpose of metadata is to support other tasks. So the design of a metadata system should be related to the design of applications which are metadata-aware. The functionality provided to support the use of the metadata is more important than the structure used to store it. This

leads to the conclusion that metadata design should be seen as an instance of the more general systems design process, and should take advantage and make use of progress that has been made in that area.

A great deal of effort in recent decades in the Computer Science domain has been spent on the development of methodologies to support the development of application software. Most of this has been focussed on the Object-Oriented approach, and within statistics we have the application S (and R) as an example based on these ideas. A number of different methodologies (procedural specification methods for software development) grew from this base, followed in the late '90's by a consolidation and standardisation effort that produced UML, the Unified Modelling Language[7]. UML is now in version 2.

UML is widely supported by tools based on the standard, from initial design through to implementation of applications. It covers all stages and aspects of system design, so that (at least in theory) it is possible to completely specify an application in UML, and then have that design automatically translated into an executable program in a suitable language. In that sense it is language-independent, though it is probably most frequently used with Java.

Designs in UML include data structures (as needed for internal use in applications as well as for exchange), but also cover the behavioural and procedural aspects of applications. Thus, when thinking about the information structures that are needed for a metadata application, it is also possible to specify the behaviour and processes that are expected to be associated with this information. UML specifically supports the idea of abstraction, so it is not necessary to immediately specify all the details of some process: it is sufficient to specify just the outline, with further details supplied later at an implementation stage.

## Conclusion

Metadata needs to be used, and this use takes place in metadata-aware applications. So the design of metadata systems should take place in the context of (at least some of) the applications which will make use of such a metadata resource.

UML is an ideal medium in which to specify the design of such metadata systems, as it allows behaviour and process functionality to be designed, alongside the data structures needed to represent the metadata resource.

A number of recent metadata proposals have been designed in this way, and demonstrate the feasibility of this approach, specifically the SDMX[8] (Statistical Data and Metadata Exchange) standard for the exchange of aggregate data between statistical authorities, and Version 3 of the DDI[9] (Data Documentation Initiative) Codebook standard for documenting data resources.

## Coverage of Statistical Metadata

As we start to think more specifically about the tasks for which statistical metadata can be used, we also need to think more widely about the areas of statistical processing that should be covered.

Statistical metadata started with data documentation in codebooks, containing label information about variables and codes within a single dataset. Extensions to this that have been implemented include multiple datasets, aggregated data, sample design and responsibility links. Specifications for derived variables and complete questionnaire specifications have also been attempted.

As part of the MetaNet[10] project, Froeschl and colleagues[11] produced the first version of UMAS, a Unified MetaInformation Architecture for Statistics, intended to cover all aspects of statistical processing. This contains many good ideas, but is an ambitious project that needs further work.

The Opus Project[12] also extends metadata well beyond data description, though with a more specific objective. It attempts to document statistical analysis processes based on statistical models and represent the statistical models used in the analysis of data, and thus to provide a link back from conclusions and results based on statistical information (evidence) to the assumptions and data used.

## 2. Why Metadata?

What is metadata?

Bo Sundgren claims to have coined the term ‘metadata’ in his PhD thesis of 1973, but the term did not come into general use until considerably later. It was generally used by people working with statistical data by the late ‘80’s, but it was at least another 10 years for the blank expressions to disappear from the faces of people from most other disciplines.

In the Metanet project we developed the following definition of statistical metadata.

*Statistical Metadata is any information that is needed by people or systems to make proper and correct **use** of the real statistical data, in terms of capturing, reading, processing, interpreting, analysing and presenting the information (or any other use). In other words, statistical metadata is anything that might influence or control the way in which the core information is used by people or software.*

It extends from very specific technical information, used, for example, to ensure that a data file is read correctly, or that a category in a summary table has the right label, or that the design of a sample is correctly taken into account when computing a statistical summary, right through to descriptive (intentional) information, for example about why a question was worded in a particular way or why a particular selection criterion was used for a sample.

Thus, metadata includes (but is not limited to) population definitions, sample designs, file descriptions and database schemas, codebooks and classification structures, processing details, checks, transformation, weighting, fieldwork reports and notes, conceptual motivations, table designs and layouts.

With the rise of the Internet the concept of meta-data has become much more widely recognised. The major factor for this (though not the most interesting application) has been the Dublin Core standard for recording descriptive information about other resources. This has been taken up as an enabling technology for intelligent searches on the web.

What is metadata for?

This is probably a more important and interesting question than the question about the basic definition.

Statisticians usually start with data description – the set of labels for annotating output from analysis of a statistical dataset. This usually includes names and labels for variables and labels

for codes within the dataset, and is sufficient (for an informed user) for the creation of reports and the transfer of datasets between applications.

A usual and frequent extension is to support secondary uses of original data by other analysts. Then the metadata has to help the user to understand the data. Information is needed about the context and provenance of the data. This can include proper understanding of the meaning and intention of words in codes and questions, the methods of data collection, etc.

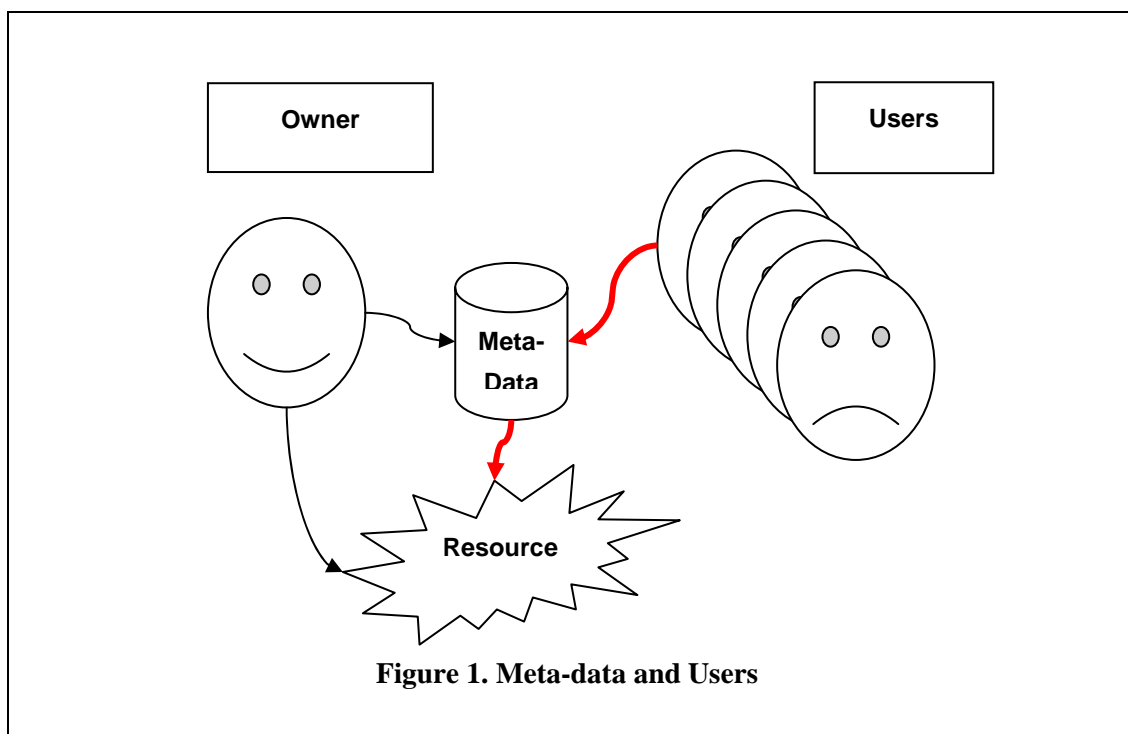
As data usage becomes more complex – for example, the integration of data from multiple sources, or more complex statistical analysis and modelling – the demands on the meta-data increase, with the need for deep understanding of coding systems and mappings, and for confidence in the quality of the underlying data and in the manipulation and analysis methods used.

Metadata supports use

From the perspective of use, we can formulate a definition more focussed on the way in which the meta-data supports the use of the underlying resource.

*If I have created a dataset, the metadata for that dataset is the information that I need to give to you for you to be able to use the dataset correctly for your chosen purpose (or to realise that the dataset does not support that use).*

More generally, metadata is the information that the owner of a resource needs to supply to potential users of that resource, so that they can use it correctly.



### Metadata as a resource

With this approach to metadata we need to move the focus away from the underlying structure and on to aspects that support its use. Firstly the correct content must be available, and secondly the correct functionality must be supplied to support the processes in which the metadata is used. For content we need to address issues of the structure that is to be allowed for the representation of the required information, and for the functionality we need to consider the processes in which the metadata will be used. The metadata should be seen as a

resource to be used, and the methods and means of use are as important as the specification of the structure of the content of the resource.

This approach echoes developments that have been taking place in recent years with regard to the Internet.

What is a resource?

Significant effort has been invested in recent years to extending the way in which the Internet is used. Its underlying ability to deliver content has been recognised some time but the move to address the issues of delivering functionality as well. The first within this direction is in the context of Web services. The idea here is that specific functionality can be centralised by some service provider on Web, and users can then communicate with this service, using an appropriate protocol, to make use of the processing methodology that is supplied. The second idea is the one often referred to as Web 2, which attempts to associate much more understanding (semantics) with objects and resources found on the Web. This is clearly closely related to the Metadata ideas that we are discussing.

This approach raises significant and deep questions. What is a resource? What is a service? How are these resources and services identified and used?

The answers to these questions are well beyond the scope of this paper, but it suffices to say that solutions are being discovered and developed within the computer Science and Internet domain. Protocols and standards have been developed, at both generic and specialised levels, and are now supporting many activities on the Internet. Many are in specialised areas and so not widely appreciated. Two that are very familiar to most Internet users are the use of search engines, such as Google, and payment schemes for authorising purchases over the Internet.

The use of Web services requires considerable trust between the client and the server. This can only be achieved when there is great precision and formality in the specification of the interaction between them. Standards are essential in this, and must address understanding and interpretation as well as structure.

Potential client systems need to be able to discover the existence of services, and to discover the standards required for communication with them. To this end there has been a parallel development in the concepts of registries and repositories. A registry holds links to services that have agreed to meet specific standards, and can be queried to discover specific servers offering particular types of service. A repository is similar to a registry, but is likely to contain fewer types of resource and more actual content. A registry can solve the discovery problem for particular services, but does not of itself ensure common understanding between the client and the service of the detailed implications of the service specification.

### The object paradigm

Underlying most recent development work is the object-oriented approach to systems. This was developed some time ago and came into prominence in the late 1990s. It is fundamental to most current programming methodologies, but also is applicable in any system design context.

Within this approach, the focus is always on an *object*. This is a collection of information, which has a type, and which has behaviour associated with it. The general type of an object is called a *class*, and this determines the general form of the content of that type of object, and the functionality associated with it. Instances of an object contain actual information specific



to a particular resource. The content of an object is generally a set of other things, including both other objects and primitive values.

Object instances have individual identity, so can be uniquely identified, and can respond to requests to perform tasks. These can be specific tasks related to the type of the object, or more general tasks such as reporting an object's own type, or listing the tasks which it can perform. Object types (classes) can share structure and behaviour through inheritance, in which one type of object is defined as a specialisation of another, usually with some enhancement to particular areas of structure or behaviour.

### Object modelling

The object paradigm sets out various principles about how systems should be designed. Following from this a number of methodologies were proposed for the process of system design (or modelling, the computer science use of this term). These different proposals were finally brought together in the Unified Modelling Language (UML), which has continued to develop and is now at version 2.1.1. A number of design and development applications (and environments) have been created which use UML: as *their* underlying model, and these can allow designs (models) to be exchanged.

UML is a very extensive specification, and is not easy, either to learn or to use, but it is capable of producing precise specifications of significant systems. It is not entirely watertight, but used carefully it can go a long way towards the production of an unambiguous specification of the structure and behaviour of a system. It is thus an important tool in addressing the problem of communicating understanding between the authors and users of systems.

Specifications in UML are formally structured, so can be transformed into other forms. For example, it is possible to generate executable code (programs) from certain types of carefully written UML specifications. It is also possible to transform the structural specifications in a UML class specification into the schema needed to create equivalent structures in XML. This facilitates the exchange of information about specific resources between different applications or implementations based on the same UML specification.

### XML for Exchange

XML is an extremely important technology for object systems, because it allows complex structures of information to be expressed as linear text in an XML Document. This is sometimes referred to as *serialisation* of object structures. Text documents are easy to move around over communication systems, so XML solves the problem of the transport (or exchange) of object structures between systems, whether different applications or different instances of the same application in different places.

The allowed content for a particular type of object in XML is defined using an XML Schema. The system is powerful and flexible and allows considerable detail to be specified about what is allowed at various points within a specific type of document. Even more generality is possible in UML, and because UML covers the design of complete systems it can also be used to specify the behaviour or functionality associated with particular elements of the object structure – this is not possible in XML schema.

It is thus recommended that UML be used to define the structure and behaviour of systems (including metadata systems), and then to generate from the UML the XML specifications

needed to control the exchange of information about specific resources between systems, in the form of XML documents.

### Levels of abstraction

Different types of application and context need different levels of abstraction in their conception. In dealing with practical problems we need to operate at a very concrete level, but when designing functionality for these operations it is often better to work at a more generic level. For example, users of the metadata about a dataset want to know about the meanings of the codes used within that particular dataset. However, designers of programs that use the metadata need to think more generically about what a code is and how it can be used appropriately. Similarly, those responsible for the integration or for the comparability of information from different data sources need to think at a more generic level about code sets, about structures of coding systems, about the concepts that underlie the questions asked, and so on.

There is an even more abstract level that is often needed. This relates (among other things) to the functionality and concepts that are needed to define things at the generic level. For example, with XML, the basic XML language is defined in a very abstract way, without reference to actual information resources. The XML language is used to construct XML documents, and some generic functionality is defined, regardless of the content of the documents. So software can read and write XML documents without needing to know the details about the content, and generic facilities support navigation through the structure, using the names of elements or their position. Similarly, a document can be displayed without knowing what it represents. In addition, the idea of a schema, with which to define the structure of a particular type of document, exists at this generic level. A schema then contains the specifications for the structure of a particular type of document, so this specification is at the intermediate level of abstraction. Actual instances of documents of this type (consistent with the schema) are the concrete level.

It is very important to be clear about the appropriate level of abstraction needed for the design or discussion of some structure or functionality. One of the most important outcomes of the MetaNet project was the realisation that most misunderstandings or confusion in discussions arose from a failure to recognise or agree on the level of abstraction appropriate for the context or application.

## 3. Model Representation in OPUS

### Metadata for Statistical Models

The OPUS project was concerned with the integration of information from multiple sources, through the use of statistical modelling. A component of the project addressed the issue of recording what had been done (a complete audit trail), so that subsequent analysts could review all the steps taken, and so that users of results derived from the statistical models could (if so inclined) obtain information about the uncertainty associated with such estimates. We refer to these two areas as Provenance and Reliability.

Most of the rest of this section discusses (at a high level) the components of the structure that we designed. The following section shows some of the display functionality that has been implemented. The uses that this resource could allow include:

Detailed exploration of the form of statistical model used, including the (conceptual) variables and parameters involved, and the relationships between these elements (both derived and stochastic).

Exploration of the quality of the input data used, through links to the actual datasets used in model fitting.

Investigation of details of the model fitting steps employed, including the software and algorithms used, plus progress and final states of iterative procedures.

Exploration of the confidence associated with parameter estimates (in a classical analysis) or the posterior uncertainty (in a Bayesian context) at the end of fitting steps.

Risk analysis for decisions based on results from the modelling, including sensitivity analysis to aspects of the model specification.

General investigation of model reliability and sensitivity.

Implementation has so far addressed only the first of these uses, and that not to any deep level. Most of the others have not yet been elaborated.

### Objectives for Data Structure

Our objective has been to design a structure (which we refer to as StatModel) for the representation of information about statistical models, together with appropriate functionality for the presentation of that information.

The data structure must include information about

- the structure of the statistical model used;
- the processing steps used to calibrate the model against data (model fitting);
- datasets used in model fitting steps;
- parameter estimates and uncertainty resulting from fitting (a model state);
- results derived from a model state

This information will come from other applications. Different applications embody different conceptual models of their methodologies and their application domains. Our aim with StatModel is to be sufficiently generic to be able to encompass these different views. This requires exploration of the mappings from application-specific views to the StatModel view.

Our immediate purpose for this information is to allow users to explore the specification of statistical models. However, some developers have suggested a need for a structure to allow the exchange of statistical models between applications, and we hope that StatModel may be a contribution to that.

We also hope that applications will want to use StatModel as their native structure for storing information about models and processes. So we allow for structural extensions to the model, both as new generic requirements are identified and so that applications can store information specific to their requirements.

When developing an understanding of an analysis domain it can be useful to store models that are incomplete, or that are defined at a level of abstraction above that needed for the computation of results. So the structure must not impose completeness rules unless they are always appropriate. Of course, this implies that applications using the structure must be able

to check whether an instance of the structure is complete enough for the purposes of the application.

We are not attempting to produce a complete statistical metadata system, so assume that metadata and presentation functionality for other components (such as classification structures and dataset documentation) will be available through external links.

## Users

Many users of statistical results have only limited understanding of statistical models and methodologies. For them, information about the statistical models needs to be presented in ways that relate to the reliability of the results for use in their working context, rather than the more abstract terms of the model itself. Furthermore, different domains build on different conceptual models for the form and description of relationships and dependencies, and an ideal presentation system should work within these.

In contrast, those who actually develop statistical models are more likely to have a good understanding of the more abstract concepts involved. For them a more generic presentation of model information may be adequate, or even preferred.

We thus envisage that significant use of the StatModel approach to support the use of statistical model results in a domain will require a presentation application that is specific to that domain. Such presentation systems are not too difficult to build using web-based methods. As an example, the Nesstar system [13] has been used in a number of large-scale dissemination projects to present basic statistical results and related materials to groups of users in specific domains.

We have concentrated in creating generic displays that can be used by specialists, and that can also be building blocks for the construction of more specific presentations.

## Structures

The structure and semantics of the classes used to hold information about a statistical model are specified in UML using the *hyperModel* Workbench application [14]. The main components of this structure for a single model are shown in Figure 2, and are elaborated in [15]. Information about multiple models can be stored together and they can make cross-references.

A *model* must contain a *model specification*, which is where the variables, parameters and relationships which specify the model are stored (the structure of these elements is elaborated later). The variables correspond to the statistical idea of Random Variables, that is they relate to suitable data subjects (for whom actual values may be observable) but we are interested in the stochastic distributions of the values, not the values for individual respondents.

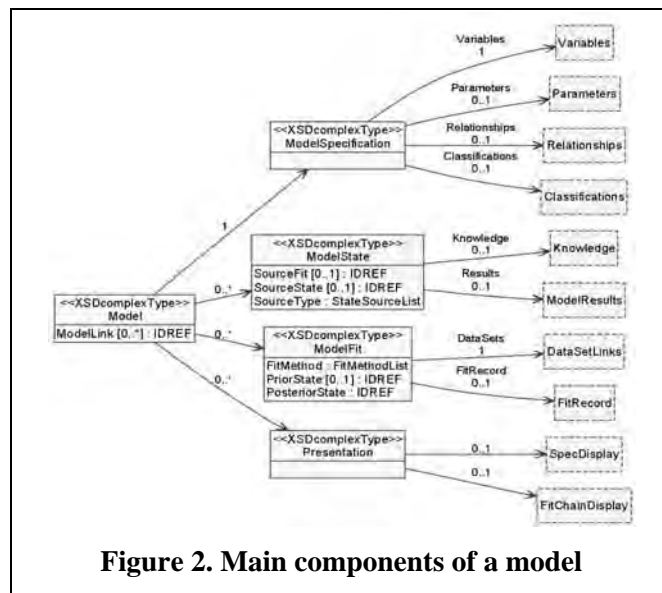


Figure 2. Main components of a model

Parameters are properties of the underlying system. They are real (fixed, though perhaps changing over time), but they cannot be directly observed. We extract evidence about them from data, but there will always be uncertainty about the true value. This uncertainty is represented by uncertainty distributions, which have the same mathematical properties as probability distributions, but a different interpretation.

At the early stages of development of a model it is sufficient to specify variables and parameters at a conceptual level. Their intention must be clear, so that influences between them can be identified, but details of their representation can be left for later. For example, a variable that relates to the income of a respondent should probably make reference to an appropriate definition of income, but does not need to be specific about currencies or whether the representation is exact or grouped. More measurement detail is needed when the details of the relationships are added, as these will include mathematical expressions.

Both variables and parameters can be array structures in which all cells are of the same type. Classifications are used to define the dimensions of such arrays.

Relationships specify how variables and parameters influence each other, and can be deterministic or stochastic.

A *Model State* contains knowledge about all the parameters in a model that are not determined by relationships, in the form of their uncertainty distributions. Following Bayesian methods there can be multiple states of the knowledge about a model. Any results derived from a model are based on a particular set of uncertainty distributions, and so can be linked to a specific state.

A *Model Fit* documents a step in which some application is used to extract evidence about the model from data. Such a step usually draws on knowledge from a model state (the Prior state) and always produces knowledge for a new state (the Posterior state). Model fitting processes are thus chains consisting of alternating fits and states, where each state is the output of one fit and the input to the next.

## Relationships in Statistical Models

Relationships show how the various elements of a model depend on and influence each other. The specification of the set of relationships is the essential core at the heart of statistical

modelling. This is a highly skilled technical activity that needs to be informed by both statistical ideas and a deep understanding of the domain to which the model is to be applied.

Relationships imply links between the variables and parameters in the model. These can be displayed in an Influence Graph. In this, each element is a node and the links connect from all the input elements of relationships to the corresponding output element. Where the resulting graph is acyclic, the model falls into the class known by statisticians as Graphical Models [16].

A relationship can have multiple inputs. The output of a relationship is always a single element (variable or parameter), and an element can only be the output of (be defined by) one relationship.

All dependent variables must be the output of a relationship with parameters and other variables (in order to specify their derivation or their stochastic properties). Variables that are not so defined are called ‘independent’ or ‘exogenous’. Parameters can also be specified by relationships with other parameters: those that are not are called ‘terminal’ or ‘free’. Uncertainty distributions must be supplied for them in any model state.

### Presentation Functionality

As discussed earlier, our main use for the information is to give confidence to users of the results of statistical modelling. Since these people will not be statistical specialists the presentations will need to be tailored to the level of understanding of the users, and to their conceptual views of the domain in which they operate. Generic presentations are also useful, but only for specialists or as building blocks.

We take as our model for presentation the Nesstar system. This is a system for building distributed access and dissemination systems for statistical datasets and results. Presentation is through a web interface based on templates and components. Components are provided to give access to any information that is stored within the system, including results derived dynamically from accessible data. But by using the web paradigm the presentation of the results can be customised to the target domain, and any other relevant information accessible on the internet can be incorporated in the presentation.

We have not attempted to construct a general presentation application for the metadata about statistical models. Instead we have concentrated on the development of generic presentation components that can be used as building blocks for such systems.

A generic facility for listing all the components of a StatModel instance is directly useful to specialists, and provides a resource from which appropriate elements can be extracted for more specific listings. A graph display applet is used to explore the influence relationships in the model, and to show the fitting processes used. Mathematical derivations are displayed using MathML [17, 18].

## 4. Opus Presentation Examples

### Generalised listing with style sheets

An XML style sheet has been developed to display all the components of a StatModel instance document, using a web browser. Figure 3 shows a fragment of the listing for a single complex model that has been used as a test case. This model explores the distribution of the

time spent travelling (within a day) on different modes of transport, and the extent to which this is affected by the characteristics of the traveller (including where they live). It includes interactions between different mode choices.

Name	Output	Input	Form
Derived ModeIndic		NMWMTime	<b>Derivation:</b> $\text{Step}(\text{NMWMTime} - 1.1) + 1$ For the 5 mode use variables, recode (1 to 1), (2, 3 & 4 to 2). This gives a vector of indicators showing whether each mode was used.
Derived BaseTimeProb		BaseTimeRate	<b>Derivation:</b> $\frac{\text{BaseTimeRate}[j,k]}{\sum_n \text{BaseTimeRate}[j,n]}$ For $j \in \text{Mode}$ , $k \in \text{NMWMTime}$ Convert Base Rates to Base Proportions for time distributions. This gives (for each mode) the travel time distribution for respondents who only use one mode and are in the base groups for all respondent attributes (usually group 1).
Derived RespTimeRate		BaseTimeProb, Alpha, Beta, Gamma, Delta, Kappa, Phi, Xi, Omega, ModeIndic, Age, HomePop, Sex, TravelDay, WorkMode, WorkMode2, BreathProb, HeartProb	<b>Derivation:</b> For $j \in \text{Mode}$ , $k \in \text{NMWMTime}$ $\left\{ \begin{array}{l} \text{BaseTimeProb}[j, k] \\ \times \text{Beta}[j, 1, \text{ModeIndic}[1]] \times \text{Beta}[j, 2, \text{ModeIndic}[2]] \times \text{Beta}[j, 3, \text{ModeIndic}[3]] \\ \times \text{Beta}[j, 4, \text{ModeIndic}[4]] \times \text{Beta}[j, 5, \text{ModeIndic}[5]] \\ \times \text{Alpha}[j, \text{HomePop}] \times \text{Delta}[j, \text{TravelDay}] \times \text{Phi}[j, \text{Sex}] \\ \times \text{Omega}[j, \text{WorkMode}] \times \text{Omega}[j, \text{WorkMode}2] \\ \times \text{Kappa}[j, \text{BreathProb}] \times \text{Xi}[j, \text{HeartProb}] \times \text{Gamma}[j, \text{Age}] \end{array} \right\}$ If $(k = 1)$ then else $\{\text{BaseTimeProb}[j, k]\}$ Adjust Base Proportions for time distributions to give Respondent Rates. The probability of NOT using a mode depends on the use of other modes and the other attributes of the respondent. Where a mode IS used the time distribution does not depend on the other factors. Each Beta is a factor by which the probability of NOT using a mode is altered if another mode IS used. The other parameters are all adjustment factors when a respondent attribute is not the base group.
Derived RespTimeProb		RespTimeRate	<b>Derivation:</b> $\frac{\text{RespTimeRate}[j,k]}{\sum_n \text{RespTimeRate}[j,n]}$ For $j \in \text{Mode}$ , $k \in \text{NMWMTime}$ Convert Rates to Proportions for Respondent time distributions.
Stochastic NMWMTIME		RespTimeProb	<b>Distribution: Categorical:</b> <b>Probabilities:</b> RespTimeProb The second index of RespTimeProb gives the probabilities of the four categories of travel time (of which the first is non-use).
Derived Beta		LBeta	<b>Derivation:</b> $\exp(\text{LBeta})$ The probability calculations use multiplicative factors, but the Bayesian fitting is done using the logged versions of these parameters. This has the effect of constraining the factors to be positive, but also tends to be more effective computationally.

Figure 3. StatModel Listing (part)

The fragment in the figure is the start of the section that lists the relationships in the model, and it shows the input and output elements used in each relationship, together with the mathematics of the relationships. It also shows how comments that describe the purpose of the relationship can be associated with it. Colour coding is used in the listing to highlight variables and parameters and to identify comments. Everything in the listing (apart from the italicised headings) is taken directly from the instance document.

The information presented by this listing is a complete presentation of the contents of the XML document, and is often adequate for exploration by the originator of the model and other specialists. It can be more helpful than the original script used in WinBUGS (the statistical application used for this example) for less experienced users, but in general a user will need more explanatory information and context.

For specific applications the style sheet can be used (by someone familiar with XSL/T) as the basis for a more focussed listing. However, this approach has limited applicability.

### Graphical display of model relationships

Components have been developed to provide dynamic displays of parts of the StatModel instances within a web browser. Figure 4 shows the testing interface which is made available

to partners and associates of the Opus project. This includes the full influence diagram for the model of which Figure 3 lists a part.

The display is a Java applet, and the graph is created dynamically from the information in the XML document. Controls are provided for a number of different automatic graph layout algorithms, the user can zoom and pan the display, and can select to manually adjust the location of individual nodes. Various graph trimming and filtering options are provided, and interesting sub-graphs can be pre-defined.

Colour coding is used to distinguish between variables (green) and parameters (blue) in the display, and the node shapes distinguish between derived, stochastic and other relationships. Constraints are shown in light grey.

Figure 4 shows all the relationships in this model. The display is complex, but some regularity can be discerned, suggesting that the complexity comes from the number of elements, rather than the basic complexity of the model.

Figure 5 shows two sub-graphs. The left one extracts just the part that relates to the response variable in the dataset, the variable NMWMTIME, which is a vector showing time spent on each of 5 transport modes, grouped into four classes. BaseTimeRate is an array giving the basic usage rates for each mode. This is a terminal parameter, indicated by the lozenge shape. These figures are standardised in BaseTimeProb and then adjusted in RespTimeRate to account for the characteristics of a data subject, including their mode usage. These rates are again standardised to give probabilities (in RespTimeProb) which are used in the stochastic relationship (indicated by an oval) to define NMWMTIME. ModeIndic is a vector of derived indicators for whether each mode was used at all, which feeds back into the calculation of RespTimeRate.

The right-hand part illustrates how some of the covariates contribute to the calculation of RespTimeRate: Age selects a Gamma parameter, Phi relates to Sex, and Beta to Mode Usage. The mathematics of these relationships can be read from Figure 3.

### Comparison of models

The influence graphs can make it

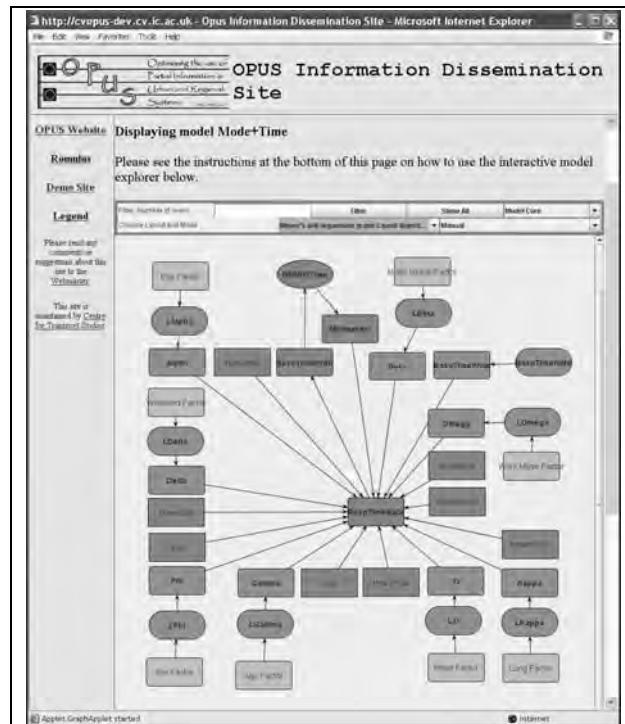


Figure 4. Full influence graph in browser

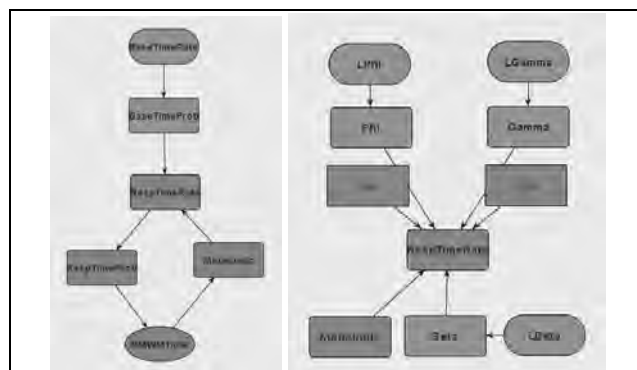


Figure 5. Sub-graphs

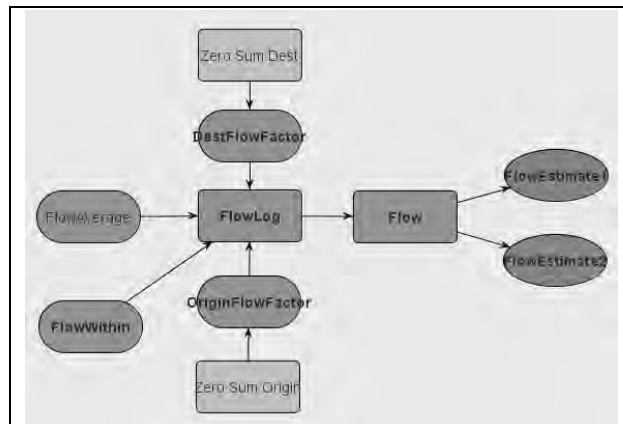


straight forward to see the differences between related models. For example, Figure 6 shows the graph for a log-linear model in which two independent estimates are used to calibrate a single model.

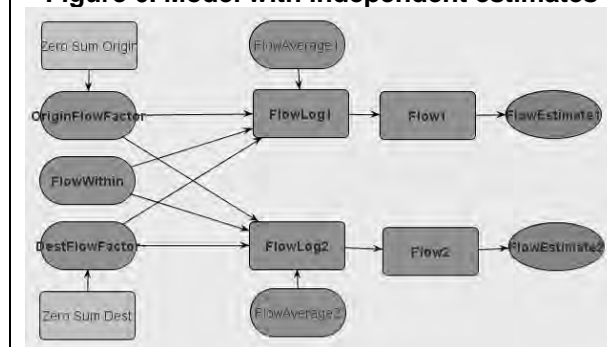
This example arises when two quite different methods (such as a household survey and roadside measurement) are used to estimate the flow between zones in a transport system, and a combined estimate of the parameters is desired.

Figure 6 treats the two estimates as independent samples from exactly the same stochastic process. In contrast, Figure 7 shows the model in which it is recognised that the two different data collection methods may influence the estimates.

Here the origin, destination and within factors are still shared, but it is recognised that the average level of flow obtained from the two datasets may have been influenced by different biases, so are estimated separately.



**Figure 6. Model with independent estimates**



**Figure 7. Model with differentiated estimates**

## References

- [1] Hughes, K., Jenkins, S. and Wright, G. *The Triple-S Survey Interchange Standard*: See [www.triple-s.org](http://www.triple-s.org)
- [2] Dublin Core Metadata Initiative. See [dublincore.org](http://dublincore.org)
- [3] UK GovTalk. *e-GMS, the UK Government metadata standard*. See [www.govtalk.gov.uk/schemasstandards/metadata\\_document.asp?docnum=872](http://www.govtalk.gov.uk/schemasstandards/metadata_document.asp?docnum=872)
- [4] IQML: Intelligent Questionnaire Markup Language. See [www.epros.ed.ac.uk/iqml](http://www.epros.ed.ac.uk/iqml).
- [5] QEDML: Questionnaire Exchange and Deployment Markup Language. See [www.qedml.org](http://www.qedml.org).
- [6] Kent, Jean-Pierre & Schuerhoff, Maarten: *Some Thoughts About a Metadata Management System*. SSDBM 1997.
- [7] UML, the Unified Modelling Language. See [www.omg.org](http://www.omg.org).
- [8] SDMX.. *Statistical Data and Metadata Exchange*. See [www.sdmx.org](http://www.sdmx.org)
- [9] DDI Alliance. *Data Documentation Initiative*. See [www.ddialliance.org](http://www.ddialliance.org)
- [10] MetaNet: *Network of Excellence for Statistical Metadata*. See [www.epros.ed.ac.uk/metanet](http://www.epros.ed.ac.uk/metanet)
- [11] Froeschl, Grossmann, Del Vecchio: *The Concept of Statistical Metadata* (2003), at [www.epros.ed.ac.uk/metanet/deliverables/deliverables.html](http://www.epros.ed.ac.uk/metanet/deliverables/deliverables.html)
- [12] The Opus Project. See [www.opus-project.org](http://www.opus-project.org)
- [13] *Nesstar*. See [www.nesstar.com](http://www.nesstar.com)
- [14] Carlson, D.A. et al. *hyperModel Workbench*. See [www.xmlmodeling.com](http://www.xmlmodeling.com)
- [15] Westlake, A. *Provenance and Reliability: Managing Metadata for Statistical Models*. In proceedings of 18<sup>th</sup> SSDBM, IEEE (2006) ISBN 0-7695-2590-3.
- [16] Whittaker, J. *Graphical Models in Applied Multivariate Statistics* (1990). Wiley.
- [17] *MathML*. See [www.w3.org/Math](http://www.w3.org/Math)
- [18] Design Science. *Math Player – display MathML in a browser*. See [www.dessci.com/en/products/mathplayer](http://www.dessci.com/en/products/mathplayer)

## About the Author

Andrew Westlake has more than thirty-five years experience contributing to the design, implementation and management of statistical computing projects in university departments and international research organisations, and is a Fellow of both the Royal Statistical and the British Computer Societies, as well as an Honorary Member of the ASC. He has been deeply involved in software development for survey and statistical applications, and has gained an international reputation for the application of database methods to statistical problems. His research projects have involved work for statistical, demographic and health departments in academic, research and government organisations throughout the world, and have varied in scale from small sample surveys up to national vital registration systems. His current special interest is in the structural and metadata aspects of databases for aggregate statistical data and statistical models. He can be contacted through [www.sasc.co.uk](http://www.sasc.co.uk).

# Triple-S: The Broader Horizon

Geoff Wright

## Abstract

The Triple-S standard has been in use for over 10 years, and now has over 50 registered implementations. For the past 4 years the Triple-S Group has been working on a significant enhancement to the standard. This paper presents some of these new features and discusses how adoption of the standard is moving beyond its original concept as a survey interchange format.

## Keywords

Standards; Survey Interchange; Survey Metadata; Triple-S; XML

## 1. Background and History

The Triple-S survey interchange standard was originally developed in order to make the transfer of data and metadata between survey packages easier. The problem of a proliferation of bespoke imports and exports between different survey packages was discussed by Peter Wills of Mercator in a paper given at the First ASC International Conference in Bristol (Wills, 1992). The Triple-S Group came together soon after and the initial version of the standard was published in 1994. The group has continued to develop and publicise the standard whilst ensuring that it is not commercially driven by one supplier.

The standard defines two text files that describe the survey data. The *Definition File* contains general information about the survey together with descriptions of the survey variables (i.e. the metadata). The *Data File* contains the actual case data for the survey. Although many consider the definition file to be the Triple-S standard, it is in fact the formal definition of both files that constitutes the standard (i.e. Triple-S is not a means for describing an arbitrary data file).

Use of the standard was very limited until the adoption of XML for the Definition File. This was described in a paper given at the Third ASC International Conference in Edinburgh (Hughes, Jenkins & Wright, 1999) and the new standard was subsequently published in 2000. This XML version of the standard now has over 50 registered implementations.

Consideration of more significant enhancements to the standard started after a presentation on the future direction for Triple-S (Wright, 2003) and at an open meeting during the Fourth ASC International Conference in September 2003. A range of options for possible improvements was produced and then discussed at a meeting with a number of developers and users in March 2004. This resulted in a reduced set of changes which were then refined

during the rest of 2004. The resulting Triple-S XML 2.0 standard was published in April 2005.

## 2. Additional Metadata Information in version 2.0

Having monitored the use of the existing standard, the Group appreciated that there were places where exporters and importers would like the ability to specify additional metadata information.

### Score attributes

It was recognised that many variables of type *single* had an associated set of scores for the values. These score values would often be used for computing statistics such as a mean rating during analysis. In order for the researcher to transfer these scores the standard was extended to permit an optional numeric *score* attribute on the *value* elements. For example:-

```
<variable... type="single">
...
  <values>
    <value code="1" score="2">Liked it very much</value>
    <value code="2" score="1">Liked it a little</value>
...
    <value code="5" score="-2">Disliked it very much</value>
```

### Non-numeric value codes

Although it is possible for all variables of type *single* to be exported as numeric codes, there are many survey programs that allow data to be coded with letters. This convention is often used for recording special status values, such as “don’t know”. In these situations an exporter would often prefer to export these values as is, without having to recode them.

To allow for this the standard now permits the optional use of “literal” codes. Any variable of type *single* can be defined with the attribute *format="literal"*. This indicates that the codes are to be treated as characters rather than numbers, even if the code is represented by numeric digits. For example:-

```
<variable... type="single" format="literal">
...
  <values>
    <value code="0">Never</value>
    <value code="1">Once a day</value>
...
    <value code="4">Less frequently</value>
    <value code="X">DK/NA</value>
```

### Date and Time values

Hitherto there had been no way to indicate that a data value represented a date or time. The new standard recognises that providing explicit *date* and *time* variable types allows exporters and importers the ability automatically to transfer these values. Whilst there are many ways to represent dates and time, we have adopted the basic ISO 8601 all-numeric format within the standard. This means YYYYMMDD for dates and HHMMSS for times.

## Separate Interview and Analysis texts

With increasingly sophisticated survey systems, the actual question text may be unsuitable when presenting the results (e.g. because it is too long, includes HTML or runtime text substitution). To provide for the situation where the researcher has created both the raw interview text and a simpler question text for analysis, the standard now supports a *mode* attribute on the *text* element. Two explicit modes are available: “interview” and “analysis”.

```
<sss version="2.0" modes="interview analysis">
...
<variable... type="quantity">
  <name>Q12B</name>
  <label>Age
    <text mode="interview">How old is %Q12A (in years)?</text>
    <text mode="analysis">Age of child</text>
  </label>
...
```

## 3. More Data Structures in version 2.0

Increasingly the use of a single flat fixed format data file is failing to provide enough flexibility for exporters and importers.

### CSV format data files

With programs like Excel being used as part of the analysis process, there is a move towards exporting data in CSV (Comma Separated Values) format. As a result the standard now supports an option for data in this format, with data values similar to the existing standard (i.e. one record per case, and one field per value).

The first line or lines of a CSV data file are sometimes used as documentation for the succeeding values (e.g. names for the columns/fields). An optional *skip* attribute on the *record* element can be used to ignore one or more initial records in the file.

### Hierarchical data structures

Probably the most significant advance in version 2.0 is the ability to specify hierarchical data structures. This has been done by building on the existing standard, rather than a substantial redesign. In effect a simple Triple-S “survey” specification is used to describe each level of the hierarchy. Then a separate Triple-S “control file” specifies how these files are linked. This means that each level can be analysed on its own as a simple flat “survey”, or if the importer can cope with hierarchical data then the entire hierarchical survey can be processed.

A hierarchical data structure can be used in a number of ways, such as:-

#### Master/trailer type surveys

This will be the most common and straightforward use where, for example, we have some questions at the household level followed by mini-surveys of each member of the household. In this example there would be one specification and data file for the household level (e.g. hhdat.xml), plus another specification and data file for the person level (e.g. persondat.xml). The control file would look like:-

```
<hierarchy>
  <level ident="hhold" href="hhdat.xml"/>
  <level ident="person" href="persondat.xml">
    <parent parlev="hhold" linkvar="hhnumb" ordered="yes"/>
  </level>
</hierarchy>
```

#### Representation of Grid Rows

A second possible use for a hierarchical approach is where the survey includes grids with either a large or unlimited number of rows. Specifying each row as separate variables within a single survey often results in a large number of very similar variable names, and a long data record which may be very sparse. Adopting a hierarchical structure could better represent this kind of data, but with the downside that many existing importers will not be able to handle this structure.

## 4. Standard data import by new analysis software

One trend that the Triple-S Group have observed is that the availability of a widely supported standard for transferring survey data and metadata easily leads to its adoption as the standard data import when writing new analysis software. This should make the writing of the data input easier, and ensure that the software can work with a large number of other products.

#### The Risk-E program

One such product which decided not to invent its own wheel is Risk-E ([www.risk-e.net](http://www.risk-e.net)), which was reviewed by Andrew Jeavons in ResearchWorld (2006). This is a decision support tool, developed by the CMR Group Ltd in association with Computable Functions Ltd, to assist business decisions taken on the basis of survey research data. Risk-E identifies those sub groups in the survey population with a significant influence on the key findings. It can then alert users to the possibility of bias in the key results because the predictor groups have not been correctly represented in the sample. The data can then either be weighted to correct the bias, or additional interviews can be conducted to correct for any shortfall in numbers.

Although most of these analyses could be done within a large general purpose statistical package, Risk-E provides these in an accessible and understandable context. In addition to ease of use, the product uses re-sampling techniques so that the statistics do not rely on assumptions about underlying distributions.

As a new standalone tool for investigating survey data the authors needed to provide a mechanism for data import, and they have adopted Triple-S. Iain Mackay from Computable Functions summed up their decision by saying “You can point the program at a Triple-S dataset and you are good-to-go immediately.” He said that the main reasons for selecting Triple-S were:-

The Triple-S question types were just what were needed for the application.

They did not want to write a data definition/preparation system – that would be the tail wagging the dog.

The use of XML for the metadata made it straightforward to process.

Many major data collection systems can generate a Triple-S dataset.

There were some drawbacks to the decision, in particular because many sources of case data don't come from major data collection systems. So the authors have developed a separate utility to facilitate building a Triple-S dataset from delimited files (CSV2SSS), which covers the majority of non Triple-S sources.

## 5. Use within the wider data analysis arena

The standard was initially produced for use in packages within the market and survey research area, and developments have always been driven by this usage. But the representation of data as records containing variables (e.g. quantity, single) extends beyond this community. Through the suggestions and comments section of the website, the Triple-S Group has been contacted by a number of users from outside this area.

We are aware of situations where Triple-S is being used either in whole or as the basis for an internal standard for metadata. Areas like Business Intelligence, Employee Satisfaction, and Mystery Shopping are benefiting or could benefit from the use of Triple-S. Not only does this provide a convenient internal standard, but also allows the data to be processed by general purpose survey analysis tools in addition to their own specialised tools.

### 3BI Consulting SA

One example of a company from outside the traditional market research area that has adopted Triple-S is 3BI Consulting in Switzerland ([www.3-bi.ch](http://www.3-bi.ch)). The company offers comprehensive Business Intelligence solutions and services, designed for internally and externally collected data. In order to have an integrated approach to data from heterogeneous sources, they use Triple-S extensively as their standard internal working format.

To achieve this they have developed a Java library to read many types of data formats. This library loads data in memory in a class structure that reflects Triple-S format, and allows them to work with and convert practically anything into Triple-S. Using this they have developed utilities such as:-

- A validation tool for Triple-S which checks the structure and permits graphical inspection of the data and metadata.

- A data loading tool that imports large Triple-S datasets into a Microsoft SQL data base.

- A data transformation tool from SPSS to Triple-S.

But more significant is the range of end-user software that they have created for clients where all the data loading and processing uses Triple-S format. This includes:-

- Software that reads Triple-S data for trend hunting.

- Analysis software that reads Triple-S data for performing pricing simulation, elasticity calculation and portfolio optimisation.

- A client reporting system for Banks.

Since the adoption of Triple-S the company has processed hundreds of gigabytes of consumer data coming from all around the world in a variety of formats, but all processed using Triple-S format. This represents a type and extent of use that we would never have foreseen when the Group started work.



## 6. Conclusion

This paper shows how the most recent revision of Triple-S has adapted to the changing needs of a standard for survey data interchange. But more interestingly we are now increasingly seeing Triple-S as a format for describing case data as input to new software and as an internal format extending outside the traditional market and survey research fields.

However these new uses of the standard do provide some problems for the Triple-S Group. New software can sometimes require metadata information that is not provided by Triple-S, and internal use can easily lead to local extensions that are hard to resist but reduce the ability to work with outside products.

Examples include shared answer lists and the representation of codes (e.g. part numbers, or postal codes) that have a large number of valid values but with a standard representation format. But overall the decision for more companies to base everything on a standard like Triple-S can lead to the generation of an extensive library of supporting code and utilities.

## References

- Hughes, K., Jenkins, S., and Wright, G. (1999) triple-s XML: A Standard within a Standard. Proceedings of the 3rd ASC International Conference, Edinburgh UK.
- Jeavons, A. (2006) Risky business. ResearchWorld, December 2006.
- Wills, P. (1992) Data Use and Reuse: The movement and use of survey data across different hardware and software platforms. Proceedings of the 1st ASC International Conference, Bristol UK.
- Wright, G. (2003) triple-s: Managing Success. Proceedings of the 4th ASC International Conference, Warwick UK.

## About the Author

**Geoff Wright** is a director of Computable Functions Ltd., an independent survey and media research IT consultancy. He has been using, designing, and advising on survey and statistical software for more than 30 years, and has been a member of the Triple-S Group since its inception in 1992. He can be contacted at Computable Functions Ltd., Enigma House, 30 Alan Turing Road, Guildford GU2 7AA, England; e-mail: [geoff@computable-functions.com](mailto:geoff@computable-functions.com).

# A divergence-weighted independence graph for finite population surveys

Joe Whittaker & Chung Kao

## Abstract

The analysis of survey data, collected on a set of response variables defined over a finite population, may benefit from a bird's eye view of the strengths of their inter-relationships. This overall analysis highlights those variables that strongly modify the conditional distribution of another variable, and by contrast, indicates those which have little affect. We introduce a weighted graph based on measures of independence strength calculated from the population that fulfils this purpose. We show that the graph may be properly defined in terms of population measures without any appeal to super populations, probability modelling or to likelihood. A sample of young women and their smoking behaviours, taken from the General Household Survey is used as an illustration.

## Keywords

Conditional independence, design-based survey analysis, graphical model

## 1. Introduction

### Motivation

The survey analyst has observations on a collection of variables defined on units in a finite population. A central interest is to provide an overall view of the inter-relationships of these variables and their magnitudes. We would like to know which relationships are relatively strong, and which are weak, or equivalently, which variables need to be included in a predictor of another variable, and which can be ignored.

### Background on graphical models

The seminal paper on graphical models, Darroch et al (1980), combined the concept of conditional independence between random variables and the notion of a graph on these variables, into a unified theory. It provides an overall view of the inter-relationships of these variables in the context of a probability model for a contingency table.

Two modelling assumptions are made, one at the level of the variables, where the joint distribution of the observed variables is presumed to admit factorisations indexed by specific values of parameters. This codes independences between variables. The second is at the levels of the units, where the contingency table is constructed from simple random sampling with replacement.

A substantial statistical literature has developed following this paper, and several text summarising parts of the theory and practise are available, see for instance Whittaker (1990), Lauritzen (1996), Edwards (2000), and with applications from social science Wermuth & Cox (1996). Much of the work is to develop model based statistical procedures such as likelihood ratio tests, model comparison, and variable selection for graphical models. In practise the high dimensional parameter raises difficult model selection problems, so that the graph (or graphs) resulting from the analysis are not particularly stable, and may have no clear structure. Our contention is that much of this endeavour is unnecessary within a design based analysis.

A graph  $\mathcal{G}$  consists of two sets, one of vertices representing random variables, and another of edges, representing probabilistic dependency between random variables. The underlying concept in such graphs is the relationship of conditional independence, between two variables or subsets of variables, conditioned by a third subset. Conditional independence corresponds to a missing edge in the graph, so that two variables that are not adjacent are conditionally independent given the remaining variables. The diagram of the graph gives an overall visual display of the variables and their associations.

### Information divergence

The Kullback Leibler (KL) information divergence (Kullback & Leibler (1951), Kullback (1959)) measures the divergence (distance) between two probability distributions for a random variable  $X$ . It is an expectation of the log likelihood ratio of the two distribution averaged over the values  $X$ . When the expectation is taken over a finite sample the measure is the sample mean of the log likelihood. Later we use this relationship to numerically calculate interesting KL measures using standard statistical software for fitting GLMs.

### Our contribution

Graphical models have successfully married the ideas of conditional independence and of graphs to succinctly represent multivariate distributions. Generally this exercise has taken place within the paradigm of statistical modelling. Here we employ the same ideas within the design based paradigm of finite population survey analysis, for instance, Cochran (1977). The key idea is to consider population measures of conditional independence as population defined parameters, so that standard design based survey analysis can then used to estimate these parameters from samples.

The graph is defined with respect to a specified collection of observed survey variables. The measures of interest are the information divergences against the independence of each pair of variables conditional upon the remaining variables in the collection. The divergence measures are used to define and display the weighted graph which provides a visual representation of the strengths of association in the collection. The connection to standard graphical models is that an exact conditional independence corresponds to a zero of the divergence measure and a missing edge in the graph.

In order to maintain the comparability of measures the variables are assumed to be categorical or are recoded to be so, and the survey analyst has to choose the variable collection of interest. The computation of the divergences can be made by using the output of binary logistic regressions. This logistic regression framework easily incorporates design weights from stratification and from post calibration when these are available.

A substantive application of the techniques is given to a study of women's smoking based on a sample taken from the UK's General Household Survey. The weighted independence graph is shown to be a useful tool for exploratory analysis of the data set. Various collections of variables are of interest, enabling comparisons of different years, and of different definitions of the smoking variable, as well extending the analysis to subsets of women defined by having a partner and so having more measured variables.

## 2. Measures on finite populations

### Finite population universe

A finite population  $\mathcal{P} = \{t | t = 1, \dots, N\}$  consists of a set of  $N$  units. The  $k$  survey variables, denoted by  $(y_1, y_2, \dots, y_k)$ , are measured on each unit indicated by a superscript  $t$ . The possible values of  $y_i$  are finite. Innately continuous variables are discretised.

The two most important population measures are means and proportions, see Cochran (1977). The population proportion  $\phi_i(r) = \sum_{t=1}^N \text{Ind}(y_i^t = r) / N$ , where  $\text{Ind}$  is the indicator function, is the proportion of units for which  $y_i$  takes the value  $r$ .

The baseline of interest is independence, defined by

$$\phi_{ij}(r, s) = \phi_i(r) \phi_j(s) \quad (1)$$

for all  $r$  and  $s$ .

One population measure of association for two binary variables  $Y_i$  and  $Y_j$ , both taking values 0, 1, is the cross product ratio (cpr). An essential property of this measure is that the variables are independent if and only if the cpr is 0, for instance Whittaker (1990). There are other ways to measure dependence and some, including the cpr, generalise to more than two categories, and to more than two variables. For a review see Agresti (2002).

### Information divergence against independence

The KL information divergence against independence in the finite population setting is

$$\text{Inf}(Y_i \perp Y_j) = \mathbb{E}_p \log \frac{\phi_{ij}(Y_i, Y_j)}{\phi_i(Y_i) \phi_j(Y_j)}, \quad (2)$$

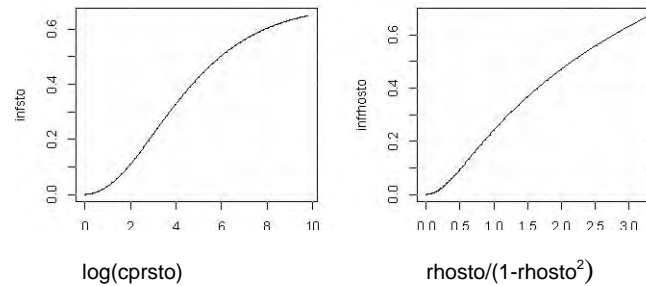
a double sum over the population. It measures how nearly the joint population proportion factorises into the product of marginal proportions when averaged over the population.

Using the general properties of information divergence, see Cover & Thomas (2002).  $\text{Inf}$  is always non-negative, and  $\text{Inf} = 0$  if and only if the variables are independent.

### A comparison of measures

An idea of the magnitude of  $\text{Inf}$  can be obtained by comparing it with the correlation coefficient and the cpr. When two Normal random variables are standardised to have zero means, unit variances and a correlation coefficient  $\rho$  then  $\text{Inf}(\rho) = \frac{1}{2} \log(1 - \rho^2)$ , for instance see Whittaker (1990). It is plotted in the Figure.

It is hard to give an analytic expression for  $\text{Inf}(\tau)$  between two binary random variables, but it is numerically tractable. With the variables standardised to have equi-probable margins and a given cpr  $\tau$ , the information  $\text{Inf}(\tau)$  is plotted the left panel of the Figure.



Both plots are monotonically increasing; both divergences are plotted on the scale of the interaction, defined by the coefficient of the cross-product in the log linear expansion of pdf and pmf, which makes the plots more linear. Both reach approximately 0.6 for the ranges plotted ( $\tau = 0..10$ ,  $\rho = 0..0.86$ ), and for moderate dependence, with  $\tau = 0.2$ ,  $\rho = 0.42$ , the information against independence is 0.1.

### Information against conditional independence

The KL information against independence generalises to three (and more) variables with a joint frequency computed from the finite population  $\mathcal{P}$ , to the information divergence against conditional independence,  $\text{Inf}(Y_i \perp Y_j | Y_k)$ . It is the measure we use to weight a graph of variables from the population. It quantifies the information against the independence of  $Y_i$  and  $Y_j$  given other variables, averaged over all values taken in the population. It has similar properties to the information divergence against independence, including positivity and a zero which implies conditional independence.

### Degenerate distributions

With  $k = 10$  binary variables there are over 1000 cells in the population and certain combinations may never occur. It is fairly easy to show that the information against independence remains well defined and stable in such degenerate distributions where certain proportions  $\phi$  are 0.

### A multinomial predictand: additivity into conditional binaries

Our intention is to assist the analyst with a tool that measures interaction strength between the measured survey variables. The common framework presupposes that each study variable can be coded as a set of binary variables. For a continuous variable where each binary variable corresponds to a cut there may be some loss of information. For discrete (ordinal and categorical) variables there is no loss and the information measure for multinomial variables may be obtained by additivity over the sub-populations defined by the binary variables.

## 3. Divergence weighted independence graph

### The $k$ -dimensional divergence weighted graph

We define a graph to portray the independence relationships manifest in a subset of  $k$  survey variables, and defined entirely in terms of population measures. A  $k$ -dimensional divergence weighted independence graph is the graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ , with vertices  $\mathcal{V} = \{i | i = 1, \dots, k\}$

corresponding to the variables, edges  $\mathcal{E} = \{(i,j) | i,j \in \mathcal{V}\}$  and weights  $\mathcal{W} = \{w_{ij} | i,j \in \mathcal{V}\}$  where  $w_{ij} = \text{Inf}(Y_i \perp Y_j | Y_{\text{rest}})$ . The graph is complete, in that all edges appear, and a natural form of display is to set the width (and/or tone) of each edge proportional to the weight.

We make some remarks. The ‘rest’ are the remaining variables, so that  $w_{ij}$  is the extra information for predicting one variable from another after conditioning upon the rest. If the edge is excluded from the graph whenever  $w_{ij} = 0$  and other weights are ignored, the graph becomes identical to the classic conditional independence graph of Darroch et al (1980).

We choose to use conditional measures of divergence rather than marginal measures for two reasons. Firstly the resulting graph approximates the conditional independence graph and makes its separation or Markov properties available. Secondly using the conditional divergence implies strong neighbours of a vertex (those with high edge strengths) are always required for the best prediction of the vertex.

#### Example: a graph on 4 variables

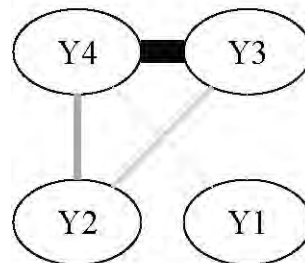
Consider a toy example, with  $k = 4$  binary variables and with joint proportions over the population given by a linear relation to the variable number. In standard order, the population proportions, as a percentage, in the 16 categories linearly increase and are

0.7, 1.5, 2.2, 2.9, 3.7, 4.4, 5.1, 5.9, 6.6, 7.4, 8.1, 8.8, 9.6, 10.3, 11.0, 11.8

The divergences against conditional independence (multiplied by 100) are summarised in the matrix

-	Y1	Y2	Y3	Y4
Y1	0.000	0.031	0.065	0.110
Y2	0.031	0.000	0.202	0.367
Y3	0.065	0.202	0.000	1.051
Y4	0.110	0.367	1.051	0.000

The matrix is symmetric, as it should be; the values on the diagonal are zero, all other elements are positive with a largest value of 1.051 between  $Y_4$  and  $Y_3$ . Setting the thickest width and darkest tone of an edge to 1.10 gives the divergence weighted independence graph below.



The relative sizes of the divergences are now apparent. The smallest divergence is between  $Y_1$  and  $Y_2$ . The variable  $Y_4$  interacts most strongly with the other variables, and so is most predictable, while  $Y_1$  is the least predictable. The graph is complete but thresholding the edge strengths to specific values leads to approximations of the population proportions.

The appearance of the graph is sensitive to the setting of the thickest width.

## Directed graphs

If there is a given ordering to the variables so that it only makes sense to predict one variable from another if it is 'later' in the ordering makes sense to modify the definition of the weights. The conditioning variables  $Y_{\text{rest}}$  are replaced by only conditioning on the predecessors of the predicted variable in the given ordering. This gives an appropriate weighting to a graphical model on a directed graph, often referred to as a Bayesian net. Furthermore it generalises to a chain graph.

## 4. Methodology

### Embedded optimisation: logistic regression

Logistic regression, for instance Hosmer & Lemeshow (2000) is a well known statistical technique for regressing a binary response variable from a linear predictor formed from a linear combination of predictor variables. The deviance is twice the maximised log-likelihood ratio test statistic for fitting the combination compared to that of the saturated model.

Consider the deviance difference for fitting a given binary response variable  $i$ , on the combination of the remaining variables with, and without, variable  $j$ . When the set of predictor variables are composed of categorical explanatory variables and all their possible interactions, it can be shown that the deviance difference is the divergence against the conditional independence, scaled by  $1/2n$ . That is, the divergence may be calculated numerically as an embedded optimisation problem.

A linear approximation to the divergence may be taken by suppressing the interaction terms in the linear combination. It may then be shown that  $\text{Inf}_{ij} > \max(\text{dev}_{ij}, \text{dev}_{ji})/2n$ . The right hand side gives an estimate of the left.

### Algorithm

A unit based numerical computation of the divergences can be obtained from the glm function for logistic regression, using the binomial family, with the logit link, and specifying the unit weights. The pseudo code using the linear approximation goes as follows.

Initialise:

Choose a collection of variables of interest  $\mathcal{V} = \{1, \dots, k\}$ , a subset from the full set of survey variables. Arrange this data in the form of an  $n \times k$  data matrix, together with a stratification weight vector  $v$ . Set  $\text{dev}_{ij} = 0$ , the deviance difference between the complete model and model without variable  $j$ .

Cycle:

For each variable  $i$  in  $\mathcal{V}$ , for each variable  $z$ , say, in the binary representation of  $y_i$ , fit the glm of  $z$  on  $y_{\text{rest}}$  where the  $i$ -th variable is excluded [and store the deviance]. For each variable  $j$  but not  $i$ , fit glm of  $z$  on  $y_{\text{rest}}$ , where both the  $i$ -th and the  $j$ -th variables are excluded [and store the deviance]. Store the deviance differences attributable to excluding each variable  $j$  from the explanatory set. Increment the deviance differences over each  $z$  to give all the deviances. End cycles.

Collate results:

Finally compute  $w_{ij} = \max(\text{dev}_{ij}, \text{dev}_{ji})/(2n)$ , to give the estimated divergences between variables on the right scale.

## Graphical display

A graphical representation of this information is based on defining the properties of the nodes and edges. The location of nodes usually depends on the scientific context of the study, and consequently this rules out the use of edge-length to display divergence information. We choose to make both edge-tone and edge-width linearly proportional to divergence information, with edge-colour constant. For a single graph, the maximal values of edge-tone and edge-width can be set to the maximal edge divergence. However for comparisons it is important to fix the maximal value in advance.

The layout routine used is part of the dot software of GraphViz (2004). The node locations are constrained to 'rows' of the graph specified in terms of equivalence classes. This determination of the node location allows easy comparison of graphs from using the same subset of variables, but perhaps different subsets of units.

Other alternative displays might be suitable. For instance, edge-strengths may be divergences truncated at a putative significance level. This would correspond to a crude model selection procedure in which an edge is excluded on the basis of its 'last to enter' distribution.

## 5. Example: women's smoking

Socioeconomic influences on women's smoking in early adulthood, Graham et al (2006), examined the socioeconomic patterning of cigarette smoking among women aged 20-34. It established that 'smoking risk was strongly patterned by the socioeconomic life course: by socioeconomic circumstances in childhood, educational pathways and current SEP. Women's domestic life course also emerged as an important, and separate, pathway of influence on smoking status, with women who became mothers in their teenage years and women who were lone mothers more likely to smoke, and to smoke heavily, than those in better circumstances.' Further research at Lancaster on women with partners, taken from data collected by the General Household Survey (GHS), reinforced these findings. The methodology was based on fitting logistic regressions to response variables that measure smoking behaviour. Here we display some of the divergence weighted graphs that were useful in an exploratory analysis of the data set.

### GHS

The material reported here is taken from the GHS website, GHS (2007). The GHS is a multi-purpose national survey, which has been running since 1971, that collects information on a range of topics from the general population living in private households (hh) in Great Britain. It is a repeated cross-sectional study (with data collected from individual face-to-face interviews) that randomly samples addresses from the postcode address file throughout the country. The GHS design is a stratified two-stage sample design. The primary sampling units are postcode sectors which are allocated to 30 major regional strata. The stratification factors are based on an area classifier and selected indicators from the 1991 census. The response rate is usually upwards of 70%, and for instance, in 2005, 17184 selected hhs led to 12271 interviews. The GHS reports a weight variable with each household to be used to compensate for differential non-response.



### Extracted data from GHS:

The extracted records of all 5271 women in the 20-34 age range in the three surveys in the years 2001/02, 2002/03, and 2003-04, with information on roughly 20 variables, and additional information on their partners if they live at the same address, are summarised by

	sampyr			
pind	01-02	02-03	03-04	total
0	792	779	807	2378
1	996	851	1046	2893
total	1788	1630	1853	5271

### Measured and recoded variables

A subset of variables of interest, recategorised for this analysis and excluding cases with missing values, are

smk3, edage3, edlev3, sep4, tenure3, chnum4, age4, pind,

with similar information on the partner when the indicator pind = 1. The category definitions are

smk3: reported smoking status : 0=current; 1=ex-smoker; 2=never

age3: age in years of women : 0=20-24; 1=25-29; 2=30-34

page4: age in years of partner : 0=20-24; 1=25-29; 2=30-34; 3=35+

sep3: socio-economic position: : 1=managerial/professional; 2=intermediate;  
3=routine/manual

edage3, pedage3: age left FT education : 0=16-; 1=17-20; 2=21+

edlev3, pedlev3: education level : 0=A-level+; 1=GCSE-level; 2=other

marstat4: marital status : 0=single;

1=married (living with partner); 2=married (separated from partner); 3=divorced or widowed

tenure3: household tenure : 0=owners; 1=social renters; 2=private renters;

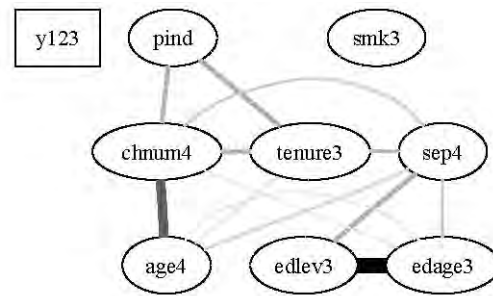
chnum4: number of children (16-) in hh : 0=no child; 1=one child; 2=two+ children

psmkev: partner's cigarette smoking status : 0=never smoke; 1=ever smoked;

Certain variables have not been used, for instance, light/heavy smoker and age of mother at 1st child.

### The divergence weighted independence graphs

There are 8 variables in the graph. The square label just indicates that the data combines the information from the three survey years. The edges are displayed in relation to a maximum information divergence of 0.2.



The remarks to make are that :

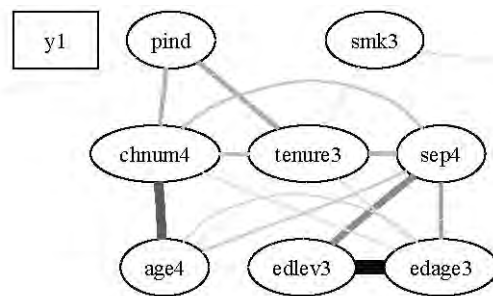
The smoking variable *smk3* is relatively weakly connected to other variables, but its neighbours are (*pind*, *tenure3*, *pedage3*).

The strongest edges are entirely expected: between educational level and age of leaving education (*edlev3*, *edage3*); and those between mothers age and number of children, (*age4*, *chnum4*).

The graph contains two subsets of moderately interacting variables that (*pind*, *chnum4*, *tenure3*) and (*sep4*, *edlev3*, *edage3*), which again is predictable.

Rerunning the same analysis but calculating the deviance from logistic regressions with weights proportional to the GHS variable *weight01*, led to very similar edge strengths.

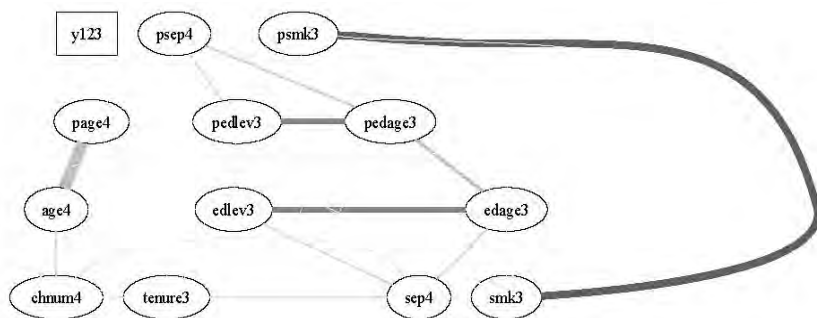
Rerunning the same analysis for the first sample, with roughly 1/3 of the data gives



which demonstrates the stability of the graph, but also indicates the divergences are subject to some sampling variation.

### Including partner information

For those women who have partners in the same household more variables are available.



Obvious partner variables, such as age, are directly connected to the corresponding woman variable. Here the strengths are scaled in relation to a divergence of 0.3. This graph confirms the conclusion that smoking status among women who live with a partner, is mainly

associated with partners' smoking status, Graham et al (2006), Reassuringly the previously observed triangular interaction, (sep, edlev, edage) for women is mirrored in the partners.

## 6. Discussion

We have shown how to extend conditional independence ideas and independence graphs to sample survey analysis. In particular we believe such graphs provide a useful contribution to the exploratory analysis of sample survey, where an overview of interaction and dependency in measured variables is required. When the order is pre-specified the conditioning sets for the independence statements can be modified to handle directed graphs and their extension to chain graphs.

The challenge has been to develop these ideas within the paradigm of design based analysis. We have shown by example that the logical relations determined by definition of variables and the sample design, may be treated by selecting sub-samples and using weighted regression. To complete this exercise we need to address the repeated sampling properties of the divergence measures.

There is no probability modelling, no fictitious distributions, such as the multivariate normal, nor are such models imported through super-population ideas or latent variables. There is no likelihood inference. An advantage over standard graphical modelling is that the computational complexity and inferential vagaries of model search are avoided

The computation employs a common framework of binary logistic regressions. It scales according to  $n$  and  $k$  squared, and so is now feasible for large  $n$  and moderate  $k$ , due to the availability of good algorithms in software such as R and GraphViz.

## References

- Agresti, A. (2002). *Categorical Data Analysis*. Wiley, New York, 2nd edition.
- Cochran, W.G. (1977). *Sampling Techniques*. Wiley, New York, 3rd edition.
- Cover, T. and Thomas, J. (2002). *Elements of Information Theory*. John Wiley and Sons, 2nd edition.
- Darroch, J.N., Lauritzen, S.L., and Speed, T.P. (1980). Markov fields and log-linear interaction models for contingency tables. *Ann. Statist.*, 8:522--539.
- Edwards, D. (2000). *Introduction to Graphical Modelling*. Springer, 2nd edition.
- General Household Survey, (2007). [www.statistics.gov.uk/StatBase/Product.asp?vlnk=5756](http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=5756).
- Graham, H., Francis, B., Inskip, H., and Harman, J. (2006). Socioeconomic life course on women's smoking status in early adulthood. *Journal of Epidemiology and Community Health*, 60:228--233.
- GraphViz, (2004). Graph Visualization Project. AT&T Research. {[www.graphviz.org](http://www.graphviz.org)}.
- Hosmer, D. and Lemeshow, S. (2000). *Applied Logistic Regression*. Wiley, New York, 2nd edition.
- Kullback, S. (1959, 1968). *Information Theory and Statistics*. John Wiley and Sons, Inc., New York.

- Kullback, S. and Leibler, R.A. (1951). On information and sufficiency. *Ann. Math. Statistics*, 22:79--86.
- Lauritzen, S. (1996). *Graphical Models*. Oxford University Press, Oxford.
- Wermuth, N. and Cox, D. (1996). *Multivariate Dependencies*. Chapman Hall, London.
- Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Wiley, Chichester.

## About the Authors

Joe Whittaker, [joe.whittaker@lancaster.ac.uk](mailto:joe.whittaker@lancaster.ac.uk), is a Senior Lecturer in the Department of Mathematics and Statistics at Lancaster University. Chung Kao is a Phd student in the same Department.



## **‘Going beyond the fence’: Using projective techniques as survey tools to meet the challenges of bounded rationality**

**Derek Bond & Elaine Ramsey**

### **Abstract**

If the concept of bounded rationality is accepted then there is a clear need to attempt to record the complex and often non-rational behaviour of decision makers. To address this challenge there has been renewed interest in the application of ‘mixed methods’ to evaluate socio-economic policies. However, it is widely recognised that the normal ‘mixed method’ approaches – using standard quantitative survey techniques supported by qualitative methods such as semi-structured interviews – often fail to measure or investigate issues ‘outside of the fence’. The aim of this paper is to consider whether these challenges can, in part, be addressed by including projective techniques within the ‘mixed methods’ approach.

Projective techniques have been used in psychology for many years and have more recently been adapted for use in market research. However their use as a qualitative survey tool in general socio-economic research has been extremely limited. This could be due to a number of reasons including: substantial overheads, scepticism about the general validity of the techniques and the possible ambiguities in the interpretation of the responses. In this paper the results of an international pilot study into their use in assisting the evaluation of policies is outlined. The study concerned the response of small businesses to governments’ policies aimed at encouraging the adoption of ICT. The lessons learnt from this pilot study about the use of projective techniques as part of a ‘mixed method’ survey methodology are then explored. In particular the paper considers how ICT might be used to reduce the overheads involved in implementing the technique. For example, is it feasible and appropriate to use the web to deliver the instruments? The paper concludes by proposing a research agenda that needs to be addressed before projective techniques might become part of the standard qualitative survey toolkit in general socio-economic research.

### **Keywords**

Policy Evaluation, Projective Techniques, Bounded Rationality, ICT

### **1. Introduction**

With the decline in the importance of the positivist’s paradigm and the increasing acceptance of the limitations of the traditional Neyman-Fisher approach to statistical inference the need for innovative ‘mixed methods’ research in the social sciences is becoming increasingly recognised. This is especially true in the area of policy evaluation. Here the inability of traditional quantitative and qualitative survey techniques to uncover satisfactorily the attitudes

toward and impact of policies is increasingly being discussed. One way of explaining these shortcomings is to place the discussion within the concept of 'bounded rationality' (Edwards (1954) and March and Simon, (1958)). Adopting this framework is equivalent to accepting that much decision making is not entirely rational. Thus, traditional techniques aimed at measuring rational behaviour are of limited use. By implication what is needed are measurement methods that work outside of the 'bounded-rationality'. One possible option here is the use of projective techniques.

Projective techniques, epitomised by the (in)famous Rorschach ink-blot test have a long pedigree in psychological research and more recently have been used in marketing and educational research. Recently their usefulness in policy evaluation has been tested by an international pilot study on attitudes towards and impact of Government policies to e-business. The results of the pilot were encouraging though it raised many issues that need to be addressed before such techniques could become part of a standard mixed method approach to policy evaluation. This paper considers how using computers to assist in the implementation and analysis of projective techniques could help to address some of these issues.

Whilst projective techniques have been used in educational research to measure students' reactions to computer assisted learning there appears to be little research or literature on the use of computers in projective techniques. The aim of this paper is to suggest issues that might form the basis of a research agenda for the area. In the next section the concepts of bounded rationality and projective techniques are discussed. The opportunities and challenges that provide the area of policy evaluations are then examined. This is followed by the main part of the paper which details the role that computers might play in the application of such techniques. The paper concludes with a discussion of the way forward for such an approach to policy evaluation.

## 2. Bounded Rationality and Projective Techniques

The concept of 'bounded rationality' was introduced and developed in Edwards (1954), Simon (1957) and March and Simon (1958). The development of the concept was, in part, in response to the practical limitations of neoclassical economics', rational and utility maximising 'economic man'. Arguing that rational decision making is bounded by the limitations of the human mind and the structure within which the mind operates, it is most unlikely that 'economic man' can exist. Bounded rationality suggests that if social scientists want to understand 'real' decision behaviour, they have to focus on perceptual psychological and cognitive factors that 'cause' imperfect actions. Bounded rationality can be viewed as an attempt to dismantle the 'black box' mentality often used to explain stimulus/response strategic decision making. Such processes are extremely complex, ambiguous and multifaceted, (Mintzberg et al, 1976; Mason and Mitroff, 1981). Therefore what is needed is the ability to measure the heuristics used in decision making. A possible approach is the use of projective techniques.

The concept of projection originated with Freud's work on paranoia in 1911. The techniques were initially used for personality assessment and psychoanalytic treatment (for a light hearted review of its early use see Bellak, 1992) with the usefulness of the techniques in consumer/market and education research being more recently established (see Branthwaite (2002) and Caterall and Ibbotson (2000) as examples). Despite the early work of Rogers and

Beal (1958) their use in a more general business research context has been extremely limited; a recent application is Ramsey et al, (2006).

The main attraction of projective techniques is their apparent data elicitation capabilities. They are subject-oriented, non-verbal and indirect self-reporting techniques that have the ability to capture responses from people in a less structured and more imaginative way. Viewed as a 'face-saving' method, projectives provide the defence mechanisms that enable individuals to unconsciously attribute their personality traits and impulses to others. They can be viewed as complex and sophisticated qualitative data gathering techniques that go beyond standard interviewing, (Patton, 2002). They are used to uncover, rather than measure, feelings, beliefs, attitudes and motivations that participants would otherwise find difficult to articulate (Boddy, 2005). There are conflicting views on how successful projective techniques are at data elicitation. Whilst some (see for example, Rogers and Beal 1958; Catterall and Ibbotson 2000; Lilienfeld, 2000) praise them, others have derided their scientific shortcomings with critiques viewing them as naive and inappropriate (Westen et al, 1990; Yoell, 1974).

There are many different types of projective techniques. Stimuli range from those that are very ambiguous and unstructured at one extreme, to those that are very structured, clear and definite at the other. Rorschach inkblots, present a wide range of alternative choices and the subjects may choose their own interpretations, is an example of an ambiguous stimuli. Semi-ambiguous pictures and sentence completion techniques represent an intermediate position on the stimulus continuum. The more structured projective techniques, such as word association, can be seen as being related to traditional categorical survey methodologies. It is generally argued that the more unstructured and ambiguous projective techniques are the more the subjects will project their emotions, motives, attitudes and values. At the same time the presentation of ambiguous stimuli in isolation of research context can be challenging for the participant to complete.

### 3. Using Projective Techniques

Increasingly projective techniques are seen as part of a 'mixed methods' research methodology. They help the analysis of the research issues from multiple angles, sources and variables of expression, (Lincoln and Guba, 1985; Patton, 2002). A simplistic view is to see projective techniques as the final stage of a three stage process. The first stage being a standard self completed survey questionnaire at a 'macro level'. The second being in-depth interviews at a micro-level using face to face interviews. By implication then provide a nano-level approach delving into issues more deeply than is possible with the other approaches. Each stage is more resource intensive and hence less respondents are involved.

In policy research it is generally accepted (Bennett and Robson 1998; Storey 1994 and Kai-Uwe Brock 2000) that quantitative analysis provides only a limited understanding of the contextual issues and circumstances. Qualitative techniques, such as in-depth interview, are seen to begin to add richness to the research process, via explicit verbal dialogue. However the depth of information will depend on how the respondents rationalise the 'line of questioning' (Hoskin (2002). Qualitative approaches often prove to be insufficient to obtain much insight into the issues raised by quantitative studies. In face-to-face interviews respondents' often constrain their discussion of the issues and seem to provide answers that are of a 'politically correct' nature because they are what they



expected the interviewer to want. In many studies the non-verbals observed give the impression of being uncomfortable about discussing the issues. Early indications are (Ramsey and Bond, 2007) that projective techniques can overcome these data elicitation problems in policy research. They seem able to tap into the deep and hidden layers of the psyche (sub-conscious, cognate thought processes) that are inaccessible via direct questioning; and this is where the 'sweetest' bits of information can be found. One potential bone of contention in the use of projective tests is the issue of response barriers. Fisher (1993) found that indirect questioning reduced social desirability bias on variables subject to social influence. While Broderick & Renwill (1996) found that cartoon completion methods produced less respondent embarrassment. Thus by imagining how others might think or feel in a given situation, response barriers are overcome because the respondents have a research mechanism that allows them to depersonalise, distance and disassociate their cognate thoughts.

The projective technique typology of Lindzey (1959) is useful in deciding on the techniques to use. This typology focuses on 'the nature of the response' and details five projective alternatives that include choice ordering; associative; completion; construction; and expressive techniques. Choice ordering is generally viewed as being of little added value as it involves ranking and ranking choices can be obtained from survey questionnaires. Construction and expressive techniques (see Burns and Lennon, 1993; Gordon and Langmaid (1988); and Lindzey (1959) for explanation of research approaches associated with construction and expressive techniques) are demanding for participants to complete. This leaves association and completion techniques as perhaps the main types to be applied in policy research.

Associative techniques are epitomised by word association methods. These utilise language selection procedures giving researchers a snapshot of the cultural frame that structures meaning. Through the use of scenarios to help build upon a particular frame of reference, these techniques are based on the literal connections that people make when they are building up a narrative. They are simple tools that enable researchers to elicit negative and positive perceptions that may be held towards the concept of interest. The weakness of word association is the inability to project perceptions to a third person. Words can have explicit as well as ambiguous meanings. However, word association can be extremely useful as it demands little from the subjects.

Completion techniques require the participant to complete an incomplete sentence, story, argument or conversation. It (completion technique) depends on creating a context. For example, bubble cartoons are viewed as a simple way to make the context visible. For this method a researcher would produce a cartoon drawing and fill in a speech bubble to initiate a conversation, argument, thoughts, perceptions, feelings and emotions. The participant can respond by completing the speech and thought bubbles provided within the cartoon. It requires more complex and controlled intellectual activity on the part of the subject, penetrating issues to a 'deeper' level than word association, with careful consideration given to scenarios built around the cartoon.

Word association tests give a simple way for respondents to consider their perceptions. Whereas the way the respondents interpret cartoons revealed much about their own emotions, feelings and perceptions. Visuals help to focus on the 'heart' of the subject, perceptions. This serves to minimize the 'politically correct' discourse that interviews tend to produce. Projective techniques make little or no demand in terms of literacy or academic skills on the part of the respondent. They are predominantly visual, largely independent of particular

languages, and in some cases do not involve speech at all. Their scope is therefore much wider than that of self-reporting or rating procedures (Berkman & Gilson, 1986). Projective techniques appear to be able to address both the widely varying decision processes and individuality of respondents (Zheng et al, 2004). Thus the use of these techniques could provide some of the tools toward a fuller understanding of the way people respond to policies.

#### 4. Limitations of Projective Techniques

Whilst it seems that projective techniques have an active role to play policy research a pilot study (c.f. Ramsey and Bond, 2007) highlights many issues that need to be addressed. The issues include: design of stimuli, reliability and validity, methods of distribution, methods of recording and storing responses, methods of analysis and interpretation.

##### Design

Researchers rarely explain how the projective instruments are derived and the stimulus materials designed (Schlackman, 1989). However, designing stimuli can be complex and many issues need to be considered. For example, visual stimuli could lead to gender biases. Therefore, Martin and Kennedy (1994) suggest that researchers should employ simple 'stick' people in cartoons to avoid the risk of gender biases. This recommendation may also alleviate the demands on researchers' creative/artistic talents, is in keeping with the ambiguous nature of projective techniques and may have the potential to make the cartoons more culturally amenable to global samples. Ramsey et.al. (2006), however, suggest that since all individuals will innately house many sub conscious biases, stick people will not have the capacity to negate gender biases. They argued that since reality is socially constructed cartoons should contain 'real' representations of people. However, the cartoons should not be over complex, dull, abstract, hard to believe, difficult to apply or irrelevant. They suggest the following criteria be applied in their construction: interesting (illustrated, concrete examples), intelligible (choice criteria clear), plausible (history, consistency, worldview), relevant to key issues and ultimately useful to achieve the aims and objectives of the research. It is therefore very laborious/time-intensive finding appropriate characters/drawings to create various cartoons

##### Reliability and Validity

One of the main limitations, noted by Seitz (2001), is the risk that data outcomes can reveal more about the interpreter than about the psychological motivations and personality structure of the respondent. Thus a considerable degree of subjectivity applies in the interpretation of responses to projective techniques with researchers frequently disagreeing among themselves (Sampson, 1986 and Churchill, 1991). The general weakness is that if two equally competent interpreters interpret the data in a 'significantly' different manner; this may call into question the reliability and validity of the technique (MacFarlane and Tuddenham, 1951). Krathwohl (1993) suggest that explanation credibility; demonstrated results, translation validity, and credible results would instil confidence in readers that the research instruments are fallible indicators of the latent constructs they are intended to 'assess'. Ramsey et.al. (2006) suggest the research mechanisms in Table 1 to ensure reliability (Burns and Lennon, 1993) and validity and hence, the interpretability of the findings gleaned from this qualitative stage of the research (Creswell, 2003).

## Delivery

Traditionally individual face-to-face or focus groups are the tried and tested methods for administering projective techniques. It has been suggested that one can gain most from the techniques in a face to face situation and if the researcher presents with a warm and friendly rapport they will be best equipped to elicit the best quality information from their participants (See for example, Will et al, 1996). However, such an approach is resource intensive and given that the respondents of policy evaluation studies are often extremely busy senior business persons such an approach can lead to low participation rates. As an alternative, Ramsey et.al. (2006) distributed the instruments via post for completion. It was thought that the postal method would be less intrusive and provide participants with the alternative to complete the tests at a time 'convenient' to them and without the added 'peering pressure' from the researcher anxious to record every word and body movement. However, the researchers did not attempt to assess the effects of using postal distribution had on their results. They reported that from a process perspective, the distribution strategy did not present the research team, or the participants with any major problems. Participants were able to follow instructions associated with filling in the speech and thought bubbles of the completion tests and no complications were apparent, with a very satisfactory response rate being achieved.

Table 1. Research mechanisms to ensure reliability

Type of validity /reliability	Definition	Assessment Criteria
(i) Internal Validity (Demonstrated results)	Measurement accuracy	Rigour of development approach
ii) Content Validity (Translation validity)	Does the content reflect the intended concept	Pre-pilot-test projectives Findings from stage 1of research process
iii) Construct Validity (Explanation credibility)	Empirical and theoretical meaning, relative to constructs	Pre-pilot work Data analysis
iv) Reliability (Credible results)	Measure is reliable to the extent that independent but comparable measures of same construct agree	Data analysis of both qualitative instruments facilitated triangulation

## Analysis and interpretation strategy

A major disadvantage of projective techniques is the complexity of the data and the corresponding skills required of the researcher (Burns & Lennon, 1993). Responses are not easily codified and need careful evaluation by researchers who are both trained and skilled interpreters of information. It has also been argued by some (Colwell, 1990 as cited in Boddy, 2005) that to enable the competent analysis of projective techniques, one should perhaps have a degree in psychology or the behavioural sciences. In terms of analysis and interpretation of projective tests, Boddy (2005) has also voiced concerns about the lack of insight on how the

data derived from projective techniques are analysed. Ramsey et.al. (2006) employed analytical methods to establish rigour in the interpretation and analysis of the projectives study. In brief, a systemic quasi-quantitative analytical approach using context and matrix analysis was utilised to maintain rigour and also to enable cause and effect relationships among variables to be established. Attempts at a rigorous analysis are highly resource intensive in nature.

## 5. The Role of ICT in Projective Techniques

Despite projective techniques finding a niche in the area of assessing the impact of ICT in education there appears to be little or no research on the use of ICT to aid their delivery. This is odd as in most survey areas ICT has provided a way of addressing resource issues and many of the problems highlighted in the Ramsey et.al. (2006) pilot study are related to resource issues. Therefore, it seems reasonable to consider how ICT might help make projective techniques more effective and efficient research tools.

### Design

The Ramsey et.al. (2006) pilot study highlighted the resource issues associate with the design of projective instruments. There does not appear to be a depository either online or offline for such instruments. Within the United Kingdom the UK Data Archive at the University of Essex holds extensive documentary material relating to the datasets they hold. However, this documentation does not appear to contain any material relating to projective techniques. There would appear to be a need for such a repository if projective techniques are to be used.

The ready availability of computer aided design (CAD) packages should mean that it is possible to develop visual stimuli easily. However, as with much computer produced output there is an obvious quality issue here. Visual stimuli produced using CAD could be of very low standard and contain many cognitive flaws.

Computers also provide the possibility of design new stimuli. Rather than just static cartoons the possibility of producing video clips etc. need to be investigated. The appropriateness of a 'please say what happens next' type stimuli in completion tests has yet to be investigated.

With the development of context translation tools the possibility of using CAD to develop multi-lingual stimuli is becoming a possibility.

The way respondents record their responses also needs to be investigated. Not all respondents will be ICT literate and care must be taken in the design of the instruments to ensure that cognitive issues are considered

### Reliability and Validity

The issue of the reliability and validity of the projective techniques employed can be assisted by the use of ICT. Delivery of instruments using ICT (see below) could be more objective and remove interviewer biases. Moreover, with responses digitised it should be easy to ensure that reliability and validity are maintained. Different evaluator's views of the response can be quickly shared and compared. Also with digitised responses they can, hopefully, be analysed in a more objective way using computers reducing the risks of subjectivity in the interpretation of the findings.

## Delivery

Traditional face to face delivery of projective instruments is resource intensive. The use of postal instruments seemed to work in the Ramsey et.al. (2006) pilot study however, this was not assessed. Using ICT for the delivery of the instruments offers many opportunities for resource savings though there are obvious risks. Delivery could range from interviewers providing respondents with instruments on a notebook computer in their presence to completely remote delivery with no interviewer interaction using the internet. The growth of VOIP and video over internet together with the growth of smart-phones, that can handle video and internet connections, offers many opportunities that need to be investigated.

Using the internet to deliver the instruments could also be problematic. When faced with a stimulus that requires a carefully considered response, it is likely that some respondents will use internet tools, such as Google, to assist in them deciding upon their response. Such possibilities, whether they need to be considered and what procedures need to be developed all need to be considered.

At a naïve level it is possible to envisage that using ICT to deliver the (multilingual) stimuli could lead to a more complete coverage of responses. In policy evaluation this is a major issue and relates back to the reliability and validity issues. Unfortunately, in practise this is not likely to be the case. The problem is while ICT is likely to reduce the resources needed for administering to each respondent the enlarging of the number of respondents geographically is likely to give rise to all sorts of cognitive and cultural issues. These aspects need to be investigated further before any firm statements can be made.

One of the main disadvantages of a computer mediated distribution strategy is that it assumes that people have access to technology and the expertise to use it well enough to complete the research tasks set them. The projective instruments may appear confusing to the person they are directed to. However, this can be overcome to a certain extent if used as part of a mixed methods approach. In this way the participant already has knowledge of the context in which the research is being done. Also it would be pertinent to include a letter of instruction with regard to how the projective tests should be completed.

## Analysis and interpretation

The pilot study by Ramsey et. al. (2006) found attempts to rigorously analyse the data derived from projective techniques is resource intensive. This is not a unique problem of projective techniques, most qualitative methods produce large amounts of data that are time consuming to analyse. In recent years the development of packages such as Atlas/ti and Nud\*ist/NVivo has helped reduce the resources needed for the general analysis of qualitative data. Little work appears to have been done on how such packages can deal with the data from projective instruments. This is again an area for further study.

## 6. Conclusions

In this short paper the issue of using developments in ICT to address some of the resource issues associated with possibly using projective techniques in policy evaluation studies has been considered. For many people involved in traditional quantitative and qualitative research the usefulness and appropriateness of projective techniques have yet to be proven (c.f. Boddy, 2005). In this paper, building on the findings of a pilot study by Ramsey et. al. (2006), the

issues of how developments in ICT might address some of the concerns over the appropriateness of projective techniques has been addressed.

The main conclusions are that ICT could play an important role in most areas of the use of projective techniques including:

- Design of instrument
- Reliability and Validity
- Delivery
- Analysis and Interpretation

In all these area though much more research is needed before the appropriateness of (ICT based) projective techniques for policy evaluation can be assessed fully.

## References

- Barry, K., (2003), 'Factors limiting the transfer of knowledge and know-how: a European perspective – the survey process in the digital age', ASC Conference 2003.
- Bellak, L. (1992), Projective techniques in the computer age, *Journal of Personality Assessment*, Vol.58, pp.445-453.
- Bennett, R., and Robson, P., (1998), 'External advice and business links', in A. Cosh and A. Hughes (Ed.) *Enterprise Britain, Growth, Innovation and Public Policy in the Small and Medium Sized Enterprise Sector 1994-1997*, ESRC Centre for Business Research, Cambridge.
- Berkman, H.W. and Gilson, C. (1986), *Consumer behaviour, concepts and strategies*. 3rd ed. Boston. Kent.
- Boddy, C. (2005) Projective techniques in market research: valueless subjectivity or insightful reality? *International Journal of Market Research*, Vol.47, No.3.
- Branthwaite, A., (2002), 'Investigating the power of imagery in marketing communication: evidence-based techniques', *Qualitative Market Research: An International Journal*, Vol.5, No.3, pp. 164-171
- Broderick, A.J., and Penwill, S.J., (1996) 'Cartoon-completion or bust? Researching sensitive issues in social marketing', in: *Proceedings of Marketing Education Group Conference*, Track 6, pp. 11-18 Glasgow, University of Strathclyde
- Burns, L.D., and Lennon, S.J., (1993), 'Social perception: methods for measuring our perception of others', *International Textile and Apparel Association Special Publication*, Vol.5, pp. 153 –159
- Catterall, M., and Ibbotson, P., (2000)., "Using projective techniques in education research", *British Educational Research Journal*, Vol. 26, No. 2, pp 245-256
- Churchill, G.A., (1991), *Marketing research methodological foundations* (5th Ed.) Chicago: Dryden Press.
- Colwell, J., (1990), 'Qualitative market research: a conceptual analysis and review of practitioner criteria'. *Journal of Market research Society*, Vol. 32, No.1, pp. 13-36

- Creswell, J.W., (2003), *Research Design: Qualitative, Quantitative and Mixed Methods Approaches*. (2nd Ed), Thousand Oaks, California: Sage
- Edwards, W., (1954), 'The theory of decision-making', *Psychological Bulletin*, Vol. 51, pp. 380-417
- Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research*, 20, 303-315.
- Freud, S., (1911), *General psychological theory: Papers on Metapsychology*. New York: Collier Books
- Gordon, W., and Langmaid, R., (1988), *Qualitative market research*. Gower: Aldershot
- Hoskin, F., (2002), 'How do we ask questions and how do we know what it all means?' [online] available at URL: [http://www.nop.co.uk/news/PDE/BIG\\_Conference.PDF](http://www.nop.co.uk/news/PDE/BIG_Conference.PDF) accessed on 9th September 2002
- Kai-Uwe Brock, J., (2000), 'Information and technology in the small firm' in S.Carter and D. Jones-Evans (Eds.), *Enterprise and Small Business: Principles, Practice and Policy*. Financial Times: Prentice-Hall.
- Krathwohl, D.R. (1993), *Methods of educational and social science research: an integrated approach*. White Plains, New York: Longman
- Lincoln, Y.S., and Guba, E.G. (1985), *Naturalistic Inquiry*. Beverley Hills, CA., Sage.
- Lindzey, G., (1959), "On the classification of projective techniques", *Psychological Bulletin*, Vol.15, pp.158-168
- MacFarlane, J.W., and Tuddenham, R.R., (1951), "Problems in the validation of projective techniques". In: H.H. Anderson and G.L.Anderson (Eds), *An Introduction to Projectives*, Englewood Cliffs: Prentice-Hall.
- March, J. and Simon, H. (1958) *Organisations*, New York: John Wiley and Sons
- Martin, M.C., and Kennedy, P.F. (1994), "The measurement of social comparison to advertising models: gender gap revealed", In Costa, J.A. (Ed) *Gender Issues and Consumer Behaviour*. Thousand Oaks, CA, Sage.
- Mason, R.O., and Mitroff, I.I. (1981), *Challenging strategic planning assumptions*. New York: Wiley
- Mintzberg, H., Raisinghani, D., and Theoret, A., (1976), "The structure of unstructured decision-making", *Administrative Science Quarterly*, Vol.21, pp. 39- 58
- Patton. M.Q. (2002), *Qualitative Research and Evaluation Methods*, (2nd Ed).pp. 551-573
- Ramsey, E. and Bond, D. (2007): 'Evaluating Public Policy Formation and Support Mechanisms for Technological Innovation' *International Review of Applied Economics*, Vol.21 No. 3 July forthcoming.
- Ramsey, E., Ibbotson, P., and McCole, P., (2006). 'Application of projective techniques in an e-business research context: a response to "Projective techniques in market research – valueless subjectivity or insightful reality?"'. *International Journal of Market Research*, Vol. 48, Issue 5,

- Rogers, E.M., and Beal, G.M., (1958), "Projective techniques in interviewing farmers", *The Journal of Marketing*, Vol. 23, pp.177-183
- Sampson, P., (1986), 'Qualitative research and motivation re-Search'. In, R.M. Worcester, and J. Downham, 1986. *Consumer market research handbook*. 3rd ed. Amsterdam. Elsevier.
- Schlackman, B., (1989), "Projective tests and enabling techniques for use in market research", in S.Robson and A.Foster (Eds) *Qualitative research in action*, (pp. 58-75), London: Edward Arnold
- Seitz, J.A., (2001)., "A cognitive'perceptual analysis of projective tests in children"., [online] available at URL: <http://www.york.cuny.edu/~seitz/bio.html> accessed on 9th September 2002
- Simon, H. (1957) "A Behavioral Model of Rational Choice", in *Models of Man, Social and Rational Mathematical Essays on Rational Human Behavior in a Social Setting*. New York, Wiley
- Storey, D.J., (1994), *Understanding the small business sector*, London: Routledge,
- Westen, D., Lohr, K.R., Silk, L., and Gold Kerber, K., (1990)., "Object relations and social construction in borderlines, major depressives and normals: athematic apperception test analysis". *Psychological Assessment*, Vol.2. pp. 355-364
- Will, V., Eadie, D., and MacAskill, S., (1996), "Projective and enabling techniques explored". *Marketing Intelligence and Planning*, 14, pp. 38-43
- Yoell, W.A.,(1974)., "The fallacy of projective techniques"., *Journal of Advertising*, Winter, pp33-35
- Zheng, J., J., Caldwell, N., Harland, C.M., Powell, P., Woerndl, M. & Xu, S, (2004), 'Small firms and e-business: cautiousness, contingency and cost-benefit', *Journal of Purchasing and Supply Management*, Vol.10, pp.27-39

## About The Authors

**Derek Bond** is a senior lecturer in the School of Business, Retail and Financial Services at the University of Ulster at Coleraine. In addition he is a consultant on call to various Government Departments and has worked as an EuroStat expert in the accession countries. For the last four years he was chairperson of the International Statistical Institute's Standing Committee and Regional and Urban Statistics and committee member of the International Association for Official Statistics. He can be contact on telephone 02870324350 or e-mail [d.bond@ulster.ac.uk](mailto:d.bond@ulster.ac.uk)

**Elaine Ramsey** is a lecturer in the School of Business, Retail and Financial Services at the University of Ulster at Coleraine. Her main areas of expertise are in the areas of SMEs; Tradeable Services; and E-Business. Her particular area of research interest relates to regional economic development and public policy support mechanisms associated with enhancing the competitiveness of small service firms. Research outcomes concern the influence of government policy on innovation, enterprising activity and 'e' entrepreneurial orientation of SMEs. Elaine can be contacted on telephone 02870323335 or e-mail [e.ramsey@ulster.ac.uk](mailto:e.ramsey@ulster.ac.uk)





# Qualitative Data Exchange: Methods and Tools

Louise Corti

## Abstract

This paper will present recent findings from a project that is exploring the feasibility of developing data exchange models and data conversion tools for primary research data collected in the course of qualitative research.

## 1. Summary

A standard format for representing richly encoded qualitative data is necessary because: it ensures consistency across datasets; it supports the development of common web-based publishing and search tools; and it facilitates annotated data interchange and comparison among data collections. Importantly, it should enable data and linked products to be imported and exported directly into and out of CAQDAS packages, avoiding the reliance on just a single product, and offering the opportunity to share analytic workings outside the confines of any particular software. The model and format have been requested by many other social science data archives round the world who are starting to collect qualitative research materials. Likewise there are researchers collecting multi media data in their own research who want to publish and showcases these data online, yet lack a consistent model of representing data that has a longer interpretive shelf life than basic html.

The project is developing, refining and testing models for data exchange for qualitative research data based on XML schema based on METS, while incorporating existing schema. These include: the Text Encoding Initiative (TEI), the Data Documentation Initiative (DDI) and Dublin Core and a system for identifying segments based on stand off annotation. The test data selected for this project are from the social sciences, but these formats are typically found across all domains of primary research. A key sample dataset has been taken the ESDS Qualidata archived project the Edwardians, that hold a range of data types, included data marked up in TEI.

The first step has been to liaise with and seek support from CAQDAS software vendors on a minimal data exchange model. A workable data exchange scenario rests on them providing export to and import from a common intermediate format. At the 2007 CAQDAS conference held at the University of London, a meeting of the majority for the vendors was held at which an initial model and schema was proposed. This was based on what might be termed core CAQDAS functionality, or a set of common denominators that could be replicated in the DExT data model. DExT has termed them **Data Constructs**.

The intermediate standard at the time of writing is likely to be called DEx-XML and QDEx-XML for the qualitative model, with any tools based on the model being known as DExT. The

full METS profile is being termed DEx-METS. The project has defined five major constructs (functions) that are common to all or most of the CAQDAS packages:

*SEGMENTS: Identify Subsets of the study (e.g. Text or Line selections)*

*CODES: Assign Values to a Subset of a study, eg a segment*

*HiCODES: Create a Value Hierarchy (e.g. Codes arranged in a coherent hierarchical structure)*

*FileCLASS: Create a File Hierarchy/file classification (e.g. Files arranged in a coherent hierarchical structure)*

*MEMOS: Assign Notes or Comments (to a segment or a code)*

Once the CAQDAS data have been transformed from vendor XML via XSLT stylesheets to the QDExT-METS, an enhancement interface enables additional metadata will be added. This includes the capacity to identify all the related files in the whole 'study' (given the possibility of complex multi-media collections) and looking at how the files relate to one and other. A viewer that transforms the DEx-METS XML will enable key mark-up of the collection to be demonstrated, though a fully functional viewer or interactive tool will not be produced under this project. It is hoped that if the model and schema are adopted other projects will consider developing user-oriented tools.

The project has engaged the services of an XML consultant from the Open Data Foundation (ODaF) who has helped finalise the first draft of the data model, the XML schema and a data processing interface (based on JAVA).

DExT is also working with intermediate data model and stand for the common statistical packages, SPSS, SAS and STATA. All the models and tools being created under this project are open source under the terms and conditions of the JISC award. The model will be published on Gforce and any tools on Source Forge. A beta release should be available by Autumn 2007.

## References

DEXT [www.data-archive.ac.uk/dext/](http://www.data-archive.ac.uk/dext/)

DEXT WIKI [dawiki.essex.ac.uk/DEXTWiki/](http://dawiki.essex.ac.uk/DEXTWiki/)

Open Data Foundation (ODaF) <http://www.opendatafoundation.org/>

METS [www.loc.gov/standards/mets/](http://www.loc.gov/standards/mets/)

UK Data Archive [www.data-archive.ac.uk](http://www.data-archive.ac.uk)

The Edwardians [www.esds.ac.uk/qualidata/online/data/edwardians/](http://www.esds.ac.uk/qualidata/online/data/edwardians/)

## About the Author

Louise Corti

DExT

UK Data Archive

University of Essex

Colcehst5erf CO4 3SQ

44 1206 872145

corti@essex.ac.uk



## ***'Framework'*, Computer-assisted Qualitative Data Analysis Software and its Role in Increasing Quality and Transparency in the Analysis of Qualitative Data**

Kandy Woodfield & William O'Connor

### **Abstract**

Qualitative researchers need intelligent tools that can assist them in understanding large bodies of unwieldy data. An exponential growth in the use of computer assisted qualitative data analysis (CAQDAS) software has occurred since the mid-1980's. Commentators regularly return to the question of whether, and how, the use of CAQDAS software affects the analytical process and the quality of outputs. This paper revisits those debates and explores how a new generation of software might impact on quality. The paper uses the example of *'Framework'* a pioneering new approach to computer assisted qualitative data analysis due for release in 2008 to debate the key issues. Developed by the Qualitative Research Unit at the National Centre for Social Research (NatCen) in the mid 1980's *'Framework'* is a matrix based tool for qualitative data management. By the time of the ASC 2007 International conference the software will be close to release in both standalone and network versions. The paper will examine some of the innovative parts of the application and discuss various features of the new software that will assist researchers in producing robust, credible qualitative evidence. In doing so the wider issues relating to quality in qualitative research and how far the use of software packages in the analytical process can support and enhance transparency will be explored. Finally, the paper will assess the new generation of software available and demonstrate how innovative technological features are likely to give these new packages a premier place in the qualitative researcher's toolbox.

### **Keywords**

Computer-assisted qualitative data analysis; Framework; quality; qualitative research

### **1. Introduction**

The field of applied policy research remained dominated by quantitative methodologies until the mid-1980's. Qualitative researchers often faced sceptical audiences who were unconvinced about the validity and robustness of findings from qualitative research. These challenges created an impetus for qualitative researchers to search for ways of demonstrating analytical rigour and representing the foundations of quality in the approaches they had chosen. Qualitative data analysis was often viewed as an elusive 'black box' with researchers describing how findings were 'intuitive' or had 'emerged from the data', neither of which

provided that sceptical audience with transparency about the analytical process nor demonstrated how findings had been validly derived from the data collected.

Since this time the qualitative research community has made great strides in addressing quality issues and there are now a plethora of quality frameworks which attempt to provide a means for audiences to make informed judgements about the quality of the research they are reviewing (see for example, Spencer et al, 2004).

Nevertheless, the analytical process remains shrouded in some mystery, not least because of the myriad approaches to analysis adopted by researchers. The emergence and, some would argue, burgeoning dominance of CAQDAS software, has added another layer to the quality debate and what 'analytical adequacy' (Fielding, 2002) consists of in qualitative research.

## 2. The quality debate

Quality is a heavily contested concept amongst qualitative researchers with little consensus about what constitutes quality, and some resistance to developing common frameworks for establishing quality. Unlike in quantitative research there are few shared canons for judging quality and those that do exist tend to relate to the design or data collection phase of studies, not the analytical process. However, judging the 'analytical adequacy' of a piece of research is not a simple science and criteria can not simply be applied in a checklist fashion, they require an informed application by the reader to ensure an appropriate judgement is reached.

Spencer et al's 2004 framework for the assessment of quality in qualitative research specifically addressed what constitutes 'analytical adequacy'. They argued that to be robust and credible research findings need to demonstrate evidence of:

- The use of data management methods which are suited to the analytical task and which preserve the context of the data which is critical to allow audiences to judge the authenticity of the findings
- Portrayal of detail, depth, complexity and diversity in the data, including discussion of outliers and cases not conforming to dominant patterns
- Discussion of explicit and implicit explanations and underlying factors accounting for those explanations
- Identification of patterns of association, for example between circumstances and behaviours, or attitudes and experiences
- Non-numerical treatment of qualitative data
- Clear portrayal of the analytical building blocks which have led to the findings, providing the reader with the ability to critically appraise whether the findings are grounded in the data and credible interpretations
- A transparent audit trail demonstrating evidence of how the data has been handled and interpreted by the research team

### CAQDAS software and the quality debate

As previously noted, the process of qualitative data analysis has been somewhat overlooked in quality debates and it is still common to see researchers themselves spend little time reflecting on how they have reached the conclusions they have drawn. To some extent this

has been exacerbated by the growth in the use of specialist software with some researchers now describing their analytical approach by naming a piece of software. It is common, for example, to read 'the data was analysed using NVivo, N6 or Atlas Ti...' with little further explanation of what that means in terms of the treatment of the data and how findings were generated. In this sense, less experienced researchers appear to mistakenly equate the use of software with adopting an interpretative approach. A useful analogy would be if we were to describe the process of drafting this paper as simply revolving around the word processing software used – 'this paper was written using Word...' rather than the underlying intellectual work undertaken. As Fielding and Lee (2002) noted research has indicated there is the potential for CAQDAS users to confuse 'a technical resource with an analytical approach'. Similarly, there is evidence that some users believe becoming equipped to manage a software package is akin to being a 'good' analyst (MacMillan & Koenig:2004). These confusions have led to concerns that 'learning the software' is taking the place of 'learning the craft of qualitative data analysis' with newer researchers focusing on 'making order' in their dataset (by coding and labelling) rather than recognising the importance of 'making sense' of what that data means and conducting robust interpretative analysis.

Many of the concerns voiced in the early days of CAQDAS were related to specific features of the early packages. These included the fear that the packages would remove data extracts from their context and 'lift' the researcher away from close contact with the stories they had collected in the field. For instance, the early focus on 'code and retrieve' features whereby data was labelled according to a conceptual 'Framework' allowing the package to then retrieve all extracts coded similarly worried researchers who feared the segments of data would lose their analytical potency when removed from their context. Similarly, some argued that restrictive output formats (primarily, the retrieval of coded segments of verbatim data) would make detailed and creative interpretation more difficult, particularly if data reduction was not permitted. Some of the packages themselves made a feature of their ability to output to SPSS or Excel formats which led to concerns that they could encourage inappropriate numerical interpretations. Overall, there was a tendency to fear that the growth of 'code and retrieve' packages would lead to analytical rigidity rather than encourage analytical pluralism (Coffey, Holbrook and Atkinson:1996) particularly as many of the early packages were associated with one specific approach to qualitative data analysis – grounded theory.

However, growing sophistication of the software available, the use of new technologies and an expansion in discussions about quality more generally has meant that many of these concerns have dissipated over time. The use of hyper linking and other data linking technology means that there are now much stronger links to the original data, similarly new technology has helped to widen the range of display and output options encouraging more creative approaches to interpretation. A growth in the use of digital technology has also led to packages being able to sustain more innovative approaches to data handling and most packages now offer much more than the traditional 'code and retrieve' function. So how can this new generation of packages help to support transparency and 'analytical adequacy'?

### 3. Supporting quality and providing transparency – the role of CAQDAS packages

The latest generation of software packages offer many features which can strengthen and support a quality approach to analysis. Firstly, they provide a transparency of methods which



provides the audience, or sponsor, of the research the opportunity to scrutinise what has happened to the data, in what order, and to understand the reasons for those analytical decisions. Traditional paper based approaches to the analytical process often relied on researchers recalling decisions they had made and being able to make those interpretations explicit, now most packages offer systematic audit trails of changes made to the conceptual 'Framework' or to groupings and classifications of the data. This allows the analytical process to become open for review and critical appraisal in a way previously not possible. The audit trail functions vary in sophistication but certainly provide a much greater ability to scrutinise the analytical process.

Similarly, most packages offer a memo or notebook function which allows the user to attach analytical comments to their data as they move through the process, these can provide an important trail of interpretative decisions which can be shared and scrutinised. The use of the packages then provides a level of explicitness to the analytical process not previously seen. Equally the sophisticated archiving potential means that data becomes easily stored and open for review in an accessible way.

The packages also provide a systematic and consistent approach to data searching meaning the analyst can be confident that they have explored all the corners of their dataset rather than simply 'cherry picked' the most striking, dominant or articulate accounts and explanations. Systematic searching without fear of human error can be done swiftly and repeated using different search combinations providing a powerful way to explore the data set from a range of angles. New artificial intelligence and mapping functions can help the analyst to explore patterns in the dataset and search for explanations, generating new theories and hypothesis throughout the analytical process.

As the software has become more sophisticated, tools increasingly provide and encourage analytical pluralism by facilitating different forms of analysis to be undertaken across a range of studies. For example, packages now allow researchers to perform: within and between case analysis; descriptive, associative and explanatory analysis; and, longitudinal analysis. This flexibility counters earlier concerns that the growth of CAQDAS might lead to analytical rigidity, although the creative use of all functions of a package relies as ever on the skill of the user in taking advantage of the full potential of the package. Fielding and Lee's research (2000) with users of CAQDAS packages indicated that much of the full creative potential of these packages was under utilised by researchers who tended to remain wedded to the core code and retrieve functionality rather than exploring other aspects of the software.

Finally, and of growing importance in qualitative research with a greater emphasis on collaborative and cross-disciplinary working, newer versions of packages can offer sophisticated team working facilities allowing groups of researchers to share the interpretative journey, collaborating to produce findings.

CAQDAS software therefore offers great potential to support quality in the analytical process but the caveat remains that all software is a tool in the process and requires careful and informed application by researchers familiar with the key elements of qualitative data analysis. In the hands of the inexperienced or uninformed user even the most systematic computing tool can produce inappropriate or less than credible findings.

#### 4. 'Framework' – a case study in supporting quality and transparency

'Framework' is a matrix based tool for qualitative data analysis developed by the Qualitative Research Unit (QRU) at the National Centre for Social Research in the UK. Key to the approach is that it organises data into a series of conceptual or thematic matrices from which it is possible to conduct case based and thematic analysis. It facilitates the systematic analysis of all qualitative data sets from the straightforward through to the more complex.

'Framework' was conceived in the mid-1980's and has since then developed a reputation as a tool that supports transparency, consistency and quality in the analytical process. It is an approach which has been adopted widely by other researchers both in the UK and internationally. The conceptual framework is developed inductively from the data and the dataset is then systematically summarised into the series of matrices. Summarised data retains a direct link to the original data source. The process provides a systematic and consistent approach to data handling which allows the researcher to reduce the quantity of the data they will need to make sense of without losing any of the richness or complexity of the dataset. The analyst is left with a dataset organised by conceptual themes, reduced but not stripped of its depth, and displayed in an easily navigable matrix output. The actual process of synthesis, or summarisation, facilitates creative thinking and activates the analytical mind by immersing the researcher in their data set allowing them to gain depth without drowning in the volume of data they are working with.

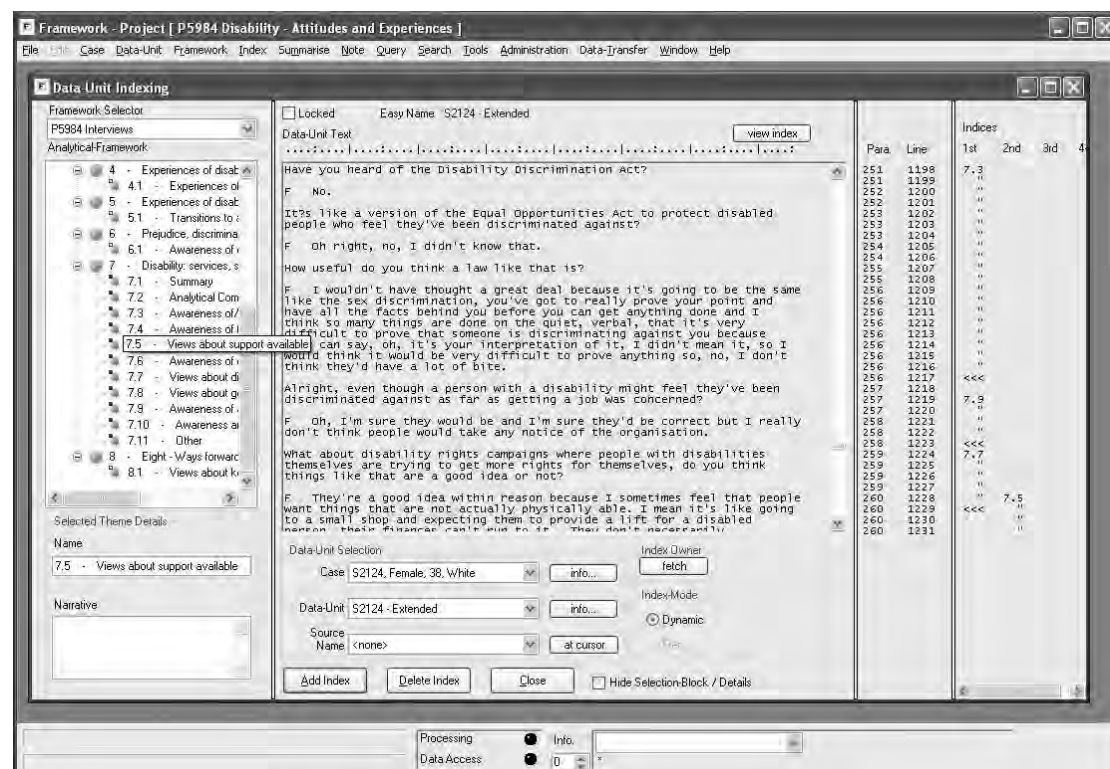
Initially an entirely paper based method, it has evolved over the last twenty or so years into a unique software package, to be released in 2008, which can be run on all Windows based computers. Key to the 'Framework' approach is the creation of summaries of verbatim data which maintain context, language and meaning. Other CAQDAS packages either do not enable this or do it in a slightly cumbersome way. In addition, the unique output of 'Framework' – the thematic matrix – in which these data summaries are displayed by theme and data unit (see below), is not adequately supported by other CAQDAS packages.

	2.1 - Definitions	7.10 - Awareness and responses to See the Person campaign	7.2 - Analytical Comments	7.3 - Awareness of views about legal and civil rights	
	<p>coping with the world around you. Might be people with hearing like mine who have a different temperament whose world would be devastated by it so it's also about the person</p>				
L04W Female, 52, White		<p>Not seen posters(10) It's good a way to get people to learn(10)</p>	<p>TIME RAN OUT AT END OF INTERVIEW SO THIS SECTION QUITE RUSHED</p>	<p>NOT DISCUSSED</p>	<p>Sur col hell in c For spe toil</p>
L04W Female, 52		<p>Have seen some of these posters before seen 'It's tragic' thought it was rubbish, they're excellent. 'See the person' it brings out the idea that yes they are people. They're done in a lighthearted way but the small print makes you see things from a different angle, they are all mentally or physically handicapped b/c that's easy to see SEE CHART A FOR LINKS TO ATTITUDES TO DISABILITY. It's v. difficult to show a blind or deaf person in a still photograph, think the campaign is a good idea(27). You're never going to change everyone's atts. Posters will make some of them more aware but some people will always think that someone who is not in their likeness isn't worth knowing (* 31)</p>		<p>Presumed that disabled people have the same as everyone else gets, not aware of any additional rights (27)</p>	<p>"At call no mai hav. The but the one you no rea</p>

While 'Framework' has been developed to suit the needs of those who wish to use this approach, it has been designed so that it incorporates the key areas of functionality popular in most other CAQDAS packages (such as allowing researchers to code their data units and retrieve coded segments of verbatim data by theme). This means that the package offers a diversity of ways of working with data and that the analytical methods used can be driven by the requirements of the research and researcher. In addition, the software has broadened the applicability of 'Framework' to a wider range of research activities including longitudinal qualitative research, secondary analysis of qualitative data and systematic reviews; basically any activity that requires the navigation of large bodies of textual data. The package is available in two formats. A networked version allows the creation of a secure repository of qualitative data that multiple users can access thus supporting teamwork. The stand-alone version for single users will have the same functionality without the data sharing capabilities, although it will still be possible to export to and import data from other stand-alone version users.

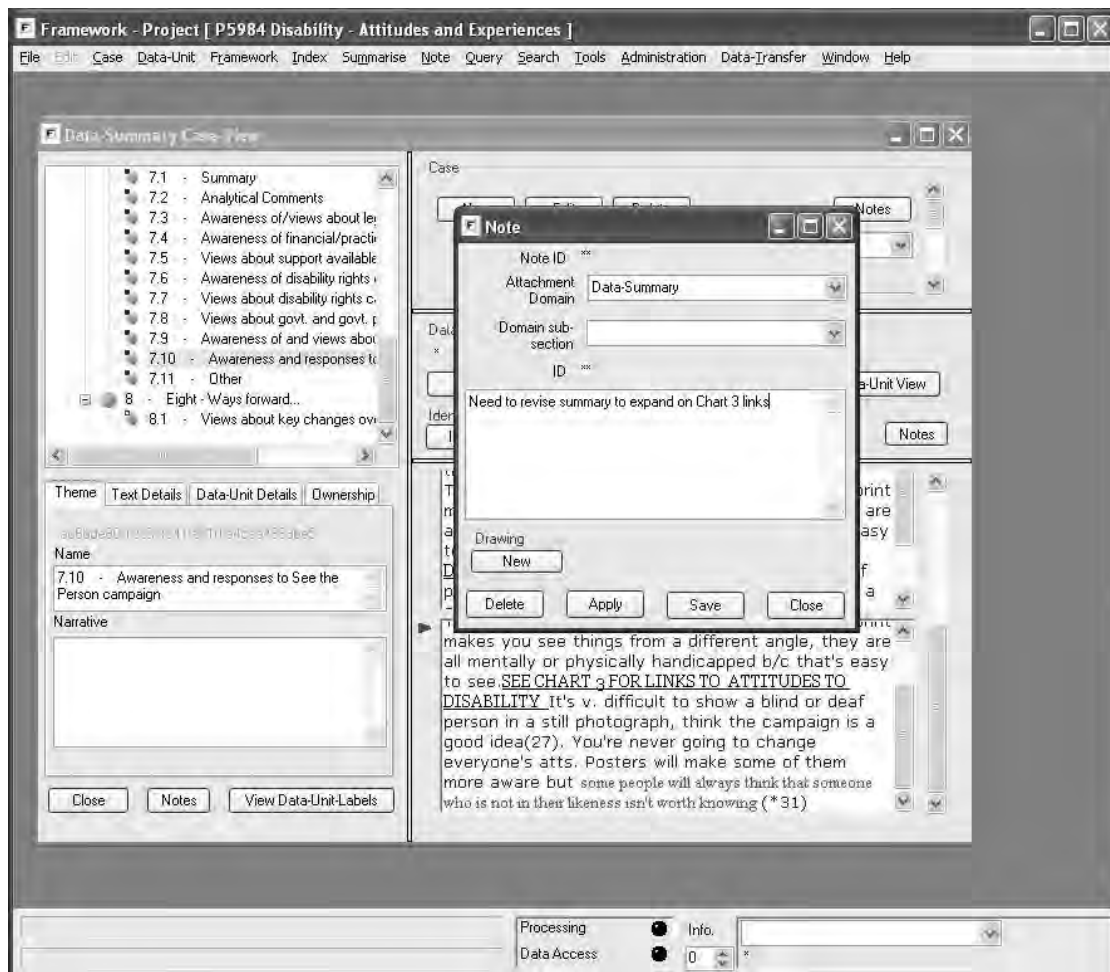
The programme has several core features allowing researchers to:

- work with all forms of textual data (from transcripts of interviews to field notes)
- develop a series of conceptual matrices for the organisation of data which can easily be modified by a 'drag and drop' facility transferring data to new categories as required
- 'index' or code their data by conceptual categories allowing for the swift and systematic retrieval of extracts of verbatim data



- summarise their data into conceptual matrices allowing for data reduction and organisation
- perform both simple and complex text searches and identify positional relationships between themes, index labels and cases

- create a range of outputs (indexed extracts of verbatim data and matrix displays of summarised data)
- develop a range of analytical queries which can produce matrix based or verbatim outputs relating to one or more conceptual themes
- apply socio-demographic or analytical labels to data units (for example typological classifications) at any point in the process to provide more precision and creativity in the interpretative process
- monitor a dynamic audit trail of revisions and changes to the dataset
- record analytical thoughts and, in the networked version, work concurrently with other team members across a network sharing analytical hunches and developments



### Core features supporting quality and transparency

'Framework' can enhance the quality of the analytical journey in a number of key ways. Firstly, the software has inbuilt system checks which enforce consistent, systematic treatment of the data. For example, it is impossible to create a summary without ensuring that a direct link to the original case, this allows the analyst to easily switch between their summarisation and that data in context. Secondly, the display functions mean that the original transcript or text is always on display whether during the coding or summarisation process and the researcher can always retrace their steps to the relevant section of the original document ensuring the context of the data is retained at all times.

'Framework' also has a dynamic audit trail facility which means that the software automatically records a visible account of changes to the analytical process, including who has been responsible for changes. For example, as the analyst develops their conceptual framework the programme will track the development of new conceptual categories and their application to the data units. This type of function is invaluable for collaborative research but equally useful for sole researchers in keeping track of the steps they have taken along the analytical route.

By incorporating the key functionality of other software (such as code and retrieve functions) 'Framework' helps to support analytical creativity allowing the researcher to customise which elements of the programme are suited to their data and analytical approach. Output formats can also be customised allowing for greater creativity and ensuring that the software is driven by the researcher's requirements rather than enforcing a rigid analytical route through the data.

As with other packages 'Framework' provides for the documentation of analytical developments in its notebook function and in the networked package users can share these thoughts in 'real time'. The networked version of the package will allow research teams, even in different locations, to share their dataset avoiding the often fraught process of merging datasets different researchers have created and provide greater capacity for shared analytical hunches, hypothesis and questions.

## 5. Conclusions

The use of new technologies in qualitative research has come a long way since the mid-1980's, no longer limited to simple word processing or text retrieval the new CAQDAS packages offer an opportunity to bolster and sustain quality and creativity in the analytical process. 'Framework' is just one such package and continuing innovation in digital technology and software will provide ongoing challenges and opportunities for qualitative researchers to engage with technological developments.

There remains, though a need for vigilance amongst users of software packages and amongst those training researchers new to the craft of analysis. It remains critical that software users learn both the key principles of qualitative data analysis and how to effectively use their package of choice. We need to retain transparency about the approaches we have used to interpret our data, if we say we used 'Framework' to analyse the data we need to ensure we have explained what that means in practice. Most of all we need to continue to provide sufficient detail of the context, depth and detail in our datasets to allow our audience to make informed judgements about the credibility of our findings and the approach we have taken. In our efforts to uphold analytical adequacy continued technological developments, such as sophisticated audit trails, may help us to demonstrate the credibility of our findings but it is unlikely that software innovation will ever be able to replace the craft and creativity that the researcher brings to the process.

A web-based demonstration of 'Framework' can be found at [www.natcen.ac.uk/Framework](http://www.natcen.ac.uk/Framework); for further information about 'Framework' please email [Framework@natcen.ac.uk](mailto:Framework@natcen.ac.uk).

## References

- Coffey, A. Holbrook, B. & Atkinson, P. (1996) 'Qualitative Data Analysis: Technologies and Representations', Sociological Research Online, vol.1, no. 1, <http://www.socresonline.org.uk/socresonline/1/1/4.html/>
- Fielding, N. & Lee, R. (2000) 'Patterns and potentials in the adoption of qualitative software: the implications of user experiences and software training', <http://qualitative-research.net/fqs.texts/3-00/3.00fielding-e.htm>
- Fielding, N. & Lee, R. (2002) 'New Patterns in the Adoption and Use of Qualitative Software', *Field Methods*; 14; 197.
- MacMillan, K. & Koenig, T. (2004) 'The Wow Factor: Preconceptions and expectations for Data Analysis Software in Qualitative Research', *Social Science Computer Review*; 22; 179
- Spencer, L. et al. (2004) 'Quality in Qualitative Evaluation: A 'Framework' for assessing research evidence', Cabinet Office: London

## About the Authors

**Kandy Woodfield** and **William O'Connor** are co-Deputy Directors of the Qualitative Research Unit at the National Centre *for* Social Research (NatCen). NatCen is an independent social research institute conducting a range of sponsored and grant funded research across the broad spectrum of applied social policy. Both authors have extensive experience of designing and conducting qualitative research studies and have jointly led the development of the 'Framework' package. In addition to conducting research and the 'Framework' development both are experienced trainers in research methods and regularly lead training for a range of different organisations. They can be contacted at: the Qualitative Research Unit, 35 Northampton Square, London EC1V OAX; telephone 020 7250 1866.



# XSight and the shaping of Marketing Analytics

David F. Birks

## Abstract

The marketing research industry in the UK is facing unprecedented competitive pressures. Many commentators argue that these pressures are forcing a polarisation of the industry with a sector that strictly focus upon competences in gathering and analysing data, and one that focuses upon adding value to data by creating strategic marketing intelligence. In response to these developments, the University of Southampton has developed an MSc in Marketing Analytics (which received 152 applicants in its first year and selected 21 students for its first cohort). This paper will set out the University of Southampton's perspective of Marketing Analytics and how it aims to integrate research skills, software training and management expertise. It will argue that Marketing Analytics goes beyond pure quantitative measures of marketing performance to incorporate qualitative marketing research and competitor intelligence. In the teaching of Qualitative Marketing Research and Data Analysis, XSight, the qualitative data analysis software from QSR will be used for the first time (at Southampton or indeed any University in Europe). The paper will reflect upon the experience of using this software with a very diverse set of international students that are predominantly numerate. Will the outcome mean 'back to the drawing board' in defining the nature and scope of Marketing Analytics?

## Keywords

Marketing Analytics; Qualitative research; Computer-assisted data analysis; XSight; Marketing education.

## 1. Introduction

Marketing Analytics is a relatively new and emerging discipline in the study and practice of marketing. The nature and scope of Marketing Analytics may be perceived in many different manners ranging from a very numerical perspective of performance and control through to a broader based means to support marketing decision-making. This paper sets out a perspective of Marketing Analytics based upon the intellectual and educational values set out at the University of Southampton. It will show how qualitative marketing research and the analysis of qualitative data fits into that perspective. In delivering the MSc to their first cohort of students, the qualitative data analysis package XSight was used. The rationale for using qualitative data analysis software will be presented with a reflection of the impact of using XSight with a very diverse set of international students who are predominantly numerate. The paper concludes by reflecting upon the efficacy of XSight and qualitative marketing research in the development of the discipline of Marketing Analytics.



## 2. What is Marketing Analytics? The Southampton perspective

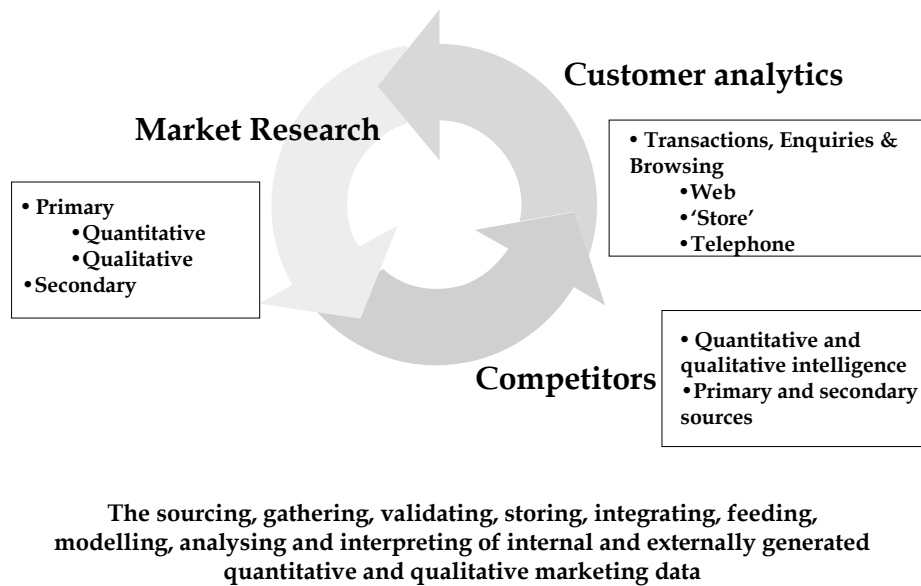
Over the last decade, there have been massive developments in research approaches, information systems, and software to help understand how existing and potential consumers behave. Evidence of such developments can be seen in examples such as: the use of loyalty cards to electronically observe and conduct experiments with consumers, the use of geodemographic information systems to map out consumer attitudes and behaviour, customer relationship management systems and the operational data of e-commerce and web analytics that describe and model consumer behaviour, the use of data-mining and modelling techniques to uncover new meaning in the relationship between marketers and their consumers, and the growth of ethnographic approaches to add richness to the huge amounts of quantitative data now available. Such developments have enabled far more ‘sophisticated’ approaches to marketing analysis and the information systems that support it. In this context, such developments have also meant a great demand for professionals that can cope with the challenges of creating meaning from this massive explosion of data.

In response to these developments, the School of Management at the University of Southampton has created a new MSc in Marketing Analytics. The degree was designed to provide an understanding of the interplay between research methods, statistical techniques and information systems in marketing decision-making. It is driven from a perspective that values how marketing data is collected, stored, analysed, disseminated, and interpreted in order to make sound marketing decisions. Figure 1 below shows how marketing data is viewed from both a quantitative and qualitative perspective. The sources and distinct disciplines of dealing with data from these sources helps to set out the challenges of Marketing Analytics.

Marketing Analytics is seen at Southampton as:

*“The challenges involved in sourcing, gathering, validating, storing, integrating, feeding, modelling, analysing and interpreting of internal and externally generated quantitative and qualitative marketing data”*

There is also a broader context in which these challenges are set which further develops a notion of Marketing Analytics. First there is an understanding of marketing management and strategic marketing, i.e. an awareness of the decisions and decision makers who will use the outputs of marketing analytics. Second, there is an understanding of information systems strategy, i.e. an awareness of the structural means to create knowledge in an organisation.

**Fig. 1 The challenges of Marketing Analytics**

To try to provide the intellectual and skill development of such a broad range of disciplines could be seen as a huge undertaking, especially in a one year course. This undertaking is greatly helped by the students bringing a strong set of analytic disciplines from first degrees in areas such as economics, computer science, engineering and statistics. The intellectual abilities and skills grounded in their first degrees help them to quickly grasp certain parts of the panorama of challenges set out in the Marketing Analytics definition. However, with students coming to the degree with primarily numeric backgrounds, a major challenge lies in developing the rationale and primarily linguistic and artistic skills demanded in qualitative marketing research. The next step therefore is to present the case as to why qualitative marketing research should be a component of marketing analytics as a discipline and as a way of thinking.

### 3. The need for an appreciation of qualitative research within Marketing Analytics

Marketers and researchers often take dogmatic positions in favour of either qualitative or quantitative marketing research. Their positions are founded upon which approach is perceived to give the most accurate understanding of consumer behaviour. Many quantitative researchers are apt to dismiss qualitative studies completely as giving no valid findings, indeed as being little better than journalistic accounts. They assert that qualitative researchers are totally subjective and ignore basic premises of sampling, with their findings based on a single case or only a few cases. Equally adamant are some qualitative researchers who firmly reject statistical and other quantitative methods as producing a blinkered view that fails to appreciate the contextual nature of consumer behaviour, one that can force a language and logic that is meaningless to respondents. They believe that to understand cultural values and consumer behaviour requires interviewing or intensive field observation. Qualitative techniques they see as being the only methods of data collection sensitive enough to capture the nuances of consumer attitudes, motives and behaviour. (Strauss and Corbin, 1998)

The arguments between qualitative and quantitative marketing researchers about their relative strengths and weaknesses are of real practical value. The arguments reveal the ‘gaps’ that can emerge when trying to create an understanding of consumers. Qualitative research can clearly help to develop an understanding of the language and logic of target respondents. It can clearly help to understand the reasons why there may be statistically significant differences between respondent types. In the validation of exploratory multivariate statistical techniques such as cluster analysis and principal components analysis, qualitative research can help to nurture and validate solutions. In short, well crafted qualitative research offers much to help shape the nature and focus of quantitative research, and vice versa. Seeking a singular and uniform approach to supporting decision-makers by focusing on one approach is futile. Defending qualitative approaches for a particular marketing research problem through the positive benefits it bestows and explaining the negative alternatives of a quantitative approach is healthy – and vice-versa.

Business and marketing decision-makers use both approaches individually and together and will continue to need both. (Cooper, 1999) As such, marketing analytics students need to appreciate the skills and applications of qualitative marketing research, especially when set in the broader context of connecting marketing research to customer analytics and competitor intelligence.

#### 4. The challenges of analysing qualitative data

Conducting qualitative research (or indeed any form of research) is not a matter of collecting data and then deciding what to do to with a pile of transcripts, wondering how to make sense of that data.

*Analysis is a pervasive activity throughout the life of a research project. Analysis is not simply one of the later stages of research, to be followed by an equally separate phase of ‘writing up results’.* (Coffey and Atkinson, 1996)

When the qualitative research decides to probe a particularly interesting issue, they are making analysis decisions. When they decide to focus upon interviewing individuals with particular experiences or a novel means to express themselves, they are making analysis decisions. For the qualitative researcher, analysis continuously takes place as they gather data. Validating the sense they see as their data builds up, also means that analysis takes place as they gather data. In this process it is worth considering what qualitative data can mean:

1. Audiotape recordings and transcripts of those recordings.
2. Videotape recordings.
3. Notes taken during or after interviewing or observations.
4. Reflections of researchers, moderators or observers involved in the data collection process.
5. Documents produced by or sourced from respondents.
6. Photographs, drawings, diagrams, i.e. still visual images.
7. Records made by respondents such as mood boards or collages.

Coping with the analysis of this array of data, qualitative marketing researchers need a framework to help them to create a ‘story’ or meaning in the narrative as they see it. Such a

framework is shaped by two factors. The first factor is the *theoretical understanding* of the researcher as they collect and analyse the data. This theoretical understanding can be viewed from two perspectives. The first is the use of theory published in secondary data, intelligence and literature. The use of theory from these sources may help the researcher to understand what they should focus their attention upon, in their questioning, probing, observations and interpretations. The second is the use of theory from a grounded theory perspective. The researcher may see limitations in existing theory that do not match the observations they are making. These limitations help the researcher to form the focus of their questioning, probing, observations and interpretations.

The second factor that shapes the direction that the researcher takes is a *marketing understanding*. The researcher needs to understand what marketing decision-makers are going to do with the 'story' they create. The researcher needs to appreciate the decisions faced in creating an advertisement, building a communications campaign or perhaps changing features of the product. Reference to theoretical and marketing understanding in the researcher helps them to present the most valid interpretation of their story to decision-makers

## 5. Using software to analyse qualitative data

Today, it is difficult to imagine a quantitative marketing researcher coping with the analysis of a survey without the use of analysis software. The idea of wading through piles of questionnaires to perform analyses seems archaic. With our present knowledge of the power of survey design and analysis software, it is clear to see that much insight and potential within the data could easily be lost by performing a 'manual' analysis. This scenario does not necessarily hold for the qualitative marketing researcher at either a practitioner or academic level. As in quantitative research, qualitative analysis software packages cannot interpret and find meaning in data. The software facilitates, and in some cases automates, the identification and coding of text. But there is sometimes a false assumption that identification and coding are simple and unproblematic, and critical evaluation and scrutiny of coded segments and code counts are not needed. By facilitating quick analyses, which focus on quantitative category relationships, the software may discourage more time-consuming, in-depth interpretations. Whilst the software is intended as a means of allowing the researcher to stay close to the data, their misuse can have the unintended result of distancing the researcher from the data. The following arguments founded over 10 years ago set out the challenges of using software to analyse qualitative data, then balanced with what should be the clear advantages of using analysis software.

**Disadvantages of qualitative data analysis software** (Wolfe et al, 1993) (Dembrowski and Hanmer-Lloyd, 1995)

1. *Mechanistic data analysis*. The software cannot replace the creative process expected of the qualitative researcher. The researcher can evaluate the interrelated play on particular words, the tone of voice or the gestures of a particular respondent. The sensitivity towards these relationships and connections can be lost in a mechanistic search for statements.
2. *Loss of the overview*. The researcher may be seduced into concentrating on the detail of individual chunks of data and assigning codes to the data. This focus may detract from the overall context that is so vital to identify and name chunks of data. Making sense of codes can be greatly facilitated by an ability to visualise the data in its entirety.

3. *Obsession with volume* Given the ability to manipulate large amounts of data, there may be a push to increase the number of interviews. This may be counter-productive in that the emphasis should be on the interrelated *qualities* of:

individual respondents

the interview process.

4. *Exclusion of non-text data.* As noted earlier, qualitative data can include notes, observations, pictures and music that make up the total 'picture' or representation of individuals. Many programs cannot cope with the integration of different types of qualitative data, their focus being upon transcript narrative.

#### **Advantages of computer-assisted qualitative data analysis (Seale, 2000)**

1. *Speed.* The speed at which programs carry out sorting procedures on large volumes of data continues to get faster. It gives the qualitative researcher more time to reflect upon the meaning in their data. It can also enable the rapid feedback of the results of particular analytic ideas so that new ones can be formulated. Analysis becomes more devoted to creative and intellectual tasks, less immersed in the routine of shifting pieces of paper.

2. *Rigour.* In this context 'Rigour' can mean the locating and counting the number of times words, phrases and contextual relationships have occurred. From this awareness, a demonstration of negative incidences that have been located can be searched for and shown. This contrasts to a more 'relaxed' approach of selecting anecdotes that support a particular interpretation.

3. *Team.* In collaborative research projects where researchers need to agree on the meaning of codes, a check can easily be made of whether team members are interpreting data in the same way. This is particularly useful as coding moves from the more descriptive and mundane codes to ones that reflect broader theoretical concerns. Researchers can more readily pass coded interviews between them, and compare the results.

4. *Sampling.* It is easy to keep track of who *has* been interviewed, compared with the intentions of who *should* be interviewed. Beyond the sampling of individuals is the concept of theoretical sampling, i.e. the inclusion of events that corroborate or contradict developing theory. As the researcher has more time to spend on creative and intellectual tasks, they can develop stronger descriptions and theories and strengthen the validity of their views by ensuring they have sampled sufficient incidences.

## **6. Using XSight in the teaching of Marketing Analytics**

It was contended earlier that in order to fully appreciate the breadth of data support demanded by marketing decision-makers, marketing analytics students need to appreciate the skills and applications of qualitative marketing research. A specific unit, delivered over three weeks, was designed to develop an appreciation of these skills and applications. Entitled ***Qualitative Marketing Research and Data Analysis***, its key learning outcomes were to understand:

The need to balance the needs and motivations of respondents and decision makers in qualitative marketing research design and implementation

How qualitative research fits in and complements the marketing research process.

How technology and software can enhance the efficiency and quality of the marketing research process

A key part of the development of these outcomes was an individual project that contributed 100% of the marks for the unit. Over a period of six weeks, they had to choose a contemporary marketing problem that could benefit from qualitative marketing research support. From this focus they were expected to diagnose the marketing and research problem and show how qualitative research could provide a solution. The students were expected to conduct two qualitative interviews, either in-depth interviews and/or focus groups. They were expected to record the data from these interviews, produce transcripts and analyse their qualitative data using the XSight software package. They were then expected to present conclusions and marketing recommendations. With such 'little' data to work with, and recognising that their first interview could be an experimental pilot interview, questions may be raised about why use analysis software and why XSight?

English was not the mother tongue of 20 of the 21 students on the degree. Of the 20 overseas students, a number conducted their interviews in their mother tongue, helping to gain access to respondents in their home country and tackling a marketing problem of relevance to that environment. Given the translation, access and timing constraints, the emphasis was not upon gathering masses of data, but appreciating the challenges of generating well focused, quality data. Having generated qualitative data, the educational challenge was to evaluate whether qualitative analysis software would support these primarily numerate and software savvy students. Support in this sense meant finding the meaning in the data to produce sound marketing recommendations.

#### Why QSR XSight?

There is a great choice of qualitative data analysis software that could be used for this exercise. The University of Southampton has a site licence for the QSR NVivo package and most academic qualitative researchers are familiar with the program. It would have been a relatively simple and cheap option to demand the students use this program. However, in 2004 QSR launched XSight with the claims:

*"XSight is a user-friendly program for qualitative data analysis software, customised for marketing researchers. Conventional qualitative data analysis software was designed for and used mainly by academic researchers. XSight was designed for marketing researchers by marketing researchers who understood the particular problems faced by their profession".*

In developing employable skills Marketing Analytics students, could this be the software that would be *de-rigueur* for their CVs? QSR International have an office in London and count major UK market research firms amongst their customers. However, no Universities in Europe use XSight in the teaching of marketing research, the global pioneers in this field being Australian Universities. QSR's CEO John Owen said.

*"Like the release of any new technology, there were some initial doomsayers – those who thought it would rob them of their creativity. XSight has never been about stifling the creative process. It's designed to do the opposite - to take the pain out of the collation and organisation phase so researchers have more time to devote to analysis.*

The words of Gill Ereaut, Director of UK company Linguistic Landscapes and respected Marketing Research author gave comfort that our experiment with the students was worthwhile:

*“Qualitative analysis software will radically change the way people learn how to do analysis in commercial research. With XSight, the process of analysis becomes not just more flexible but more visible and transparent. Newcomers, students and trainees can be shown clearly how experienced researchers search, interrogate and interpret data to reach complex insights and conclusions for clients”.*

The concluding statement of Tim Macer’s review of XSight underlined the challenge of changing the behaviour of dyed-in-the wool qualitative researchers. Changing their attitudes and behaviour in a positive manner towards analysis software could be a slow and laborious process.

*“Perhaps it is over-optimistic to think that the more creative, free-spirited and self-confessed technophobes among those practicing qualitative research will find much excitement in XSight. But for anyone interested in reclaiming their office carpet for walking on, as they analyze 40 transcripts, this software can not only de-clutter the office, but potentially the mind too, and claim back some time for the real thinking”.*(Macer, 2005)

In de-cluttering the office and mind for the emerging generation of researchers, the experts were pointing the Marketing Analytics students towards an XSight experience.

## 7. How the students coped with XSight

In lectures and in their reading of the core text for the course (Malhotra and Birks, 2007), the students were exposed to the principles of analysing qualitative data and the pros and cons of using software to complete this process. They were introduced to XSight in broad terms of what its capabilities were and how to get started. Bundled with their core text was a full working version of the software, timed to run for three months from registration. There were deliberately no XSight classes or workshops, a different approach to all the other software training on the degree. Students were expected to work through the set XSight tutorial and then apply their learning to the data from their two qualitative interviews. In the time that this was supposed to be occurring, the students posed no enquiries or problems with any aspect of what they were expected to do with XSight. This aroused suspicions that they were not doing anything (except collectively creating a complaint to the Vice Chancellor). In one instance close to the deadline for submission, one student asked when they were going to have their XSight lecture. This clearly revealed that in this instance, the student had not read any of the material given to them – and when it was pointed out, they got on with the job.

It was deduced how well they had coped with the assignment and using XSight in this process when their final reports and XSight files were received. The most striking observation was how well the XSight tutorial had supported their learning of the software. Working at their own pace, at times that suited their habits and work patterns, and sometimes in their own space, they progressed much better than in a collective workshop or class. What they produced from the exercise and by using XSight can best be represented by relating to the relevant listed advantages and disadvantages of qualitative data analysis software.

1. **Mechanistic.** XSight allowed their transcripts to be viewed. From this the process of their interview could be visualised. It was clear to see what questions they posed, how they warmed up and opened up respondents, how they probed on particular issues. Examining the transcripts, it was clear that the nature of questions and probes were not designed to allow respondents to reflect, open up and express themselves. In qualitative terms, the questioning process was superficial and did not take respondents to any depth. This generally meant short questions and answers and a rather mechanistic approach to gathering their data. This was reflected in their analysis which tended to be a simple representation of these questions and answers. Taking this representation further, XSight allowed the analytical frameworks used by the students to be viewed. By examining the analytical framework, it was possible to see the extent to which they explored for new meanings and coded data in different ways. It was easy to see that in most instances, there was little difference between the topic guide and the analytical framework. This reinforced a view that the analysis was mechanistic
2. **Loss of overview.** XSight allowed a graphic representation or map of the structure and meaning they saw in the data to be visualised. In very few examples did these representations portray all or a majority of the data. They were most adept at using the graphics to represent the structure of issues and their interconnections in discrete areas. Many of these insights into discrete areas were illuminating but pulling these areas together to create an overview was not usually there. XSight enabled them to build structure into the issues they found important. It was clear to see how they developed meaning and insight in their data. The extent to which they could convey or communicate that insight could be clearly evaluated.
3. **Exclusion of non-text data.** XSight simply portrayed the extent to which they used additional materials, allowing a view of how well they had understood the nature of qualitative data. If additional data was used, XSight made it clear to see how that data was used. At an individual interview level or in an overview it was clear to see where links were made in a verbatim, paraphrased or interpretive manner. It was very rare to see this happen; the much easier option of just working with the data in the transcripts was the norm.
4. **Speed.** One of the biggest lessons for the students was how long it took to create good quality transcripts. A key part of the exercise was for them to appreciate the problems of getting access to good 'quality' respondents. Once they had physical access, they had to appreciate the problems of a 'mental' access, i.e., getting respondents to connect with an interviewer and the topics, and ultimately opening up. Appreciating what is involved in planning this correctly and the timing impacts of this was paramount. Once they had created their transcripts, the speed with which they learnt about how to use the software and then cope with the analysis was admirable. This has to be balanced with points 1 to 3 above, with criticisms of a mechanistic, fragmented and transcript focused analysis. Such an approach may point to analysing the data far too quickly with insufficient reflection.
5. **Rigour.** XSight allowed the extent to which they had created Analytical Frameworks and Queries to be visualised. With their Analytical Frameworks, the different ways that they were coding data could be observed; the insight they were generating was plain to see. With their Queries, the ways that they were looking for differences and



similarities between respondent types could be observed. The extent to which they were validating their interpretations by searching for alternative explanations could be observed. Given the relatively small amount of data to work with and the above critique of a superficial approach to analysis, there was no evidence of extensive iterations of frameworks, queries and validation. With manual analysis of qualitative data, there are limitations in displaying such rigour. It was felt that in even this limited exercise; XSight imposed a greater rigour upon the process. By having to work through the creation of Sample Characteristics and Analytic Framework, issues of Rigour had to be addressed.

## 8. Reflections upon the use of XSight and its contribution to the development of Marketing Analytics

Having taught Qualitative Marketing Research for many years without embedding the compulsory use of a particular software package into assessed work (100% in this case); there had to be confidence that XSight would deliver on its promises. Having previously used the QSR NVivo package (with its widespread acceptance as an excellent tool for academic qualitative researchers) before personally using XSight, the benefits were clear. XSight supports the way commercial marketing researchers approach data analysis (or could do), which was also clear from expert reviews and the growth of its use with market research organisations. XSight or any equivalent packages are hardly used in University teaching of Marketing Research. This may be because of the space in timetables allowed for qualitative marketing research in University courses. Generally there is a bias towards quantitative data collection and analysis, a bias which is reflected in the majority of standard marketing research textbooks. If universities were to create more space for qualitative research and redress any imbalance of skill, would XSight help to support the thinking and skills of qualitative marketing researchers? The view in short was a definite 'yes'.

The students selected for the Marketing Analytics degree came primarily from numeric backgrounds and were familiar with survey design and statistical analysis software. Much of this thinking and training was reflected in their approach to qualitative research. There was little appreciation of what 'qualitative' meant in terms of the qualities held in target respondents, the quality of interchange, the quality of elicitation to reveal insights from these respondents, the quality of data needed to validate a particular insight, the quality of reflection needed in coding data and searching for structure in that data. XSight helped to magnify this lack of appreciation as it presented evidence of their thought process and outcomes. In subsequent classes, further emphases can be put on these qualities. Students should think of qualitative data analysis as something that pervades the whole process of targeting respondents, developing issues to question and observe and in gathering and validating data, not just as an endpoint once a mass of data has been collected. To this end, a simple exercise at the start of the course, that rewards students for working through the XSight tutorial could aid their thinking about the differences in quantitative and qualitative data analysis.

Finally the impact of XSight on the development of the subject of Marketing Analytics should be addressed. Marketing Analytics has been presented as: 'the challenges involved in sourcing, gathering, validating, storing, integrating, feeding, modelling, analysing and interpreting of internal and externally generated quantitative and qualitative marketing data' These challenges are set in the context of an understanding of marketing management and strategic marketing, i.e. an awareness of the decisions and decision makers who will use the

outputs of marketing analytics. It is because these students will have to build trust, engage with and communicate convincingly with marketing decision makers that we have sought to balance qualitative and quantitative thinking, skills and software use.

Qualitative research can clearly help to develop an understanding of the language and logic of target respondents as they think and behave about products and services. It can clearly help to understand the reasons why there may be differences between types of consumers. In the validation of exploratory multivariate statistical techniques, qualitative research can help to nurture and validate solutions. To be effective in Marketing Analytics, well crafted qualitative research offers much to help shape the nature and focus of quantitative research, and vice versa. Even though the students treated qualitative research in quite a superficial manner, the use of XSight in their training helped them to understand the links and integration between different data types. The 'comfort' of well structured and clear software in XSight, eased the journey towards an approach associated with linguistic and artistic skills. As well as seeing the nature and purpose of integrating qualitative and quantitative data, XSight gave students the confidence to progress along that difficult path of integration.

## References

- Coffey, A. and Atkinson, P., (1996) *Making Sense of Qualitative Data Complementary Research Strategies*, Thousand Oaks, CA: Sage, pp. 10–11.
- Cooper, P., (1999) 'Consumer understanding, change and qualitative research', *Journal of the Market Research Society* 41, 1, January 1999, p. 3.
- Dembrowski, S. and Hanmer-Lloyd, S., (1995) 'Computer applications – a new road to qualitative data analysis?' *European Journal of Marketing* 29, 11, November, p. 50.
- Macer, T., (2005) XSight 1.2, *Quirk's Marketing Research Review* - December
- Malhotra, N.J. & Birks, D.F., (2007) *Marketing Research, An Applied Approach*, 3rd European Edition, Prentice Hall International.
- Seale, C., (2000) in Silverman, D., *Doing Qualitative Research: A Practical Handbook* London: Sage, p. 161.
- Strauss, A. and Corbin, J.M., (1998) *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, 2nd ed. Thousand Oaks, CA: Sage, pp. 27–34.
- Wolfe, R.A., Gephart, R.P. and Johnson, T.E., (1993) 'Computer-facilitated qualitative data analysis: Potential contributions to management research', *Journal of Management* 19, 3, Fall, p. 637

## About the Author

**David Birks** is a Senior Lecturer in Marketing Research at Winchester School of Art at the University of Southampton. He was previously at the School of Management at Southampton and Programme Director for their MSc in Marketing Analytics. Prior to Southampton, David lectured at the Universities of Bath and Strathclyde. At Bath he managed the dissertation programmes of their B.Sc., M.Sc., and M.B.A degrees. He also undertook the role of Director of Studies for their PhD programme and their Executive MBAs in Malaysia and China. David

is the co-author of Europe's leading Marketing Research textbook: Malhotra, N.K. & Birks, D.F., *Marketing Research: An Applied Approach*, 3<sup>rd</sup> European Ed. Financial Times/Prentice Hall (2007). Before university lecturing he worked in purchasing, planning, marketing and research. His publications cover the fields of Housing, Statistics, Marketing and Information Systems. He can be contacted at Winchester School of Art, The University of Southampton, Park Avenue, Winchester, SO23 8DL. Tel. 02380 59 6900; e-mail [d.f.birks@soton.ac.uk](mailto:d.f.birks@soton.ac.uk).

# Is Quanvert Here To Stay?

Nick Read

## Abstract

Quanvert has continued its dominance in the survey reporting space despite the release of other products by competing providers. But is its time coming to an end? Can it continue to deliver to the needs of users, or is it about to be superseded? The paper begins with an overview of the Quanvert application, and moves on to discuss a number of significant developments, which have changed the technological landscape and created new opportunities for taking a fresh look at how to address the needs of the market. It is *more* than just about replacing Quanvert functionality; it is about understanding the business processes around it. The paper will describe what SPSS has learnt – both from customers and competitors – and how it has responded in order to bring a real alternative to Quanvert to the market.

## Keywords

Quanvert; SPSS; Dimensions; analysis; reporting; online; interfaces; integration; Enterprise Feedback Management

## 1. Introduction

### Looking Back...

From its first introduction on to the scene in the mid-1980s, the capabilities of Quanvert and its underlying fast portable inverted<sup>1</sup> data format meant that companies rapidly adopted it and distributed it to their customers for their own use. Consequently, it has become a de-facto standard for the analysis of market research data and its functionality sets the benchmark by which other applications are measured.

At the time of Quanvert's introduction, any requirement for additional tabulations meant a request to a Data Processing professional, typically a day's wait, and often an incremental charge. Quanvert enabled instantaneous turnaround and eliminated the incremental charges. This allowed for a more iterative, thorough analytic process and a faster project cycle – it was a real leap forward.

<sup>1</sup> The production of tables is very fast due to the format of the database, which is known as an *inverted database*, i.e. data is grouped by question name or variable name rather than by respondent records.

Quanvert databases are created using Quantum, a leading package for handling all of the requisite data processing tasks that a project may require, including editing, weighting, and tabulating survey data. After writing the data processing specification for the survey, the Quantum scriptwriter runs the program to create a Quanvert database, which is then ready to send to the analyst/client. The recipient then uses the Quanvert desktop application to interrogate the data, create their own tabular reports, etc.

Quanvert is specifically designed to enable the market researcher to focus on the analysis of survey results without having to be a data processing or computing expert, or a statistician. If a survey requires more thorough statistical analysis, users can export from Quanvert to SPSS or similar statistical packages.

## 2. Key Requirements for Analysis and Reporting: Mid-1980s

The success of Quanvert has largely been because of its ability to meet the needs of a particular market at a particular time. These needs are split into three broad sections below covering the Technical, Functional and Implementation requirements for analysis and reporting applications. The Functional section is further subdivided into the five key steps that an analyst would encounter when they are working through the reporting phase. Each of these sections will be reviewed, along with a brief commentary about how Quanvert fulfils the requirements.

### A. Technical Requirements

The minimum system environment to run an application is a result of dependencies on the operating system, hardware and software requirements. Quanvert was originally released in the mid-1980s for UNIX, and then as a 16-bit application running on Windows 3.1 or MS-DOS version 3.0 in the early 1990s. The underlying nature of the inverted data structure meant that there was less reliance on processing and CPU memory when aggregating the data into tabular format.

### B. Functional Requirements

Five key steps have been identified that an analyst would encounter when they are working through the reporting phase. These form the basis of the functional requirements:

1. Utilize a variety of different data formats
2. Control the way information is expressed in the report
3. Allow users of all levels to rapidly build tabulations
4. Produce reports that are readable and understandable
5. Publish reports in the right format, for the right people

What makes Quanvert successful since its launch in the mid-1980s is its ability to provide all the features to meet these functional requirements.

#### **1. Utilize a variety of different data formats**

Q: Which kinds of data formats can I work with in Quanvert?

Data can be collected from respondents using a variety of different modes or channels (e.g. telephone, online, paper). The task of processing the data is undertaken by the Quantum scriptwriter, who writes the data processing specification to define the variables based on the

requirements for the survey, and then runs the program to create the native Quanvert 'packed' database, or.pkd file, which is then sent to the analyst/client.

The Quanvert user does not need to be a data processing or computing expert, or a statistician; they can simply open the database in Quanvert and start working with the data. In addition, they also have the option of being able to merge different data sets together when working with continuous tracking data. Where further statistical analysis is required, users can export the data to either SPSS or SAS format.

## **2. Control the way information is expressed in the report**

### **Q: How can Quanvert help me deliver targeted results?**

Quanvert offers complete control over the way information is expressed in a report, and allows the user the full capability to edit variables, define a filter, apply weighting and run statistical tests to verify the results.

The ability to edit variables is a fundamental requirement for reporting because often the way in which a question is asked in a survey may be quite different to the way it is presented in a report. For example, perhaps the responses to a question need to be grouped together (e.g. as nets) in order to produce a report at more of a summary level. In order to deliver targeted insights, the user needs to have control over the way in which the variables are used. Quanvert allows users to 'band' categories together for the display of summary results. When dealing with numeric questions, it is easy to include standard summary statistics like mean scores.

When producing a report, it is sometimes necessary to focus on a particular subset of respondents in the survey data. By applying a filter, the user can restrict the number of respondents eligible for inclusion in the results. A filter is defined by selecting the variable or variables that define the filter requirement, and then choosing the categories that define the exact filter condition. For example, a simple filter might be something such as sex=female that restricts the table to including female respondents only. This is easily achieved in Quanvert by applying a filter to the table so that those respondents in the table (or base) are drawn from the relevant group.

During the collection phase of a survey, if it is not possible to interview everyone, then only a sample of the population will be interviewed. If this sample group does not accurately reflect the proportions of various groups in the total population, the survey results may be biased. Weighting is another term for sample balancing whereby the responses in the survey can be weighted to reflect the actual balance in the population. Where a survey contains weighting information, the Quanvert user can choose whether to create weighted or unweighted tables by selecting the relevant variable.

After producing a table, a Quanvert user can then run the appropriate statistical test and check to see whether differences in the distribution of counts in tables are significant, or whether they are merely due to chance. Typically, in the case of testing columns in a table, the results of the test use letters from each column to display whether there is a significant difference between the results. This task may be achieved by manually setting the required parameters in the application.

### **3. Allow users of all levels to rapidly build tabulations**

Q: How does Quanvert empower users with different skill levels and different reporting needs?

When it comes to application usability, the design of the user interface directly affects the ease of access to core functionality in the product. When producing a report, the typical process is to first tabulate the data and then decide on an appropriate way to visualize the results before publishing the report. However, sometimes it is useful to be able to preview the data before building a table. In other circumstances, it is more appropriate to focus on respondent-level data to produce profile reports (e.g. when conducting analysis for account-level business-to-business research) or summarise a number of different variables in a single table (e.g. a mean score summary table).

All of these features, along with the ability to apply mandatory weighting to the tables, are available from within Quanvert.

### **4. Produce reports that are readable and understandable**

Q: How do I ensure that my reports have a consistent look and feel?

A report may ultimately be distributed as a standalone file that exists outside the reporting application (e.g. Microsoft® Office PowerPoint®). The tabulated data that appears in the report may be the result of some rigorous data management. Therefore, the report must contain the appropriate level of supporting information so that any business decisions made based on the report are well considered; several industry-standard guidelines<sup>2</sup> outline this requirement.

In order to produce reports that are readable and understandable, various elements in addition to the data must be controlled within the application. When defining these elements for a table (e.g. number of decimal places for percentage figures, default statistical tests, annotations that include the supporting information), it is useful if these key settings can be defined once and then stored in a template so that they apply as default for the entire report. This approach eliminates the need to start from scratch each time and ensures that the same consistency can be applied to subsequent tables. Quanvert users can define these elements once and save them as part of the specification file.

### **5. Publish reports in the right format, for the right people**

Q: What are the options for Quanvert users in terms publishing the results?

The report must be tailored to fit with the needs of the target audience. If the specification of the report can be saved, then the user should be able to send/publish this to an associate so that they can view the report using the same application. However, those people who do not use the same application will expect to receive results in some easy to understand format (e.g. Microsoft® Office Excel®). The Excel Viewer utility allows Quanvert users to produce tabbed Excel® output directly from Quanvert.

## **C. Implementation Requirements**

Quanvert is installed as a standalone desktop application and, as such, the implementation requirements are relatively straightforward. The application does not utilize any client-server technology and is controlled via a simple point-and-click user interface. If a project requires

---

<sup>2</sup> For example, *ICC/ESOMAR International Code of Marketing and Social Research Practice*; *ISO 20252: Market, opinion and social research — Vocabulary and service requirements*.

tabulations to be scheduled via a batch program, it is possible to do this via setting up a *standing order*.

### 3. New Challenges

Since the launch of Quanvert on to the market in the mid-1980s, there have been a number of significant developments, which have changed the technological landscape and created new challenges in terms of being able to meet the needs of the market.

#### Moving away from the Desktop and into the Online Realm

The growth of the internet in the 1990s and the associated rise in online survey methods made it easier for data to be collected from a geographically dispersed group of people. This growth also generated a demand for distributed reporting applications where the analysis and reporting of results moved from the desktop and into the online realm without requiring the user to have any software installed on their machine.

From a management perspective, this paradigm allows businesses to harness the benefits of a centralised system and have much greater control over the distribution of results. Software vendors recognize the power of the web, and for the first time, it is possible to deploy distributed applications over the internet without having to worry about managing the client desktop environment.

From a systems administrator standpoint, deploying an application using this distributed approach enables them to sidestep the administrative hassle of trying to keep track of the different versions of the software, which is unavoidable in the desktop space. Furthermore, if the application employs a *zero footprint* where the user is only required to have an internet browser and a reasonable connection to the internet, this makes it easier to deploy for larger organizations subject to strict security lockdown policies.

From a data management position, the centralised storage of data enables *one version of the truth* across the enterprise, and avoids the situation where users may be working with outdated versions of the data. Furthermore, if the data contains any personal customer information, then the organization must ensure that it is taking all of the necessary steps to adhere to data privacy legislation. If data containing personal customer information is stored on a laptop, then the organization may be subject to hefty fines if the laptop is stolen. In February 2007, Nationwide Building Society was fined close to £1M over security breaches following the theft of a laptop.

#### Providing Enhanced Interfaces

In the last five years, the market for survey reporting applications has matured and there are a number of vendors offering their own solutions to meet the needs of a business. However, the survey reporting market does not exist in a silo, and it is becoming increasingly important for applications to provide better interfaces to support the upstream (data collection) and downstream (dissemination of results) efforts.

With typical survey research, data may be collected via multiple modes or channels. For example, these channels may include interviewing respondents over the telephone, surveying respondents over the web, and face-to-face personal interviewing. For logistical reasons, an organization may acknowledge that different vendors have strengths in providing solutions



for different channels, and they may decide to use technology from different vendors for collecting the data. In an ideal world, reporting applications would have the ability to provide better support for the upstream data collection across a variety of channels, and would not be tethered to a proprietary data format.

When it is time to disseminate the results, the reporting requirements for today's organizations have moved beyond the standard *deck* of static tables. Results need to be accessible in a way that allows users maximum flexibility to analyse the data, uncover significant findings and publish the results in a timely fashion. In a world where there is an ever increasing amount of data being collected, and where a delay with making an informed business decision can have financial repercussions, organizations need to be able to publish reports in the right format for the right people. Often, the ultimate decision maker will expect to receive the report in a concise and easy-to-digest format. In order to provide better interfaces to support the dissemination of results, applications must be able to run pre-defined report definitions and output the results to a variety of standard file formats, e.g. Microsoft® Office PowerPoint®, Excel®. In addition, if a particular result is noteworthy and requires further investigation, the application must also support the user with this inductive line of enquiry and offer the full flexibility to slice and dice the data, and perform ad hoc analysis.

### Integrating with Management Reporting Structures

The growth of the internet in the 1990s also presented an opportunity for organizations to exploit the same protocol for their own *internal* management reporting needs. The various users with different roles accessing the different applications prefer to have a single access point to all of them over the internet. Portals provide a way for enterprises to provide a consistent look and feel and manage access control and procedures for multiple applications, which otherwise would have been different entities altogether.

Along with the demand for a single integrated view of the enterprise was the need to have enhanced data visualization. Based on the visual metaphor of the instrument panel in a car, digital dashboards provide at-a-glance visualization to support the monitoring of key performance indicators. Visual devices like traffic lights, gauges, and graphics are used to represent the high-level view and the dashboard will provide some level of interactivity in the form of drill-down capabilities where an item of interest can be selected to reveal the underlying data.

Management decisions based on the results of survey research are often balanced alongside sales and marketing metrics. In order for the results to be actionable, there needs to be tighter integration with existing management reporting structures, and while a single vendor may be able to provide a suite of products, which integrate seamlessly together, the reality is that it may not be feasible for a business to rip and replace all of their current systems. There is no single boilerplate solution that will meet everyone's needs – businesses need a reporting solution that is open and flexible to the specific needs of their own customers and stakeholders in order to gain maximum return on investment.

### Question: Is Quanvert Ready for the Challenge?

As noted above, it is *more* than just about replacing Quanvert functionality; it is about understanding the business processes around it. Quanvert emerged out of the need to serve the reporting requirements of the market research industry. As a standalone desktop reporting and analysis tool, Quanvert presents a viable solution for those organisations looking to

address the analysis and reporting requirements of a niche user base working with survey data. However, changes in the technological landscape have catalysed the emergence of three core themes: (a) a move away from the desktop and into the online realm, (b) greater emphasis placed on providing enhanced interfaces to support upstream and downstream efforts, and (c) transparent integration with management reporting structures. In its current incarnation, Quanvert is *not* ready for the challenge.

#### 4. Enter Dimensions

The Dimensions survey research platform was released in 2000 and is built on the Dimensions Data Model, a data access architecture that separates data storage and handling from the rest of the survey research process, providing easy access to the data regardless of how the information is stored. mrStudio, which is part of the Dimensions suite of products, is an integrated development environment (IDE) for developing scripts for data management and the creation of utilities. The mrStudio scriptwriter can write data management scripts and then use the add-on Tables Option to create tables and charts.

Data can also be made available to the analyst/client so that they can create their own tables and charts through a drag-and-drop interface using mrTables (online) or Desktop Reporter (desktop). Further work is progressing towards enabling Dimensions technology to feed management reporting technologies.

SPSS has listened to its customers and, in response to those challenges mentioned in the preceding section, the company has realigned how it positions the analysis and reporting tools in Dimensions to meet the needs of the market.

At the beginning of the paper, the five key requirements for analysis and reporting were outlined. These must now be revisited and the definition of each requirement broadened to incorporate the new challenges above.

#### 5. Key Requirements for Analysis and Reporting: Looking Forward

##### A. Technical Requirements

The minimum system environment to run an application is a result of dependencies on the operating system, hardware and software requirements. The emergence of distributed applications in the 1990s has shifted the onus of aggregating the data and processing the results onto the server. A clear advantage with this approach is that running the application is not wholly reliant on the processing power of the desktop machine. However, on occasion, there will be times when a user needs to be able to perform the same tasks without being connected to the internet.

Reporter is currently provided as a Desktop application, and the upcoming version 5.0 release will see the launch of the web based version. Both desktop and online versions of the application take advantage of the latest Microsoft® WinForms technology.

##### B. Functional Requirements

At the beginning of the paper, the following five key steps that an analyst would encounter when they are working through the reporting phases were provided:

1. Utilize a variety of different data formats
2. Control the way information is expressed in the report
3. Allow users of all levels to rapidly build tabulations
4. Produce reports that are readable and understandable
5. Publish reports in the right format, for the right people

These steps are still valid, but the definition of each step needs to be broadened to take account of the new challenges.

### **1. Utilize a variety of different data formats**

#### **Q: Which kinds of data formats can I work with in Dimensions?**

With Quanvert, the task of processing the data is fulfilled by the Quantum scriptwriter, who writes the data processing specification based on the requirements for the survey. This created a linear dependency where data first had to be processed by Quantum before it could be read in to the Quanvert application. SPSS acknowledges that customers may already be using different data collection technologies, and that there will be times where a snapshot view of the project will be useful to track progress with data collection, rather than having to wait until the end of the fieldwork.

To address this requirement, Reporter has been designed to work with almost a dozen different file formats, including the Triple-S file format, which is the data interchange standard for transferring data between different software packages (e.g. SNAP, Nebu, Pulse Train, and ConfirmIT). Reporter can also be pointed to those projects using Dimensions for collecting data over the web, or using telephone and display topline results.

In addition to opening different sorts of data files, Reporter allows the user to merge cases from multiple survey data files, as well as merge variables from one source file with another survey data file. Furthermore, because the application is built using the Dimensions Data Model, users can merge data from different formats.

In the case where the survey data needs to be exported to another file format, Desktop Reporter builds upon the functionality already offered in Quanvert and enables users to export to half-a-dozen different data formats, including Triple-S, SAS, and the new SPSS Dimensions Data File.

### **2. Control the way information is expressed in the report**

#### **Q: How can Dimensions help me deliver targeted results?**

Quanvert presents the user with a sufficient feature set to help them gain control over the way information is expressed in the report. Options allow the user to edit variables, filter and define rules for suppression, run statistical tests and apply weighting to the data. However, while the application offers a fine granular degree of control in terms of working with the results, it does not offer any inroads for *managing the total volume of information*. In the previous section, the paper highlighted the issue with an ever-increasing amount of data being collected, and where a delay with making an informed business decision can have financial repercussions.

Reporter allows users to identify those significant results via one-click sorting. In addition, a special table can be created using the Product Difference Test. With this test, the user can apply statistical testing (using the column proportions or column means test) to all combinations of categories across a number of variables and produce a table showing on those significant combinations. There is also a one-click option for modifying an existing table to show only those results that are significant.

### 3. Allow users of all levels to rapidly build tabulations

Q: How does Dimensions empower users with different skill levels and different reporting needs?

When it comes to addressing different reporting needs, summary information is useful, but it is not always enough. Quanvert does a sufficient job of allowing users to preview the data before building a table, produce standard summary tables, or focus on respondent-level data to produce profile reports. Therefore, in terms of addressing different reporting needs, Quanvert can provide different inroads based on the requirements of the user.

However, when it comes to application usability, Quanvert is a hostage to its 16-bit architecture. The design of the application user interface directly affects the ease of access to core functionality in the product. Users who work with Microsoft® Office applications like Word and PowerPoint® expect a degree of flexibility with regard to the ability to drive the application via the user interface (drag-and-drop, right-mouse click for options, double-click to activate, show/hide toolbars, etc). Reporter is built with Microsoft® WinForms – the Microsoft®.NET framework API – which allows the developers of the software to employ the same types of standard controls found in modern application design. For the novice user, standard functionality is easily accessible via the clean interface. Users are not presented with a barrage of options to choose from, rather, more advanced features are accessed via supplementary menus. By adopting this design approach, supported by context-sensitive help (i.e. where a user can hit the F1 key and be presented with relevant help on the task at hand), it has been found that users can quickly become familiar with the core functionality of the product, and can then focus their efforts on producing the right information for the right people.

Earlier in the paper, some of the inherent advantages with moving away from the desktop mode and into the online realm were discussed. Nevertheless, there will be times when it is simply not feasible to be online. By taking advantage of the Microsoft® WinForms technology, SPSS has implemented an application design that is *modally agnostic* and addresses the needs of both the desktop *and* online user scenarios. That is, the user has access to the same feature set and the ability to drive the application via the user interface in the same way.

Implementing a product that is both intuitive and easy to learn and one that employs a common feature set regardless of whether the mode of access creates a knock-on benefit with training. The initial learning curve is gentle enough for beginners to get up and running with the standard functionality and they are not required to learn a different version of the product when switching modes.

### 4. Produce reports that are readable and understandable

Q: How do I ensure that my reports have a consistent look and feel?

There are obvious benefits with being able to define key settings and store these in a template. Quanvert users can already define these elements once and then save them as part of a specification file or template. However, because of its desktop implementation, there is no adequate mechanism for managing the distribution of these templates across a group of geographically dispersed users.

Where a group of users from the same organization is busy working on producing a report for a particular project, there needs to be proper management of the templates to enable a

consistent look and feel. Using Reporter, it is possible to set up project templates such that the group of users working with the data will produce reports have a consistent look and feel.

### **5. Publish reports in the right format, for the right people**

Q: What are the options for Dimensions users in terms of publishing the results?

Publishing the report is reliant on the ability to have sufficient control over the aesthetics, include appropriate data visualization (e.g. charting the results) and finally export the report in the right format.

The ability to manage the aesthetics of a report in terms of specifying the layout and style will enable the user to create more visually appealing output. For instance, rather than displaying a plain table of data, a user may wish to format the output and adjust the size of the font, select a suitable colour for shading the cells and include the company logo on the report. Instead of embedding the aesthetic control into the actual content of the report, modern day web design uses Cascading Style Sheets to enable the separation of document content from document presentation. This separation allows for greater control over the look and feel. Reporter uses the same technology to control the way information is displayed. Users have the option of working with one of the existing CSS style sheets, or they can build their own style sheet to align with corporate marketing regulations.

If a picture speaks a thousand words, then using a visual device like a chart might be the preferred mechanism for conveying the results of a survey. Choosing the right chart is dependent on the information to be charted and the selection of available chart types. Quanvert Reporter users may choose from a selection of different chart types in Reporter, or build their own custom chart type to use.

Exporting information from an application and into another file format is often the simplest way to make the report accessible in a concise and easy-to-digest format, e.g. Microsoft® Office PowerPoint®, Excel®. In survey research, this output may have been appropriate for standard studies where the project was designed as a standalone piece of work where analysis of the results takes place *within the boundaries of the study*. However, initiatives like Enterprise Feedback Management<sup>3</sup> acknowledge that in order for results to be actionable, there needs to be tighter integration with the existing management reporting structures. Report results in the form of table definitions must be managed centrally with the ability to publish the results to a corporate portal, where the information may be combined with other sources of data to present a more complete view.

## **6. Conclusion**

Quanvert was specifically designed as a desktop tool to enable the *user* (e.g. market researcher, analyst) to focus on the analysis of survey results in order to produce *reports*. However, since its introduction in the mid-1980s, there have been a number of significant developments, which have changed the technological landscape and created new opportunities for taking a fresh look at how to address the needs of the market.

First, the definition of *user* has expanded out from the traditional desktop space and into the online realm. Meeting the needs of a globally distributed audience requires a different

---

<sup>3</sup> For more information, refer to the Gartner Research paper *Make the Transition From Surveys to Enterprise Feedback Management*.

approach. Organizations acknowledge that moving away from the desktop space and centralising the system offers clear advantages in terms of managing the distribution of results, reducing the time to rollout an application, and with instilling greater faith in the data by enabling *one version of the truth* across the enterprise. SPSS accepts that there will be times when it is simply not feasible to be online, and has embraced a product design that is both intuitive and easy to learn and one that employs a common user interface regardless of the mode of access.

The reporting process does not exist in a silo, and SPSS recognises the importance of providing better interfaces to support the upstream and downstream efforts. Since data is collected via multiple modes or channels, the interfaces have been expanded to work with almost a dozen different file formats, in order to support the upstream data collection. When it comes to supporting downstream dissemination of results, SPSS has listened to the market and established that the definition of what constitutes a *report* has broadened to include presentations, documents, online reporting, web based delivery, non-web based electronic reporting, static reporting and interactive data visualisation. Reporting applications can run pre-defined report definitions and output the results of a survey directly to a variety of standard file formats as well as offering the full flexibility to slice and dice the data, and perform ad hoc analysis.

Historically, the task of producing analysis and reports was handled by heritage applications like Quantum (for processing the data) and Quanvert (for analysis and reporting). This process relies on a linear and manual series of tasks where the data is first tabulated into *decks* of tables, key findings are manually extracted, and the insight is ultimately turned in to single document and delivered to the client as a static report. Because management decisions based on the results of survey research are often balanced alongside sales and marketing metrics, there needs to be tighter integration with existing management reporting structures. Some organizations are already forging ahead with centralized systems that allow them to fully engage with their current (or prospective) customers through targeted feedback programs. The information is then utilized throughout the organization to drive business improvement.

In conclusion, the key tenet of this paper argues that when trying to determine if Quanvert can continue to deliver to the needs of users, the scope must be widened to take into account the broader significant developments, which have changed the technological landscape. It is *more* than just about replacing Quanvert functionality; it is about understanding the business processes around it. In the new world of requirements, one product in one environment is simply one component of a *full* reporting solution. While SPSS provides a next-generation product that Quanvert users will embrace, it is only part of a broader solution that (a) combines with customers' Microsoft® Office® automation tools and statistical engines to form a well-integrated analytic workbench, and (b) plays well with customers' online and offline reporting capabilities, digital dashboards, OLAP tools and so on.

Many successful organizations – including businesses and sports teams – agree that analyzing data, and incorporating the results into the decision-making process, leads to better decisions<sup>4</sup>. In order to drive the widespread use of data in decision-making, a full reporting solution will provide the right information, in the right format, to the right people.

<sup>4</sup> See *Competing on Analytics: The New Science of Winning* by Thomas H. Davenport and Jeanne G. Harris

## References

- ESOMAR** (2005) “ICC/ESOMAR International Code of Marketing and Social Research Practice”
- Gartner Research** (2005) “Make the Transition From Surveys to Enterprise Feedback Management”
- ISO** (2006) “ISO 20252: Market, opinion and social research — Vocabulary and service requirements”
- Davenport, Thomas H. and Harris, Jeanne G. “Competing on Analytics: The New Science of Winning” Harvard Business School Press, 6 March 2007
- Treanor, Jill “Nationwide fined £1m over laptop theft security breach” **Guardian Newspaper** 15 February 2007.

## About the Author

**Nick Read** is a Product Strategist working in Research and Development at SPSS. He can be contacted at SPSS Limited, 65 Maygrove Road, London NW6 2SP; tel. 020 7644 6000; fax 020 250 1524; e-mail nread@spss.com

# Surveys and Technology – Polishing the Crystal Ball

Hugh Neffendorf

## Abstract

This paper provides a look at new technology developments which are relevant to survey computing and what they might offer. It is the latest in a series of similar presentations (about every ten years) which have been given at previous conferences.

Hugh Neffendorf is currently working on the impact of new technology on use and dissemination of official statistics, so the crystal ball is not too dusty. However, it seems to be getting larger and cloudier in some areas, as the pace of technology develops, as the old limitations diminish and as the world becomes hungrier for information. It is noted that the advance of hardware and operating systems can now be taken for granted, while the social implications of technology become more prominent. This presentation will offer a view of a survey environment in which most things now appear possible, but not all will deliver their promise or produce reliable information.

## Keywords

Technology, future, survey computing, Web 2.0

## 1. Background

This paper continues an irregular series of future-gazing; how will emerging technologies affect surveys and survey computing?

The author's attention to reporting on technology developments began in 1982 with a paper on survey software packages available for the new phenomenon of microcomputers. This included some future-gazing. In those days the expression Personal Computer hadn't been coined. His resulting list of packages was then maintained in a series managed by the SGCSA (now the ASC), but PCs became so commonplace and mainstream that it was no longer necessary to distinguish them from other computers. Also in 1982, he made a presentation at a SGCSA conference on the future of technology in survey computing.

In 1992 the author presented a technology overview at the first SGCSA International Conference in Bristol. Again, in 1993, a paper on the topic was printed in the Journal of the Market Research Society, and later a presentation, 'The Future Environment for Market Research: Technological Change' was an invited talk at the MRS Conference.

Finally, in terms of technology updates, the author presented a new development in hand-held personal interviewing, termed HAPI, at the ASC conference in London in 1999. A live



demonstration included features of interviewer handwriting and voice recognition plus portable address searching from an inbuilt gazetteer.

The pace of technical development continues to accelerate, and it is time for another look to the future.

## 2. Looking back helps with looking forward

It can be embarrassing when people recall your early projections and remind you of them later. One of the famous ones, in the 1982 presentation, under the heading “Hugh’s Homilies”, was that the IBM microcomputer with PC-DOS (to become the PC) would not succeed! What was meant, of course, was that technically it did not deserve to succeed, compared to other offerings. The power of image, hype and marketing was not overlooked in subsequent projections!

The 1982 presentation also noted that computing was being de-mystified but that most application needs were still met adequately at that time by relatively simple processes and conventional software. It was predicted then that SPSS would still be the leading software in 2000. An interesting brief mention was made about potential for subliminal data collection and ethical issues.

In the 1993 MRS paper, some highlights of the previous decade were considered to be:

- Computer aided personal interviewing
- Integrated survey packages
- Adaptive survey methods
- Fast tabulations
- Online panels and databases
- Automated fieldwork administration
- Online data entry and coding
- Survey geocoding
- Friendlier analysis software
- Much improved presentation of results

While some disappointments included

- Optical character and voice recognition
- Expert systems
- Hand-held computers for interviewing
- Productive use of database technology
- Better survey processing and analysis

Some emerging technologies to note in 1993 were:

- Digital multimedia
- Home entertainment advances
- Subliminal data collection
- new point and click operating environments
- online activity (still a crude concept then)
- open software and standards
- use of the new generic tools (e.g. spreadsheets), although specialist software would continue to be necessary

- more secondary analysis of public data, such as the population census
- data fusion and modelling
- geographic information

It was considered that voice, character and image recognition were still a long way off, and expert systems were still not expected to bear early fruit.

At that time, the predictions were still very technology oriented. Some of the things that fascinated us 10 or 20 years ago are now so commonplace that they hardly need mention today. Early discussion was about hardware or operating system advances, but they are barely at issue now.

The early papers started to mention communication technologies, online working, user interfaces and human factors, but not strongly. What is amazing, considering its overwhelming implications now, is that the author and the audience didn't see the Web coming!!

### 3. What's here and what's around the corner?

The developments that matter are still those that present the opportunity to work faster, more accurately and with better presentation. But these are now overlaid with the possibility to obtain more insight, to use data linkage rather than just interviews or observation and to explore data more deeply.

We're used to the hardware now, and operating systems are not a big debate for us, mainly a sideshow. Yet much of the application software on our desktops is much the same as before! But what is available through the Internet and through data resources is quite different.

It is always difficult to second guess technology, but history tells us we will underestimate change. Here are suggestions of a few likely changes in conventional technology in the next ten years or so:

- We will still have PCs, still have Microsoft, but with pressure from others such as Google to create better environments
- Continuation of Windows look & feel – moving to more browser-like interface
- Increasingly dynamic and animated presentations of results
- Much faster computing; grid and parallel computing
- No meaningful restrictions in terms of speed, storage, bandwidth
- More ubiquitous Web – little distinction between Web and desktop; it is likely that users will be online seamlessly in a few years time
- Widespread wireless access; Wimax allowing VOIP (Voice over Internet Protocol) anywhere and communication over kilometres rather than metres
- More open source, plus other resources to plug in or use (e.g. Google Earth, free applets and APIs)
- More online data and search engine approach to finding data (both on Web and desktop); searching for data in tables and cubes, not just documents
- But, rising importance of copyrighted and Digital Rights Management data (an asset and a problem!)
- Survey metadata coming of age and being used widely in analysis applications (or is this just wishful thinking?)
- More use of modelling, data fusion and understanding of uncertainty and risk in decision taking (more wishful thinking?)

- Integrated software and data across the range of communicating devices – PCs, phones, handheld devices, home entertainment, Internet access points
- Voice recognition will be used in online surveys, but with a limited vocabulary
- Greater concerns about security and privacy distorting data collection and boosting indirect ‘dataveillance’ approaches
- Web2.0 – see below

The Web is all over us. There are obvious conventional uses, delivering traditional interviewing or results through the online mechanism, but the dramatic evolution of Web 2.0 and social networking is the big news.

#### 4. Talking about ‘mySurvey’ generation

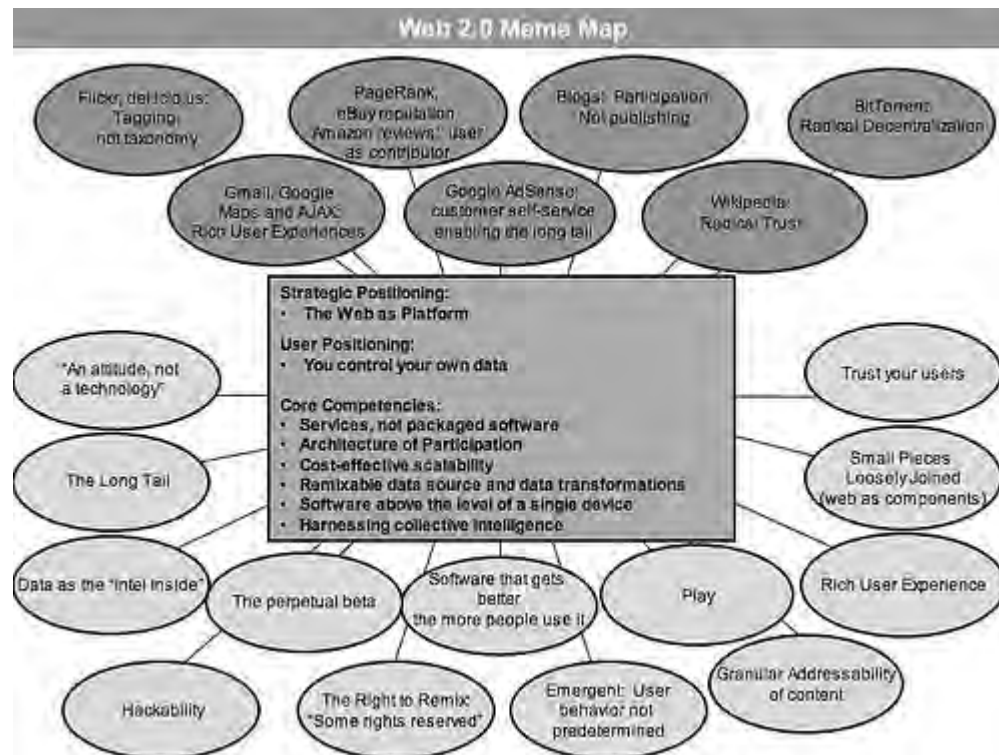
Web 2.0 is a widely used term for a range of Web experiences in which user-generated content dominates and user communities are established. The term was reputedly coined in a conference brainstorming session involving Web guru Tim O’Reilly.

O’Reilly’s dialogue and a subsequent article formulated a sense of Web 2.0 by example:

<b>Web 1.0</b>		<b>Web 2.0</b>
DoubleClick	-->	Google AdSense
Ofoto	-->	Flickr
Akamai	-->	BitTorrent
mp3.com	-->	Napster
Britannica Online	-->	Wikipedia
personal websites	-->	blogging
evite	-->	upcoming.org and EVDB
domain name speculation	-->	search engine optimization
page views	-->	cost per click
screen scraping	-->	web services
publishing	-->	participation
content management systems	-->	wikis
directories (taxonomy)	-->	tagging ("folksonomy")
stickiness	-->	syndication

The list could go on and on. But the message was that user involvement would often replace supplier-based delivery of content.

In the same article, a meme map was presented, which also gave an indication of things to come:



Enterprise 2.0 is a business view of the Web2.0 phenomenon. It has adopted a design concept acronym SLATES, defined as:

\*S\* = Search

\*L\* = Links - most linked most relevant

\*A\* = Authorship - everyone has something to contribute

\*T\* = Tags - content categorisation

\*E\* = Extensions - use algorithms to find user patterns and make recommendations

\*S\* = Signals - alert users to updates and content

The pointer for survey computing specialists is that an adaptive approach and user engagement are becoming a central part of business computing.

The 'mySurvey' generation has vastly different expectations of dealing with organisations based on experiences with eBay, Amazon, Facebook, Google, Flickr, del.icio.us, etc, all of which recognise the importance of innovation in keeping their customers happy.

Existing types of services appear to be fragmenting. Instead, people are:

- Finding
- Sharing
- Playing
- Talking
- Recommending

An interesting manifestation of Web 2.0 is 'mashups', a term derived from the music industry in which anyone can create a site by borrowing elements from other sites. Some of the big providers have introduced APIs (application program interfaces) to support mashup development, not always for altruistic reasons! Mashups have vast potential to present results and gather new data, although one wonders about the validity of some data association.

In Tim Berners-Lee's view of the 'Semantic Web', software can read and interpret Web data sources, with potential for automated mashups. This implies a strong element of attention to standards and various advances on html – xml, gml, etc. The DDI and SDMX developments are particularly relevant for survey computing.

The major impact of Web 2.0 is not so much that blogs, wikis, and other current manifestations of social networking will have an important role in survey computing, but rather that client and respondent expectations and technical familiarity will change rapidly. This presents a challenge to those involved in survey research but also, clearly, great advantages.

## 5. Bias in new surveys

It will be increasingly tempting to use the Web to carry out surveys. This is happening today, but mainly as conventional surveys, administered online. However, use of Web 2.0 concepts will also emerge in survey research. But do we understand the implications for bias in results, both from unrepresentative samples of the population and from different behaviour of respondents in that environment?

Today's main users of social networking tend to be computer geeks and youths. It is very clear that most 'normal' people still shun blogging, wikis, social bookmarking, etc, despite attempts to introduce them. Partly this is a learning process, but many people feel uncomfortable in a sharing environment, and some always will.

These are not new concerns. For example, CATI surveys present some of the same issues. The respondent population has always been restricted to those with phone access and, while most people now have this, mobile phone penetration makes it much harder to develop representative samples. And it is understood that people behave differently on the phone than in a face-to-face interview.

It is likely that Web surveys will cause different levels of 'satisficing', where people try to anticipate what answers might be wanted, or 'cognitive dissonance', where people can feel discomfort about apparent inconsistency between their answers and other information (in mashups?), about their own behaviour or that of others. Socially desirable answers may also become more prevalent as green or politically correct issues are increasingly presented, while unconstrained control of the Web may allow more malicious misreporting of true responses. Such effects need research.

A growing form of 'polling' is reflected in the SMS flocking effect, where alerts on issues can lead to a tidal wave of 'votes' influenced by factors that may be emotive or evanescent. The Big Brother or X-Factor TV shows are prominent examples.

For these types of reasons, many client agencies will still request random probability samples and face-to-face interviewing. However, it will be overwhelmingly tempting to use social networking in survey research, and in many ways it will bring advantages in respondent stimulation, quasi-personal interviewing, Delphi-style research approaches (in which people interact and modify opinions iteratively) or through interaction with other processes, like online purchasing or product reviews.

Potential for abuse of social networking results can make it difficult to distinguish quasi- and real research. There is clearly a risk that the trust in survey results could decline with the

impact of social networking and that this will amplify the trend toward ‘dataveillance’ using indirect surveillance approaches.

An increasingly important data development will be the use of micro-simulation techniques to develop representative but artificial populations or samples by randomly sampling from known statistical distributions of behaviour (stochastic modelling). In a Web context, Alan Cooper described ‘personas’:

- Personas are hypothetical archetypes, or ‘stand-ins’ for actual users that drive the decision making for interface design projects.
- Personas are not real people, but they represent real people throughout the design process.
- Personas are not ‘made up’; they are discovered as a by-product of the investigative process.
- Although personas are imaginary, they are defined with significant rigor and precision.
- Names and personal details are made up for personas to make them more realistic.
- Personas are defined by their goals.
- Interfaces are built to satisfy personas' needs and goals.

## 6. Loyalty?

An interesting indirect development is through the emergence of customer relationship management (CRM), allied with consumer lifestyle behaviour databases.

Tesco is a notable leader in this field, with its loyalty card through which vast amounts of customer purchasing data is tracked and used to plan stock control and campaigns, even with personalised or local store differentiation. Tesco makes it clear that personal information is never disclosed, but the effective use of this data has helped to make it a market leader.

Is the data from millions of customers representative? It probably is in terms of what Tesco needs to know, or near enough for targeting purposes.

## 7. Really Big Brother?

The CRM example may begin to feel like we are being watched. However, it is the tip of the iceberg relative to possibilities of ID cards, personal biometrics, RFID devices, generic lifestyle databases, home entertainment tracking, etc. And we are beginning to notice the growth in CCTV and possibilities associated with road pricing technologies. Concerns go beyond the increasing difficulties with classical survey methods and quality of data to issues of manipulating community responses and commercial or political abuse of results.

Is this a problem? That depends on one’s outlook; whether it offers the potential to improve services and products or is seen as a potential threat to privacy and personal freedom. Interaction with anti-terrorism initiatives and legislation is tending to cloud the service versus privacy debate even further.

But it is a form of survey research, or a technological replacement for survey research as we know it.

With that thought, the crystal ball can be put away for another decade!

## References

- Hugh Neffendorf, Statistical Packages for Microcomputers, *The Professional Statistician*, June/July 1982 (also published in *The American Statistician*)
- Hugh Neffendorf, Future technology in survey computing, unpublished presentation, SGCSA Conference, 1982
- Hugh Neffendorf, Survey Computing in the 1990's - A Technology Update, 1<sup>st</sup> SGCSA International Conference, Bristol 1992
- Hugh Neffendorf, Survey computing in the 1990s: a technology update, *Journal of the Market Research Society*, 1993
- Hugh Neffendorf and Jon Hall, A big survey can be a HAPI survey, ASC Conference - The Impact of CAI on Large and Complex Surveys, London, 1999
- Tim O'Reilly, What Is Web 2.0? Design Patterns and Business Models for the Next Generation of Software,  
<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html?page=1>, 2005
- Tim Berners-Lee, James Hendler and Ora Lassila, The Semantic Web - A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities, *Scientific American*, May 2001
- Alan Cooper, *The Inmates Are Running the Asylum: Why High-Tech Products Drive Us Crazy and How to Restore the Sanity*, 1999
- Clive Humby, Tesco Clubcard, and the value of its information, Presentation to RSS Statistics User Forum 2006

## About the Author

Hugh Neffendorf is Managing Director of Katalysis Ltd, a consultancy in management, technology, official statistics and information services. He is also Visiting Professor at the University of Southampton, Department of Geography. For ten years he was Chairman of the ASC/SGCSA.

# Case Study: Evolution of Web Interview Capabilities in a Large Commercial Setting

Adi Cohen

## Abstract

In 2005 Yahoo! embarked on a new research initiative by creating a full-time team of research professionals to focus on a web interviewing infrastructure. While still in its formative stage, the team has so far accomplished the installation of a state-of-the-art web interview application, built an internal programming service bureau and launched an in-house research panel.

Yahoo! is a data-driven organization that is also quite cost conscious. Creating efficiencies in its research operations is highly valued for generating the maximum number of usable insights from a limited research budget. Raising organizational competency for running in-house (“on the network”) surveys means use of costly full-service vendors can be reserved for instances where they add the most value (conducting research “off network” or taking advantage of their unique competencies). This case study will give insight on the end-client view of the changing MR technology industry and will illuminate the choices that commercial end-clients make when working directly with the MR suppliers.

## 1. Background

Yahoo! Inc., the first large-scale directory of the Internet and major Internet brand, is still arguably the most visited website in the world. The business model of Yahoo! provides a wide arena for survey research and analysis. Ways in which survey data drives our business include:

- Measuring ROI for advertisers
- Gathering feedback from customers on usability and new product development
- Conducting surveys to support our own marketing efforts

Furthermore, the company’s “free” access to an enormous community of customers is enviable, as is our ability to target survey invitations based on a customer’s behavioral and demographic profile.

Another asset of the company is its size and reputation. One might be tempted to underestimate this in relation to survey research. However, given the state of the art of survey technology, the ability to influence product is significant. For example, our need to provide self-service ad-hoc reporting solutions to our internal users drove an overhaul of our vendor’s reporting module. This kind of company-specific influence resulted in a significant upgrade for the benefit of clients of all sizes and in all industries. It demonstrates how the relationship of the web survey software industry with non-traditional, non-Market Research Supplier



clients has been helpful for product development, broadening product appeal and furthering the ultimate goal of “democratizing research.”

Another essential point is that we are an Internet technology company, NOT a market research software development company. Yahoo!’s engineers and product managers represent expertise optimal for the development of a worldwide dot-com, not for building specialized business applications like web interview software. Taking the application in-house meant that the relationship between the MR vendor and the non-MR client took on a different flavor, and both parties had a steep learning curve.

The evolution of Yahoo!’s web-based survey capabilities provides a glimpse into both the challenges and successes that arise, even in the most favorable conditions. One can imagine that similar challenges exist in other settings as well.

## 2. Size and Scope of research at Yahoo!

Many of Yahoo!’s research objectives can be achieved using sophisticated web-based interviews. There are over 100 customer researchers spread throughout Yahoo! and their web survey activity varies from those who are active every week to those who gather data once a year. Since mid 2005, when it was first created, the Research Operations group’s role has grown to managing more of the tools and services on which they all rely.

Researchers use the web interview tool for screener surveys for various qualitative methods, including in-lab usability studies, “captive web navigation”, focus group sessions, online salons, etc.

Quantitative opinion research at Yahoo! varies from simple self-selected feedback forms to sample-based surveys which require major planning and analysis effort. In a recent one year period, we conducted over 1,000 in-house surveys of nearly 10 million different website users worldwide. Even with the new in-house system, our overall goal is to send the work where it can be executed most effectively and efficiently; especially complicated or specialized projects are still outsourced to a vendor that has a process and/or project management staff already in place and the necessary domain expertise.

For example, while some highly customized advertising effectiveness research is done in-house, usually on a one-off basis, the bulk of such research on the Yahoo! network is done by third party specialists. These firms benefit from efficiencies of scale from doing so many identical studies both on and off our domain; furthermore, end-clients of such research prefer the external validity of using a neutral party. There are also instances where it is better if the respondent does not know Yahoo! is the client, and these also continue to be contracted externally.

Lastly, Yahoo! is an international company, addressing research in global markets in multiple languages. Any researcher anywhere in Yahoo! can contact the Research Operations team and request access to the survey tools and programming support. While research tool accounts and operations are managed centrally, the researchers themselves are decentralized.

A critical factor in a global company is the ability to provide surveys with multi-language capability. This is one of the evolutions that came about when the company changed to the new platform, as described below.

### 3. Historical approach to research at Yahoo!

Yahoo! has always enjoyed a relatively easy access to sample, at least with regard to its own customers. From its earliest days, even before the drive to prove web advertising as a lucrative business model, Yahoo! didn't need anything beyond simple feedback forms. There was no need to go outside the company for that level of technology – but it required a technical person to be involved in every project.

As Yahoo!'s popularity grew, so did the volume of survey projects. It became more efficient to build a tool. So in the late '90's, Yahoo! engineers developed a tool that would allow people to get surveys done without needing engineering support for every project.

This home-built application worked well for simple one-page forms. It suited the purpose of those who wanted something that was easy to put up on the website and easy to attach to user level data (i.e., how people use the network, where they click, etc.). It was also well suited to the extremely high volumes of traffic that a typical feedback form might receive on the Yahoo! network. The application provided information to product managers about how people were using products, or the information was aggregated and shared with advertising customers.

However, sophisticated marketing research requires conditional logic and other automation of good survey design (e.g., randomization, piping, branching, etc.), which were not built into the intelligence of the home-built software. When more sophisticated surveys were needed, the researchers had two choices: Accept the limitations of the existing tool and modify the methodology to suit; or, if they had the budget, and a significant project, they could outsource the work. Money often became a determinant of whether the project got done.

As more researchers went outside the company for their projects, it became clear that this was not cost effective. For one thing, we were spending quite a bit of money for small projects simply because we didn't have the in-house technology. In addition, there is a lot of overhead (and lag time) involved in working with outside vendors, including getting all the internal groups involved to create purchase orders, approve the project, manage the project, etc. Compared to being able to draft a survey and put it up using one's own tool, the cost of using consultants became increasingly impractical and the time-consuming process of going outside was at odds with the fast-paced, Internet model of the company.

As the pool of researchers grew (in size and sophistication), there were more users and demand for world-class tools. This coincided with maturing of MR web interview software to meet the requirement of the researchers as well as those of an enterprise software customer. Therefore, at a critical point, the internal customers for more sophisticated research led the search for an outsourced product that Yahoo! could bring in-house to replace their simple home-built tool.

### 4. Implications of Self-Service Tools

Yahoo!'s vision is to empower our researchers by making the research platform as self-service as possible. I don't think we are alone in this vision. As respondent sources and interview technology options proliferate and develop, I believe we will see enterprises of all sizes take a do-it-yourself approach to soliciting customer feedback. They will want as much as possible to put the tools of research at analysts' immediate disposal. For smaller businesses the issue might be cost savings. For very large organizations, centralized functional

departments can sometimes be too slow moving and it becomes important to democratize access to critical business functions such as research. At the same time, there can be a danger to universal access to such tools, as we will see later in this paper.

Yahoo! searched for a tool that would be extremely usable, and yet flexible enough to do all that the research people need, and reliable enough to deploy in a customer-facing business critical environment.

This combination is challenging to find. If the tool is feature-poor and lacks capacity for high interview volumes, researchers would have to outsource many of their projects. On the other hand, when the tool is able to do anything, the power can come at the expense of usability. Researchers are not traditionally technical. They don't spend a lot of time learning the ins and outs of software. Writing scripts and navigating complex interfaces can be a complete impediment to usage.

As is common in other areas, there is a gap between the ideal and the reality. Our experience with the tool turned up several factors that undermine the pure self-service model:

- a. The reality of usability versus capability
- b. Centralized support was needed
- c. Consequences of making surveys too easy.

#### a. Usability vs. Capability

The reality of the tool that was selected was that it was robust and well-designed in many ways, however, researchers tended to stumble on certain tasks that stood in the way real self-sufficiency. Unless the researcher takes the time to really learn the product, which most don't, they needed frequent support to use it. While this continual need for tech support had an unexpected benefit, giving the small central operations team more opportunity to oversee survey activity on the website, it did not match the original product vision. Furthermore, the operations team became a bottleneck.

#### b. Centralized Support

As part of getting a new self-service tool, Yahoo! created a centralized team to oversee the operations and support the research area. The team is responsible for training, supporting users, and helping them with development and implementation, if needed. In short, even while providing a self-service tool, a centralized operations support group was needed.

This created an interesting dilemma. Partly by design and partly as a holdover from start-up roots, the company is inherently decentralized, with each business unit having autonomous control of their budget, and expected to supply their own resources, whether it was for research, marketing or engineering. Yet, the research operations group is centralized. Virtually everyone can request our services since we made the decision to eliminate the bureaucratic layer introduced by internal cross-charges. Therefore, it has been the case that the demand for support can outstrip the capacity to deliver it.

To help ensure that surveys are done by qualified research professionals, access to the centralized service is restricted to researchers only. A business unit that wants to put up a web survey using our system needs to find a researcher. Access, rather than money, has become the de facto prioritization regime.

### c. Unforeseen consequences of making surveys too easy

In addition to getting the new tool, which made creating sophisticated surveys easier, there were internal advances in the ability and speed to administer highly targeted recruiting. With both targeting and the tool easy, there is not much in place to evaluate if a particular survey project is important. While reducing the incremental costs of producing surveys, there is the very real cost of respondent burnout and erosion when people are over-solicited or asked to complete an onerous survey. Sometimes, the cost to the overall user experience can be undervalued when the ability to survey becomes so cheap and easy to do. This is one of the areas we are working to address.

## 5. Steps to creating the new approach

Key steps towards acquiring, adopting and maintaining the new systems:

- A. Draw up criteria and make selection
- B. Vendor-client technical partnership
- C. Manage the users (training, accounts, etc.)
- D. Learn on the go
- E. Create the support infrastructure

### A. Draw up criteria

A list of criteria was drawn up based on the needs of the researchers and what was already possible or not possible in the existing tool (see appendix). It was essential that the tool be optimized for web surveys, and not just one that creates web forms as a by-product or add-on to another function (e.g. CATI software or the previous in-house tool). The tool would have to be robust enough to cover the breadth and wide-range needs of the organization. It needed to support advanced questionnaire design. Ideally, sample management and online reporting would be well integrated.

Another criteria was that the vendor had to be sophisticated enough to work with Yahoo!. More than just software, the successful vendor would have to provide support and be there as needed.

Evaluation of ease of use versus comprehensive feature set was also critical. In the end, Yahoo! chose to go for capability over simplicity. The evaluation, selection, and implementation process took about a year.

Conformit was eventually selected as the qualified enterprise survey tool. While it required more training for the end-users than the old application, it fulfilled many of the criteria we had articulated.

### B. Vendor-Client technical partnership

The Yahoo! research environment presents a different kind challenge for a web survey system. Compared to MR companies with staff dedicated to working with survey technology, the IT resources we could access had no such specialty. Initially, there was a steep learning

curve for both parties, but ultimately we learned that the Client and Vendor need to partner closely.

This might seem obvious, but it can be easy to underestimate just how much back and forth will be required when a complex survey system is installed in an already complex enterprise setting.

The technical side of the installation required intense project management. This included software installation, integration with systems, selecting/installing and configuring hardware, situating hardware, and insuring that domains resolved to the right machine.

Strong security policies within the yahoo.com network presented unique challenges. While the vendor was present as a consultant during the process, internal engineering resources were also needed to be on demand. Only a dedicated operations team can manage such a complex system throughout its lifecycle and with organizational requirements for up-time, upgrades, new issues, systems integrations and business continuity planning.

### C. Manage the User Side

In order to provide a self-service product, the research team had to manage accounts and train the end-users. We needed an intuitive self-serve tool because we couldn't invest a lot of time in training people.

Typical researchers are quite happy to use a web survey builder on a self-service basis, however, they are rarely able to invest much time in learning complex tools and scripting languages. The heavy tool users may, over time, become proficient through constant support. However, light users require just as much up-front support and then there is limited return on that investment due to the typically low level of use onwards.

### D. Learn on the Go

The solution was to be less rigid within our workflow model. We gave up on the pure self-service model and created an internal Service Bureau in partnership with an offshore supplier of trained survey programmers. This hybrid model for survey production has proven successful for a couple reasons:

- It is cost effective. What may take an inexperienced researcher 2 days of work to program takes the full-time programmer ½ a day. The result is that we use a less expensive and more efficient resource while gaining time for researchers to do analysis and dissemination of results; i.e. what they were really hired for.
- Having a sophisticated tool with a good graphical user interface (GUI) becomes a real benefit. Researchers can 'see' how their survey is programmed and access other tools (such as data export and online reporting), but they don't need to figure out how to do the complex logic and web page layout.

The operational cost to outsourcing was the need to develop a reliable process for communicating with an offshore programming company, and training our researchers to employ it. This downside was definitely outweighed by the benefit of eliminating an inefficient in-house training program.

On-going training is still a challenge, but now we focus on those skills for which our research community already has proven aptitude: process and project management rather than technical specifics.

#### E. Create the infrastructure

As part of the decision to move to a more complex tool, it was also clear that Yahoo! needed to create a world class research infrastructure. The elements of this infrastructure include:

- A state-of-the-art web interview application
- A centralized research support group, including an internally managed programming service bureau
- An “in-house” research panel

The research panel is a group of customers who have agreed to be contacted for research purposes. This is the newest addition to the research infrastructure. The panel becomes another resource, like the tool, that researchers access on request. The centralized research group mediates the access to the panel in order to ensure there is no over-contact, etc.

Organizations in other industries have long moved to using proprietary panels. For them, it is more efficient over the course of an entire research programme. However, in the case of a company like Yahoo!, with most of our customers right at hand on our website, we don't directly gain in cost-effectiveness with panels. But there is a hidden cost in that we do disrupt the web experience of some of our customers on the network. In addition, panels offer benefits we don't get with random selection.

With a pre-screened panel, we possess background information that allows for sample selection based on attitudinal attributes, demographics and self-reported off-network behavior. Managing research against the panel becomes practical when the database of users is well integrated with the survey tool. While we have a reasonable capability in this regard currently, the future points towards more need for fusion of attitudinal and behavioral data; both for the richer insights to be gained and for deepening respondent engagement.

## 6. Benefits & weaknesses of current approach

The new approach has yielded much success, but some challenges remain.

Challenges:

- A. Platform Issues and Ease of Integration
- B. Security & Control in a Self-Service Environment
- C. Reporting and Presentation Modules

Successes

- D. Vendor/Client Partnership
- E. Multi Language
- F. Ad Targeting Platform

### A. Platform Issues and Ease of Integration

The old "homegrown" tool integrated nicely with the creation and scheduling of survey invite ads. It was built by Yahoo! engineers on the Yahoo! platforms, leveraging their knowledge of internal systems. With the externally built tool, integration of this sort becomes much more difficult; also, the bulk of the related engineering support is also now outside the company. Further complicating matters is that Yahoo! primarily develops on open-source platforms (Linux, FreeBSD, MySQL, PHP, etc.), while our vendor developed survey system is built on Windows™ technologies (Win2K, IIS, ASP/.NET, MSSQL).

Our on-going challenge is to bring the right resources to bear on issues such as working with the firewall, protecting the database from hackers, back-ups, up-time, disaster plans, etc. All of these are harder to implement in an environment with relatively few Windows™ experts and are therefore taking more time than perhaps we would have liked.

### B. Security and Control in a Self-Service Environment

While the goal of having a self-service interview tool is well intended from a business perspective, it presents additional considerations. One of the consequences is that it effectively allows researchers a way to publish directly on the website, and to collect data from our customers. Integrating the company's customer contact and data policies with the needs of the research community is an area where we've made significant progress but much work remains.

QA processes designed for the rest Yahoo! can take too long for the speed with which research happens. Researchers are under pressure to get answers quickly, and thus demand fast response from the resources. Ideally, they want to take only one day to field a survey and get results.

For the company as a whole, there are considerations as to how to insure that the company standards and privacy guidelines are enforced. Breakdowns in securing customer data deteriorate consumer trust and can lead to violations of laws, resulting in fines and restrictive regulation. Thus, even while we give market researchers the ability to self-serve content, we are continually working to ensure that security and privacy have high visibility.

Currently, every study gets a policy review insuring compliance with guidelines on privacy, dealing with minors, protecting personal information, and talking about the brand. In terms of analytical review, we would like to see the QA level for the research projects reach ever higher levels of rigour. Survey technology can help by bringing work-flow related features into the application: e.g. a managed system for peer review.

### C. Reporting and Presentation of Analysis and Results

One of the issues that we had not anticipated was difficulty in providing a simple ad-hoc reporting solution. For us the most important task is quick and easy data exploration; a fancy system for sophisticated report sharing is ancillary at best. This is where the industry is still trailing, in my opinion.

A consistent theme in our story is that a close working partnership is key to the success of MR software in the large enterprise environment. In this case, we were able to address this shortcoming with our vendor and our needs were met.

Successes:

#### D. Vendor and User Cooperation

Beyond working out critical one-off issues there is still a need to incorporate Yahoo!'s requirements into software product development cycle. In some ways, the MR software industry, while improving, has yet to match innovations of the software industry as a whole, in leveraging their client base in order to improve the product. For example, when the product is modified in response to individual customer demands, the result can be a patchwork of awkwardly juxtaposed and overlapping features and functions.

Another model, employed by companies ranging from technology giants HP to blogging software developers Six Apart, is to unify clients into a community of users. I believe this multi-way dialogue will ultimately produce far better products.

#### E. Multi-Language and Multi-Modal

An absolute requirement for a 'state of the art' survey tool is multi-language support. Having that ability turns what was once a complicated and error prone process (duplicating static surveys) to something trivial.

A multi-language systems means separating the research elements, a question's semantic intent ("internal wording") and question type, from a presentation layer, such as its "external wording" (e.g. language) and visual layout. Automated error messages are stored by language of interview. Yahoo! researchers span the gamut of languages from English to variants of Chinese, and all of them are using the same interview tool. If a researcher wants to compare results from users across the globe, the researcher only needs one survey with multiple translations.

It seems quite obvious in hindsight, but this functionality has improved user experience tremendously on both sides of the survey process.

We are looking forward to this multi-language capacity evolving to its next logical step. Yahoo! is a leader in providing great web experiences beyond the computer browser on mobile devices. As the mobile web grows survey technology also needs to follow the consumers there, just as it followed them from the phones to the Web. Likewise our survey interview tools will need to be not just multi-language, but multi-modal as well.

#### F. Ad Targeting Platform for Research

To deploy surveys, researchers can use the same targeting sophistication that is offered to our advertising clients. For example, we make use of behavioral data to allow our advertisers to target those who fit a specified profile. Likewise, researchers can use that same capability to their advantage.

One way we employ behavioral profile targeting is to do stratified sampling for intercept targeting. For any given time period at most websites, more of the page views are coming from the heaviest users. A small percentage of users visits often and spend a disproportionately large amount of time on the site while a much larger percentage is light users. Because the distribution of page views is skewed towards heavy users, a random sample would skew the sample to heavy users. With the use of our profiling systems, we can determine if the person is a heavy or light visitor, and we can reserve some of the invitations for heavy users, to create a proper distribution of users in our samples.



## 7. Conclusions

What have we learned from this case study? I believe that Yahoo!'s decision and experience with bringing in a sophisticated self-service tool points to several things that are important to the MR industry.

In all fairness, the MR software industry is slowly adjusting to the research infrastructure needs of enterprise grade clients whose primary business is not MR, yet it's a two-way evolution:

- i. the industry is creating a more usable self-service technology experience
- ii. clients are recognizing the need for operations specialists; because researchers shouldn't be spending time learning overly technical tools and they need someone to champion their needs with software suppliers

Furthermore, while technology has to develop further in some areas, in others it has advanced even faster than our ability to fully adapt them.

Additional take-aways from our experience with MR technology adoption:

- Software development expertise is as important as knowledge of the MR business.
- MR technology products must continue to evolve in order to be viable in the market; software is not a static thing, and research companies that decide to create software must have a long-term commitment in order to succeed.
- As automation continues of things traditionally done by research operations specialists, such as survey programming and data processing, ease of use must be a priority. More work is needed by software designers in this area.
- Consideration must be given to general developments in computing and web technology. MR software should be based on platforms and applications that are popular in their target market. As much as possible, they should keep pace with their customers' natural hardware and software upgrade cycles.
- The decision to take a product in-house was a correct one. End-clients such as Yahoo! want and need direct contact with their respondents, especially in a web environment; having self-service tools provides this.
- Vendors of tools can benefit from including end-clients in their development cycle.
- There is still a role for consulting companies who can provide turn-key process for specialized data collection when they offer a value layer beyond just survey programming. (E.g., expertise in research methods: CSAT, ad-effectiveness, etc.)

## 8. Plan moving forward

To just make web surveying easier is not necessarily to make it better. I am not convinced that every time we do a survey, it is the best thing to do. By necessity, researchers tend to be focused on the shortest possible path between themselves and the insights that they can share. This affords little opportunity to concern themselves with the effects of individual projects on the quality of the overall research programme.

This version of the ‘tragedy of the commons’ afflicts all survey research ecosystems. However, Yahoo!’s centralized research operations group is in an unusually good position to do a better job of shepherding good behavior.

Workflow management tools can help keep us an eye on what’s going on in the system as a whole: to know what research is taking place, what is the cumulative impact on our customer base, and what is providing financial benefit to the company as a whole. We are developing appropriate dashboard management tools, but invite the MR software industry to take leadership in this area as well.

The other area of development is the use and growth of our proprietary panel. We are embarking on the best way to utilize this in conjunction with our other sampling methods.

## 9. The Last Word

In conclusion, while the project has been functioning for just two years, the overall benefit has been very positive. Researchers are getting cost-effective solutions, using in-house software and a centralized support team. Complex research that formerly was outsourced at great cost is done with lesser fixed budgets, while simple projects can still be completed relying on scalable resources. There is clearly an improvement in the operation of research and the quality of results is moving in tandem with the quantity.

## About the Author

Adi Cohen, Yahoo! Inc. Research Operations

## Appendix

**Figure: “supplier criteria”**

<b>Feature</b>	<b>Description</b>
<b>Network Support</b>	
Application Residence:	The physical location of a given application (e.g., on the client company's server or hosted on the vendor's servers).
Technology Architecture:	Operating system required to run the online research software (Windows or Unix).
Source Code Flexibility:	The ability to edit the survey software source code.
<b>Console Mgmt/Security</b>	
Password-Restricted:	Controlled access to online surveys via pre-selected usernames, passwords, and/or personal identification numbers (PINs).
Automatic Backups:	The scheduled storage of online research data for the purpose of protecting information.
Single Sign-On:	The ability to automatically sign in internal software users via a link to the internal network.
Project Accounting:	Tracking of internal efforts in managing online surveys (e.g., hours logged, surveys written, etc.).
Desktop Accessible:	The ability to log into the online research software from any computer on a network.
Set Expiration:	The ability to set a date and time for the survey to expire.
Multi-User Access:	The ability for two or more users to access the survey at one time.
Simultaneous Surveys:	The ability to administer two or more “live” surveys at one time.
<b>Data Importing/Exporting</b>	
Data Exports:	The primary exporting file types supported by the software system, including CSV, ASCII, SAS, and SPSS.
Script Importer:	Storage of past software scripts retrievable for future use.
Question Library:	Storage of questions from previous surveys retrievable for future use.
Template Library:	Storage of templates from previous surveys retrievable for future use.
MS Office Compatible:	The ability to export/import capabilities with Microsoft Office applications.
Real-Time Database Link:	The ability to link in real-time to a database in order to pull response-specific data.

Feature	Description
<b>Sample Mgmt.</b>	
E-mail Invitations:	The ability to transmit survey participation requests via e-mail to the survey sample.
E-mail Reminders:	The scheduled transmission of e-mails to respondents who have not yet completed a research survey.
Quota Management:	Tool to monitor survey responses for tracking pre-determined quotas.
Panel Management:	The ability to manage online research panels.
Save and Return:	The ability for survey participants to save at any point in a survey and return to complete at a later time.
Incentive Tracking:	The ability to track incentives promised to survey participants.
Online Respondent Help:	Assistance for survey participants taking the survey.
24/7 Help Desk:	Full-time live technical support for online research participants.
Multilingual Selection:	The ability for survey respondents to choose between different language options.
<b>Survey Design</b>	
Professional Services:	Option to request various survey creation services.
Multiple Choice:	Question set requiring respondents to select from one or more alternatives.
Multiple Selection:	Question set allowing respondents to select multiple answers.
Open-Ended Responses:	Question set allowing respondents to write in their own responses without cues.
Complex/Multiple Skips:	Adjustment to a line of questioning based on survey respondents' answers.
Question Randomization:	Automatic randomization of survey questions.
Text Editor:	Editing tool for ensuring proper grammar and/or spelling.
Navigation Control:	Setting respondents' ability to move forward and backward within a survey.
<b>Data Cleaning</b>	
Input Validation:	Automatic validation of survey response fields.
Duplicate Elimination:	Automatic elimination of duplicate entries.
Missing Data Identification:	Automatic identification of missing data in required fields.

Feature	Description
Nonsensical Data Removal:	Automatic elimination of data entries deemed nonsensical.
<b>Data Analysis</b>	
Basic Data Calculations:	The ability to compute standard measures, including averages, means, and measures.
Cross Tabulations:	The simultaneous tabulation of two or more variables.
Standard Statistical Tests:	The ability to perform basic statistical tests, including t-stats, z-tests, and f-tests.
Data Weighting:	The attachment of values to attributes based on their relative importance to the survey provider.
Hierarchical Data Sets:	Analysis of data through increasing levels of specificity.
Conjoint Analysis:	The ability to analyze the relative significance of respondents attached to particular attributes.
Regression Analysis:	Analysis between the relationship of a dependent variable and one or more independent variables.
Banner Format:	Data tabulation where responses from sub-groups are listed for each survey question.
Respondent Filters:	The ability to sample data by specified variables.
Data Sorting:	The ability to group data based on set requirements.
<b>Reporting</b>	
Multiple Report Formats:	The ability to present survey results in different formats.
Real-Time Reporting:	The ability to view in real-time a summary of survey results.
Multi-Channel Delivery:	The ability to send survey results via different channels (e.g., e-mail, PDA, Internet).
Define Recipients:	The ability to select recipients to receive research results once the survey has been completed.
<b>Multimedia Support</b>	
Picture Displays:	The ability to show picture files within online surveys.
Audio Playback:	The ability to play audio within online research surveys.
Video Playback:	The ability to show video files within online research surveys.
Macromedia Flash™:	The ability to play Macromedia Flash™ technology within online research surveys.
Logo/Branding:	The ability to incorporate logos and branding into surveys (ASP only).

Feature	Description
<b>Natural Language:</b>	
code open ends manually	support staff for coding open ends
Natural Language software	ability to bucket open ends based on pre designated terms or phrases



# Utopia - A Complete Research Management System

Tim Brandwood

## Abstract

This paper outlines a corporate-wide system that seamlessly and intelligently caters to all business functions and in which each component draws from a single centralised source. By detailing one aspect of it - the research management system - this paper hopes to demonstrate the level of sophistication and all-encompassing nature we are coming to expect from our software suppliers. The paper discusses the advantages of this system, the process we went through to get there and our frustrations with and experiences of commercial software systems. Finally, the paper argues that commercial market research systems need to be designed for greater extensibility and considers ways forward to achieve this.

## Keyword

Project Management, CATI Systems, Scheduling, Integration, Extensibility, Early Warning Systems, Intranet, Extranet, API

## 1. Introduction

Market research agencies conduct surveys. Many software vendors sell products that we use to conduct surveys. In reality there is more to it than this. Most research agencies do much more than just conduct surveys. We run projects, pay staff, anticipate problems, provide costings, manage workloads, communicate with clients, take holidays, produce accounts, communicate internally, raise invoices, write proposals and so on and so on.....

Of course, surveys are central to what we do, but there is much more to our day-to-day activities than simply gathering data and analysing it. In spite of this, most survey software focuses only on data collection, the analysis of data or a combination of the two.

Network Research has long recognised the need for a high quality system to manage and control the research process as well as support all auxiliary business activities. We are convinced that such a system needs to be centralised, work in real-time, eliminate duplication and galvanise communication.

Our efforts to date have created a sophisticated system that has radically improved the way projects are managed and information flow both externally and internally. This paper outlines the background to this process, the system we have developed and it's advantages and discusses the obstacles and opportunities for the industry as a whole.



## 2. Utopia – Specifying The Ideal

Our current system was born out of a systematic process of requirements gathering and software review. This exercise was triggered by frustrations surrounding fragmented software processes, fragmented and disparate data sources and a recognition that the market research software industry had moved on since we last reviewed our CATI software.

The review we undertook consisted of a comprehensive and thorough requirements gathering exercise. We examined the functional needs of every department in the company and effectively detailed a corporate-wide system to meet all the functional needs of every aspect of the business. Because surveys are at the heart of what we do, we went in search of a survey system that would sit at the heart of our Utopian system rather than simply adjunct to it.

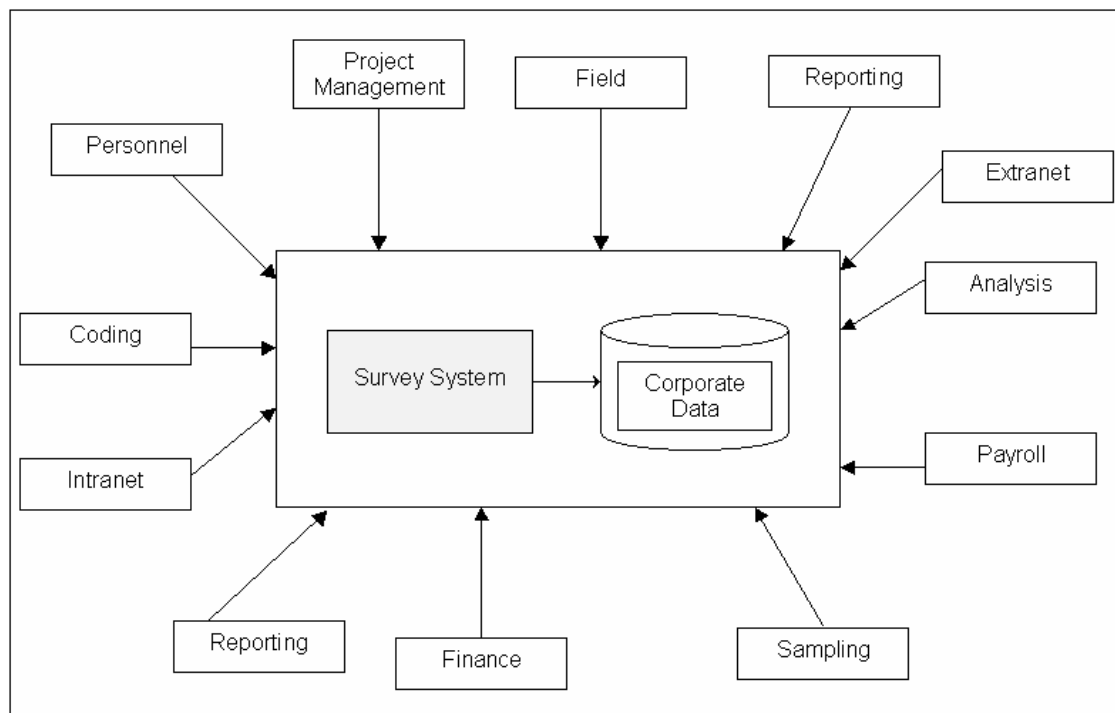


Fig 1. The survey system at the heart of the Utopia corporate system

Our requirements spec was sent to virtually every commercial supplier of market research software and we conducted a very thorough review of all the commercial tools and applications available. In an ideal world we were looking for an off-the-shelf system that would cater for all of the functional areas described above. However, we anticipated a compromise would be inevitable. In the end, we selected the system that would integrate most closely with our Utopian design whilst meeting as many of our functional requirements as possible out of the box. Under this inevitable compromise, we planned to use third party applications or our own bespoke development to fill the gaps as seamlessly as possible.

One of the key areas where all of the commercial products were found wanting, was the area of project management and project control. The efficient management and monitoring of projects is vital to all research agencies.. Similarly vital is the need for timely and relevant communication between staff associated with a project. In order to meet our own project

management needs, we undertook the in-house development of a research management system designed to work within our Utopian framework.

As explained earlier, our vision of Utopia involves more than just project management – it requires that we cater to all business functions. However, outlining the project management system we created serves as a good illustrative example of the kind of integrated system that Utopia implies.

The rest of this paper therefore outlines our project management system, the functionality it provides and the implications and lessons for the wider research community.

### 3. Project Management Under Utopia

The project management system within Utopia is comprised of three elements: an Intranet, an Extranet and a Desktop application called “Project Scheduler”.

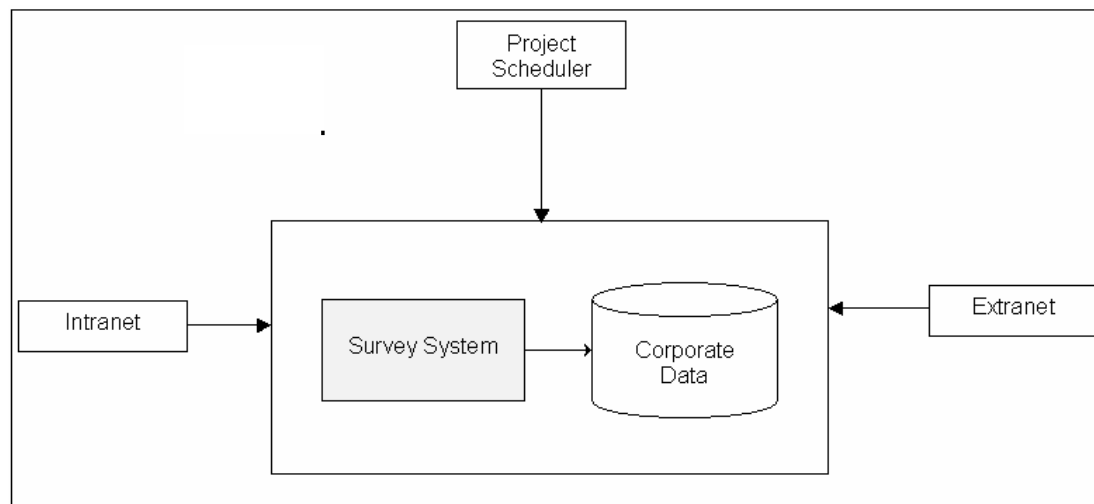


Fig 2. The constituent elements of the Utopian project management system.

The three components work from the same centralised (SQL Server) database. This means each element of the system has access to the same single source of real-time data with no duplication, fragmentation or overlap. The functionality of each component is as follows:

#### Project Scheduler

This is a desktop application available to all project staff. It is used to create, maintain and edit project schedules and timings. The system provides the input source for a definitive set of all project timings, schedules, waves and events. The schedule is detailed enough to cover all types of project timings from fieldwork periods to client meetings and from proposal writing to deliverable dates.

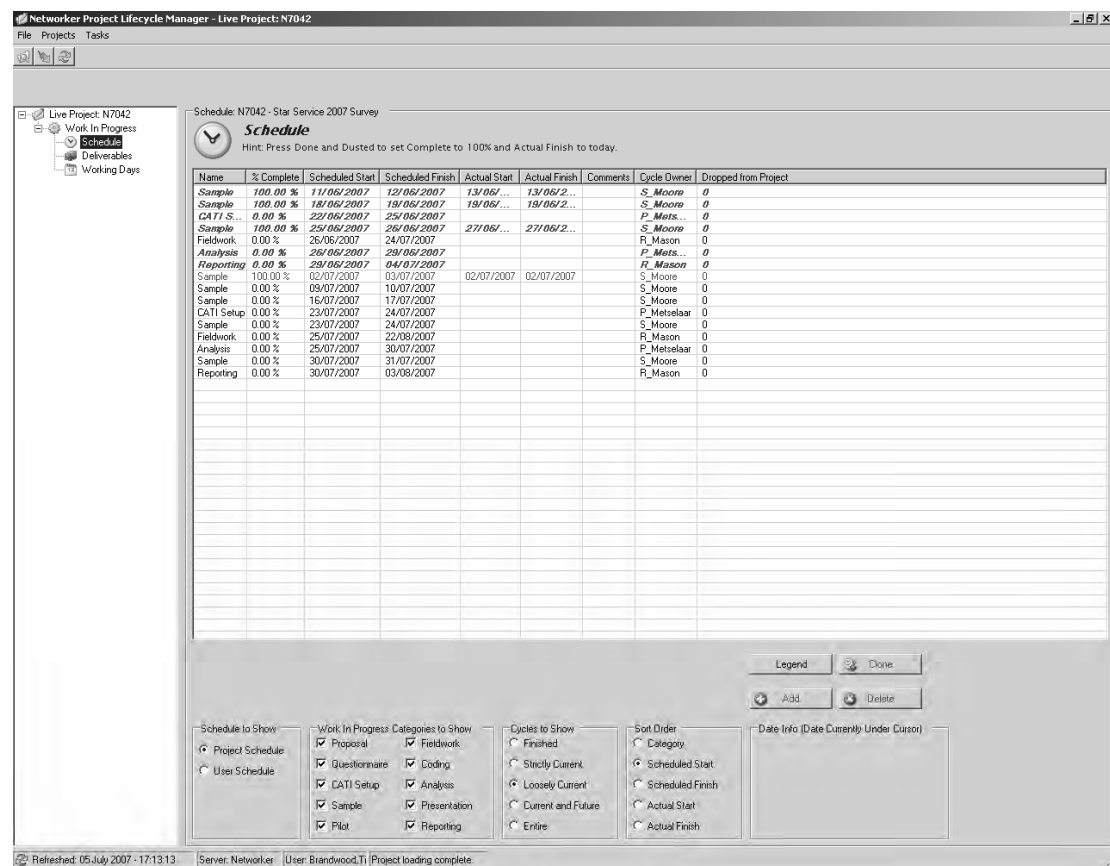


Fig 3. Project scheduling

Once a schedule is generated it is instantly visible to the company as a whole and the members of the project team are notified. There is only ever one current version of the schedule available on the system so staff accessing the schedule have the reassurance of knowing that they're always working from a definitive set of timings.

As time goes by, changes to the schedule are obviously inevitable. The system ensures that changes are made in a controlled manner. This means that only authorised staff can make changes and the system intelligently warns of bottlenecks or resource issues if the changes will cause a knock-on effect. The system also notifies relevant members of staff when changes are made. For example, if fieldwork dates are extended then coding is notified but the scriptwriter is not. An audit trail is kept of all schedule changes to assist with subsequent workflow analysis.

## Intranet

Our intranet provides a central point of access to company-wide information. As all of our data is housed in a single central source, the intranet provides users with the means to instantly reference any piece of information needed to perform their work. The intranet therefore allows staff to interact with the various parameters of a project such as project team, schedules, deliverables and documentation.

Viewing: Networker Home Page Generated: 05/07/2007 17:45:43 [F5 to refresh]

Status: All Projects Select By: All Projects Show: All Projects Project: N9000

---

### Project Basics

Project Number: N9000 Client: Prism DIY Limited

Project Name: Prism Visitor Satisfaction Contact: Steve Jones

Commissioned: 11/04/2007 Platform: Askia

---

### Project Team

Director: Richard Client Mgr: Tim

Project Mgr: Sandra Sample Mgr: Victoria

Field Mgr: Simon Analyst: Kirstin

Coding Mgr: Santosh

### Alert Summary

	-	Benchmark
Notifications	0.0	0.0
Tickets	0.0	0.0
Predictive Alerts	0	0.0
Avg Resolution Time	0.0	0.0

---

### Project Status

Project Status: Live

Currently In Field? No

Status Summary: Fieldwork finished. Awaiting analysis tables

### Field Status

Completes In Field	1408
Total Sample In Field	9597
Total Sample Used	3751
Total Sample Live	5846

---

### Costing Details

Sample Size: 1400

Project Value: 9000.00

---

### Work In Progress - Current & Recent

Project	Category	Status	Scheduled Start	Actual Start	Scheduled Finish	Actual Finish	Owner
N9000	Fieldwork	100%	13/02/2007	05/07/2007	02/03/2007	05/07/2007	RA

---

### Deliverables - Current & Recent

Project	Deliverable	Status	Due Date	Done Date
---------	-------------	--------	----------	-----------

© Copyright 2007

Fig 4. Editing intranet project parameters

The system holds the project schedule data but allows users to view this data in flexible and versatile ways. For example, it can display all the tasks assigned to a given individual, team or client. It can show all activities relating to a specific project or show all projects at once. It can filter by task type: e.g. “Show me what’s happening in field this month” or “What proposals are being written at the moment”

More importantly, the system also records and displays task completion status. So by viewing the schedule the user can see how near completion each task is. By comparing completion status against scheduled finish dates, the system can highlight things that are running behind or where deadlines have been missed. By colour coding the display the user can instantly see which tasks are areas for concern and which are running smoothly.

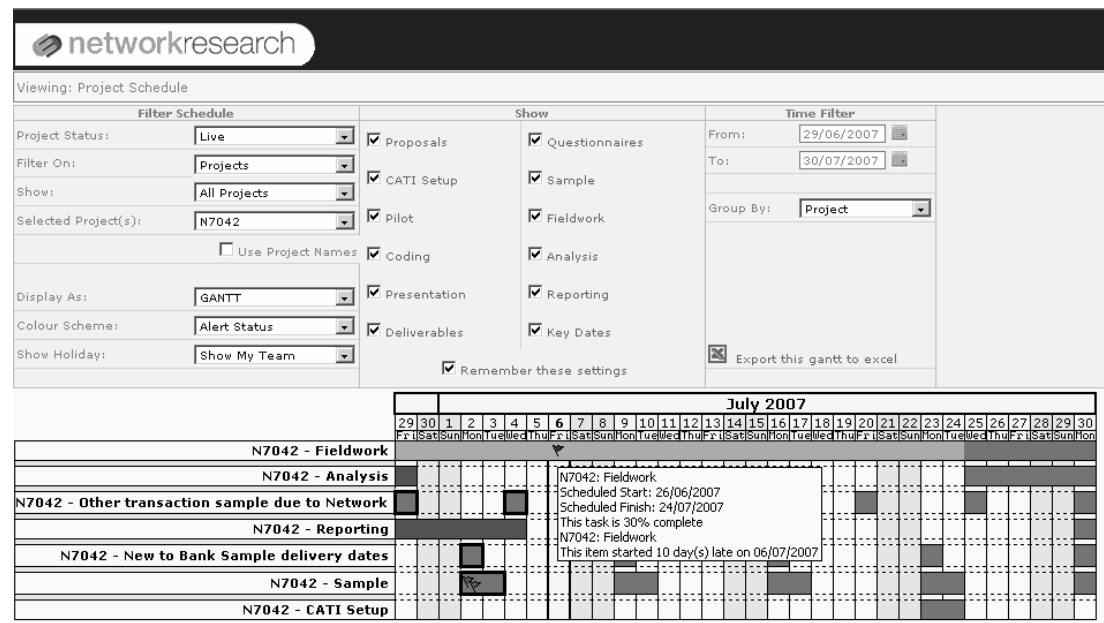


Fig 4. Viewing project schedules

In addition, the system is able to make predictions, based on current progress, about the likelihood of a task running late in the future. This early warning mechanism is invaluable in spotting bottlenecks and resource issues before they arise.

All of this has a significant effect on communication within the company. However, simply providing the information is not always enough. Users don't always have the time or inclination to check the system in search of problems. It is often more useful to "push" information them. Using a sophisticated rule based system, alerts can be issued to staff to warn them of events, problems or forthcoming tasks. Alerts may be issued on the basis of trivial conditions such as a pending holiday request or more complex conditions such as the likelihood of a task running late based on the performance of tasks higher up the chain. Whatever the complexity of the alert condition, the important point is that users are being served information rather than having to go in search of it. If no alerts are issued, the user can assume that there are no problems to deal with. In a sense, no news is good news.

Again, an audit trail is in place to monitor the alerts issued and track their frequency and average resolution time. This allows us to review not only how often issues arise, but how quickly on average they are resolved.

Viewing: Alert Browser Page Generated: 06/07/2007 15:19:54 [F5 to refresh]

Select By: **People** Show: **Paul Maskell**

Notifications		Tickets		Predictive Alerts	
Extreme	0	Extreme	3	Extreme	0
High	6	High	3	High	0
Medium	0	Medium	0	Medium	0
Low	1	Low	0	Low	0
OK	9	OK	0	OK	0
		Extranet Showstoppers	3		

Short Text	Action Text	Issued At	Showstopper?
Holiday Request Pending Giuseppe - Starting 24/12/2007	This request needs to be authorised by Paul M	19/06/2007 15:22:27	X
N4160 - Benchmarking Coding appears late	When Coding is complete, edit the schedule and set the actual finished date.	29/06/2007 00:00:06	X
N4196 - Downgrade Rese Analysis appears late	When Analysis is complete, edit the schedule and set the actual finished date.	03/07/2007 00:00:06	X
N4170 - OUBS Cert/Dip alu Reporting last day	View the project to check the latest completion status of the Reporting. When the phase is complete, edit the work in progress status for this task.	06/07/2007 00:00:06	X
N4201 - BUNPS Pilot Questionnaire last day	View the project to check the latest completion status of the Questionnaire. When the phase is complete, edit the work in progress status for this task.	06/07/2007 00:00:06	X
N7028 - Busine Coding last day	View the project to check the latest completion status of the Coding. When the phase is complete, edit the work in progress status for this task.	06/07/2007 00:00:06	X
N4066 - Call Centre Resea Coding starts Tuesday	This notification is a reminder only, no action is required.	06/07/2007 15:13:07	X
N4101 - Reporting starts Monday	This notification is a reminder only, no action is required.	06/07/2007 15:13:07	X
N4160 - Benchmarking Reporting starts Monday	This notification is a reminder only, no action is required.	06/07/2007 15:13:07	X
N4201 - BUNPS Pilot CATI Setup starts Monday	This notification is a reminder only, no action is required.	06/07/2007 15:13:07	X
N4201 - BUNPS Pilot Sample starts Monday	This notification is a reminder only, no action is required.	06/07/2007 15:13:07	X
N4226 - 2007 - Questionnaire starts Monday	This notification is a reminder only, no action is required.	06/07/2007 15:13:07	X
N7029 - Agricu Questionnaire starts Monday	This notification is a reminder only, no action is required.	06/07/2007 15:13:07	X
N7045 - M&L 2007 Fieldwork starts Today	This notification is a reminder only, no action is required.	06/07/2007 15:13:07	X

Fig 5. Alert notification on the intranet

In addition to schedule information, the system centrally holds an array of other relevant data. Information is held on costing parameters, staff holiday dates, invoicing schedules, admin forms, client contact details, currently logged in staff and so on. Keeping the all of this information in one place and integrating it with our survey system is essential if we are to intelligently link it together and provide systems that really support the day-to-day activities of users. For example, if a member of staff requests a holiday, that request is analysed by the system to ensure there are sufficient resources available to cover the absence. Any conflicts are then flagged to the line manager before approving the request.

All of these elements are essential in providing as system that intelligently links together all of our corporate data as well as survey data to provide users with a complete research management system.

### Extranet

Our client extranet allows clients to login and view, albeit with a slightly different presentation, the exact same data we see on our own intranet. Both the extranet and intranet draw their data from the same point so changes are reflected instantly in both places.

Much of the same functionality exists for clients as it does for internal staff. For example, the extranet shows project schedules, timings and events in real-time. It allows clients to see progress of tasks such as fieldwork and seeks to forewarn users about impending problems or issues.

Notifications are not “pushed” to clients in the same way they are pushed to internal staff. Users at the client end have varying and disparate roles in a project and it is impossible to automatically know which notifications are relevant to which individuals. Instead, users subscribe themselves to differing system notifications. For example, clients can request to be notified when files are published, fieldwork finishes, sample runs low or timings change. Ultimately this still results in a ‘push’ of information and the end benefit is the same as our own internal alerts mechanism, namely that the notification comes to them rather than them having to seek it out.

As another example of intelligently linked data, our extranet has a file upload facility that ensures files are always directed to the relevant member of staff. For example, if a client wants to upload a sample file, they only need to specify the project it relates to and the fact that it is a sample file. The system then checks who is responsible for sample and redirects the file to them. But what happens if that person is on holiday or off sick? The system checks for that as well and locates the best person to receive the file in their absence. This kind of functionality is especially useful when 3<sup>rd</sup> party suppliers need to send us files on the client’s behalf. Often the person sending the file has no idea who responsible for handling the file at our end. A slightly trivial example maybe, but it serves again to highlight the kind of spin-offs that are possible when you put all your data together in one place and link it up creatively.

#### 4. Advantages to Network Research and our Clients

From the preceding discussion, it should be fairly clear that we have developed quite a sophisticated and comprehensive project management tool. Most of the benefits of the system should also be fairly apparent, but before discussing the wider implications for the research industry as a whole, it is worth recapping on the main benefits for our clients and us.

##### Self-servicing of requests

Clients are often, like us, both project managers and researchers. Tools that allow clients to quickly monitor the status of projects are beneficial to them as well as to us. If clients can answer their own request by finding information on our extranet, then it’s a time saving for everyone.

##### In-built quality standards compliance

Although not detailed above, our system has also been written to support procedures and processes laid down by BS7911. As the system knows the status of each project, it also knows what documents and forms need to be provided according to these procedures. By verifying the existence of these documents the system is automatically capable of assuring compliance.

##### Paper Form Free

As our central objective was to keep all data together in a central store, it follows that we needed to phase out old style word and excel based forms. All forms can be considered as inputs into the larger system and should therefore be captured on our intranet. Whether the

user is booking a holiday, taking out a purchase order or requesting coding this is all relevant and must be linked into the wider management system.

#### Early warning system

As outlined above, the system is capable of predicting future problems, and providing early warnings to users. The survey system data is essential in achieving this as many early warnings are driven by data relating to fieldwork itself.

#### Complete transparency

Finally, and perhaps most significantly, our system displays our desire to be totally transparent to clients. If something is late then clients can see that. On the whole this instils trust and clients to see that projects are organised and under control.

### 5. Barriers and problems faced in building the system

When building the system and approaching the Utopia project in general we encountered a number of barriers and issues. Some of these relate to change management within our own organisation and some of these relate to limitations with the commercial products available. If other companies are to follow a similar path and attempt to implement a similar system, it is important that these barriers are anticipated and addressed.

#### Difficulty explaining our needs and aims to external software suppliers

During our many discussions with software suppliers, we found it surprisingly difficult to explain the importance and sophistication of our aims. Quite often software suppliers were very fixed on their “bit” without fully appreciating the wider business processes that lie beyond say, data collection or tabulation. Admittedly, survey software is a niche product of sorts. So, you might say ‘why would a CATI software supplier want to concern themselves with payroll or project management?’ To an extent this is a valid point, but this is ignoring the fact that like it or not, market research companies do have these external business processes to manage and place a high value on suppliers who can meet or help to meet these needs.

#### Difficulty in finding suitable products for integration

Given the previous point, it should not be a surprise to learn that very few products we saw offered the capacity for integration that we required. This was due to a variety of issues, such as legacy problems, perception of scope or simply a lack of resources. Whatever the reason, we found very few suppliers who could offer us a simple but comprehensive database schema in third normal form which we could read from (and sometimes write to) in any way we want, either now or in the future. Initially this may sound like an unreasonable demand, but it is ultimately fundamental. Without this kind of solid foundation none of Utopian ideas could ever be implemented. With such a database our system becomes infinitely expandable, malleable and flexible. As more and more agencies get to grips with their own implementations, it inevitably follows that the demand for this kind of open integration potential will increase. Suppliers with database formats that are closed, inconvenient, or even flat file based will find this comes as a disadvantage as customers demand greater levels of integration from their software.



### Changing working practices and gaining user acceptance

Any form of change within an organisation usually meets with some kind of internal resistance. People generally become accustomed to existing practices and need a certain amount of coercion to adopt new systems or working practices.

In the development of our systems we have generally found the following approaches aided user acceptance.

- Involvement of users throughout the development process. Early involvement of users helps generate initial buy-in. Continued involvement maintains that buy-in and helps keep the project user focused and on-track.
- Iterative development with frequent new releases. By frequently making improvements, you can maintain momentum and demonstrate to users that they are being listened to and the software is improving as a consequence. This again helps buy-in and can engender a creative virtuous circle of ideas leading to further ideas.
- Demonstrate genuine benefits to the user early on. If you can demonstrate to users that the new system or procedure genuinely makes their life easier in some way then it generally seems to gain user acceptance. Even if it's not an especially 'big bang' time saving, the point is that you're giving people reasons to like the new system rather than reasons to dislike it. This is particularly true if you can demonstrate some task, activity or output which physically would not have been possible under the old system. Obviously the earlier in the process you do this, the more beneficial it is.

## 6. Conclusions and Implications for the wider research community

Hopefully the preceding discussion has helped to illustrate the level of sophistication our agency is looking for from its systems. It should also have demonstrated that to some extent there exists a generic need within the research industry for such a system. Research companies place a high value on tools that can help with these seemingly peripheral activities.

Certainly, it can't be viable for all research agencies to implement a system similar to ours and it would be a duplication of effort if they did. Suppliers of survey software therefore need to recognise the requirement for them to either provide for these peripheral activities, or make it easier for third party suppliers to integrate with their systems and provide them instead.

For example, if a third party (like us) developed a generic project management tool, it is highly desirable for that tool to interoperate with all the incumbent tools in the marketplace. Or, as another example, why shouldn't our accounts package read directly from our CATI system to produce the payroll?

As mentioned earlier, greater openness and accessibility at the database level would open up many possibilities. However, as an industry it should be possible to go much further and provide some kind of standardised external interface that third party applications could plug into. As well as encouraging 3<sup>rd</sup> party development, this kind of standardisation and openness might even allow existing market research applications to talk to each other.

How might this be achieved? In an ideal world, all applications would be "built for integration" from the beginning. We could agree a standard database format that all research applications would adhere to. If all market research packages used an open normalised

database schema then it would allow for unprecedented levels of extensibility and integration. Clearly, and sadly, this is infeasible with so many legacy systems already in place.

So, instead a compromise might be to develop a standardised market research API (Application Programming Interface) which all survey software supports. Such an API might, for example, allow external applications to seamlessly request quota information, topline scores, hours worked by an interviewer, fieldwork dates and so on. It might even be able to manipulate sample, responses, interviewer records project information and so on.

This would of course be a major undertaking. For comparison, think how long it took the industry to adopt and agree the triple-s standard. Arguably, agreeing a common market research API is a task way beyond the scale of triple-s and therefore much more complicated and time consuming. However, providing this kind of external interface would, to the benefit of us all, allow tools like our project management system to be developed once, and then used by anyone within the industry.

## About the Author

Tim began working in Market Research in 1997 as an Executive for Millward Brown. After working up to SRE level and attaining both the MRS and CIM diplomas, he decided on a career change specialising in software development.

He has developed software for Marketing and Market Research since 2000. Tim has an MSc in Software Development and is Microsoft Certified. During his time at Network Research Tim has developed systems internally to improve information flow, streamline processes and manage projects. He is also heavily involved in the development of externally facing systems allowing clients to monitor projects and access results on-line.



# The Challenge of Geocoding Large-Scale Travel Surveys

Andrew J. Smith

## Abstract

Understanding travel behaviour is key to designing transport systems that meet both the current and future needs of the population. Travel surveys are a vital tool in providing estimates of transport demand and network usage. At their simplest level, transport surveys consist of recording the origins and destinations of travellers, sometimes disaggregated by travel mode and by time of day. More complex travel surveys consist of self completion travel diaries, on-vehicle interviews or focus group discussions where other variables may be collected such as interchanges, journey purposes, the frequency of the journey and trade-offs between cost and other factors.

Irrespective of the scale of the survey or the number of respondents, obtaining accurate location information on traveller origin, destination, and any interchanges is a key requirement of all travel surveys. This is particularly true when the survey data are to be used for transport planning purposes. Unfortunately, many surveys are carried out in less-than-ideal conditions (roadside, stations, trains, etc.) hence the data yielded by the survey can be inaccurate or incomplete. Therefore, a mechanism – **geocoding** – is required for post-survey “cleaning” of the location information to an appropriate level of accuracy.

MVA Consultancy has designed, built, and used a geocoding tool – **TARA Geocoding** – which has been applied in various forms on several very large-scale travel surveys including the London Area Travel Survey (LATS), the DfT’s National Rail Travel Survey, and the Countryside Agency’s England Day Visits Survey. It has also been employed on many smaller regional-level surveys. To date the TARA Geocoding software has cleaned more than 10 million survey addresses. It has been used both as a “back office” application and as Computer-Aided Telephone Interview (CATI) software.

TARA Geocoding utilises a three-stage process to “clean” the location information present in travel surveys, namely: **auto-cleaning**, **manual geocoding**, and **logic checks**. These processes are described in the paper together with the results from different surveys. These demonstrate how TARA has significantly improved the geocoding rate of survey data including the coding of ambiguous locations using TARA’s advanced “geosearch” algorithms.

The TARA Geocoding software, whilst very successful, is effectively a tool for fixing a data collection problem. Collection of precise address information at the roadside is a difficult task. The **challenge** is to get the data at the appropriate level of accuracy when it is collected whilst minimising the impact upon the time taken to complete the survey. Using our experience from previous surveys, we have developed a lightweight “portable” version of the TARA Geocoding application – **TARA Mobile**, allowing deployment on a handheld device. This allows the surveyor to collect accurate location data electronically and precisely at

source. This reduces the amount of post-survey processing of data required. We are currently trialling this software on a small travel-based survey and the results will be presented to the conference.

The paper will review the benefits and challenges of hand-held devices versus traditional survey techniques and discuss the implications for designing surveys where location data is sought.

## Keyword

Travel survey; surveying; geocoding; address cleaning

## 1. An introduction to travel surveys

Travel surveys are a standard technique used by Transport Planners and Market Researchers as a method for assessing why, when, where and how people travel. Travel surveys at their simplest level are simple roadside counts of the number of vehicles travelling along a stretch of road at a particular time of day. From this simple count, planners are able to make an assessment of what capacity of the highway is being used. Travel surveys can also be more complex. For example, a travel diary recording a person's day-to-day movements, journey purpose, and mode of transport can provide a wealth of detailed information about how people use the transport system.

Travel surveys are generally undertaken using a mixture of manual techniques and automated techniques. For example, traffic counts are, in 2007, frequently still undertaken by a surveyor standing on a stretch of highway with a "ticker" counter. This is a cheap, relatively reliable, and mobile solution. Where traffic counts are required on a regular basis (for example to phase traffic lights or provide congestion updates), assessments of traffic volume can be automated using under-road traffic counters or over-road cameras with Automated Number Plate Recognition (ANPR) systems. Camera-based solutions are expensive to set up and generally immobile, however once installed they provide around-the-clock reliable information.

Travel diaries are generally undertaken using a traditional paper-based approach. Respondents are given a paper diary and instructed to complete it for each trip that they make stating, for example, their origin, their destination, the time of day, their mode of transport, and their journey purpose (e.g. travel to work). Travel diaries are usually undertaken on a large scale – the complexities of the survey usually require a relatively large sample size in order to yield accurate results. Therefore, the paper diary provides a cheap solution and is usually self explanatory to complete.

Electronic diaries can also be employed for smaller focus groups where, for example, the purpose is to understand better the activity patterns of a household and the trade-offs between transport supply, allocation of time budgets and other factors including cost, convenience, work and family commitments, etc. While a useful tool in travel behaviour research there use in determining aggregate travel movements is of less value and so these tend not to be widely employed.

## 2. What is geocoding and why is geocoding important?

On a typical travel survey, the data collection exercise generates a vast quantity of data, usually in a paper-based format. For example, even simple roadside surveys may yield ten or more sets of counts (on paper). Household travel diary surveys may yield tens of thousands of paper diaries. Clearly, the data require some post-survey processing in order to turn them into a format that can be used by transport planners to make decisions.

The traditional approach to post-survey processing is as follows. First the data are keyed in to an electronic computer system. This exercise is often called “coding” and is usually outsourced to specialist data entry organisations. These organisations have the infrastructure in place to enable this work to be undertaken cost-effectively. However, not all data errors are picked up at this stage. For example, “High Street, Clandon, Surreey” may be corrected to “High Street, Clandon, Surrey”, but in reality “High Street, Clandon, Surrey” may not exist.

Correcting addresses - beyond simple typographic errors - is part of the geocoding process. Geocoding is “...the process of assigning geographic identifiers (e.g. codes or geographic coordinates expressed as latitude-longitude) to map features and other data records, such as street addresses.”<sup>1</sup> Geocoding is a vital process if survey results are to be used for transport planning purposes. The level of precision of geocoding depends upon the purpose to which the survey results are to be put. For example, if the purpose of the survey is to ascertain pedestrian routes from car parks to a shopping centre, a higher level of precision is needed than for a survey of the distance travelled by shoppers to the shopping centre. For the former, a precision to less than one metre level would be acceptable. For the latter, a precision accurate to around 10 metres would be acceptable.

Geocoding a relatively small number of addresses is a simple process. Many of us in the UK will have used, for example, the Royal Mail’s on-line service to confirm a domestic postcode. When we book theatre tickets over the telephone, we are often asked only for our house number and postcode. This is effectively small-scale geocoding. Using simple address lookup software it is possible to geocode hundreds of address locations within one or two hours. However, faced with 100,000 addresses to geocode, an automated (or semi-automated approach) is required.

## 3. Geocoding Large-Scale Travel Surveys

One of the largest travel surveys in the UK is the London Area Transport Survey (LATS) conducted approximately every ten years by Transport for London (TfL). The 2001 survey yielded almost 7 million addresses, all of which required geocoding so that the data could be used by local authorities, developers, transport researchers and other interested parties. Achieving a high-level of precision on all address locations was a key client requirement so as to maximise the potential uses for the data both inside and outside of the London area. TfL commissioned MVA Consultancy to provide a specialist geocoding solution for LATS. Geocoding 7 million addresses using a manual lookup process would take an unacceptably long time. Assuming each address takes one minute, this would equate to around 16,000 person days. To automate the process an integrated piece of software called “The Transport

<sup>1</sup> Wikipedia

Address **Resolving Application**” – or TARA – was developed to undertake the geocoding. TARA consists of three stages:

- An automatic **batch processing** of address lookup;
- A suite of semi-**manual geocoding tools**;
- A set of sophisticated **logic checks**.

The survey addresses were “denormalised”, or “flattened” into a standard format based around the UK Post Office Address File (PAF) – the *de facto* standard for UK addresses. The LATS data were then loaded into an Oracle database in the standardised format.

The three TARA process are described in more detail below.

### Automated Batch Processing

Once imported, the TARA system then passed the denormalised survey addresses automatically to a sophisticated address lookup routine fully integrated into TARA. The automated address lookup routine comprises a series of gazetteer lookups and address lookups. The address lookup routine used a product called QuickAddress Batch<sup>2</sup> which is widely used in address lookup products. The QuickAddress product provided a high resolution geocode (one metre accuracy) when a large number of key address fields were supplied (for example house number, street name, post town). However, a key remit of the project was to provide a geocode for *every* survey record at the *highest possible level of accuracy* but without making unnecessary assumptions about the source data.

QuickAddress could not provide a geocode for addresses where, for example, only the place name was supplied. For this reason, the consultants built and integrated gazetteers into the automated address lookup routine. The gazetteers consisted of sets of locations and an associated geocode, e.g. “West Clandon, 503070, 152895”. These gazetteers were produced using Geographic Information Systems (primarily MapInfo Professional and ESRI ArcGIS). Additionally, gazetteer data from other sources including airports, ports and businesses were included in the system.

Using the combination of QuickAddress and the gazetteer lookups, the automated batch processing routine was balanced to ensure that a large percentage of the 7 million addresses could be matched automatically, but the routine would not make unnecessary assumptions about locations. For this reason, the raw addresses needed to be of a reasonably high precision and accuracy to be matched by the automated batch process. For example, “8 Park Avenue, Barking, IG11 8QU” was matched by the system to “8 Park Avenue, Barking, Essex, IG11 8QU, Easting=544636, Northing=184663”. Very few addresses at a lower quality were auto-matched by the system, thereby minimising the chances of incorrect assumptions being made about the location.

From the TARA user’s perspective, the automated batch processing was triggered by a single button.

### Semi-Manual Geocoding Tools

The TARA automated batch processing routine successfully geocoded around 50% of the LATS survey addresses, thereby substantially reducing the task of manual geocoding.

<sup>2</sup> QAS Limited

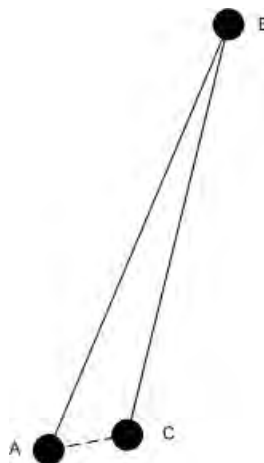
The next stage of processing consisted of **semi-manual geocoding** using a data entry form designed specifically for this system. The addresses which failed the automated batch processing were passed to a team of geocoders for semi-manual geocoding. The TARA system was installed on around 30 PCs at the consultant's office. TARA then displayed each of the "failed" addresses to the team. The geocoders were presented with the raw survey address and were provided with a suite of tools to enable them to look up the address. The tools consisted of QuickAddress (in "manual" mode), the gazetteers, and a map display (GIS). In the majority of cases (over 80%), the geocoders were able to geocode addresses successfully using these tools. In a minority of cases, the geocoder had the option of passing the address in question to a "senior geocoder" for resolution.

For example, geocoders used the train station gazetteer to geocode "Porchester road opp royal oak underground w2" to "Royal Oak Station, London, W2 6ET, Easting=525771, Northing=181580" and the map display to geocode "Porchester road near library w2" to "W2 6HU, Easting=525784, Northing=181402". In both cases, the geocoder managed to locate a suitable match using TARA and hence produce the desired level of output, i.e a 6-digit geo-referenced record.

The TARA system includes tools to manage the geocoding team such as monitoring geocoder performance (number of addresses geocoded per day), and allows a supervisor to select randomly a series of addresses to monitor the quality of the geocoders' work. Additionally, the semi-manual geocoding screen presented to the geocoders limited the time that a geocoder could spend on a single address. This ensured that predictions could be made to the client on when the job would be complete.

### Logic Checks

The final stage of TARA's geocoding process consisted of a series of **logic checks**. Many of the surveys undertaken as part of the LATS project consisted of multi-address responses. For example, travel diaries consist of a series of inter-related addresses such as "home", "work", "shops", "home". It would be possible, for each of the individual addresses to be geocoded correctly, but for the journey as a whole to be illogical. For example:



The diagram above illustrates a journey from A to B to C. In the absence of other information, the journey appears illogical as the distance from origin (A) to interchange (B) and interchange (B) to destination (C) is far greater than the overall distance from origin (A) to destination (C).



The TARA system incorporated an automated check – performed by the user clicking a single button - that measured the angle (at B). If the angle was less than a predetermined threshold, the journey was flagged as potentially illogical. The potentially illogical journeys were then passed to the geocoding team for inspection. Some of the journeys were not in fact illogical – a natural boundary such as a river between A and C, or the absence of a road between A and C may force the respondent to travel from A to B to C rather than directly from A to C. However, a large percentage of records may be incorrect due to data capture errors (by the respondent), data entry errors, or geocoder errors. The routine therefore prompted further investigation and data cleaning to ensure that the survey data were of the highest possible quality and fit for purpose.

TARA incorporates many other logic checks to improve data quality such as.:

- **Speed against mode of transport check.** For surveys where the mode of transport and journey times are recorded, this check ensures that the speed of travel is appropriate for the mode of transport. For, instance, an average speed of 40 mph would be inappropriate where the mode of transport was cycle, but would be appropriate where the mode of transport was train.
- **Distance to rail station check.** For surveys which included rail stations, a threshold was set for travel to a rail station. For example, a journey by car of greater than 50 miles would be inappropriate.
- **Trip reversal check.** Occasionally data entry errors resulted in multi-trip surveys (e.g. travel diaries) being input in the reverse order (i.e. destination then origin). A check was built into TARA where the respondent's origin address was closer to the destination station than the origin station, or where the respondent's destination address was closer to the origin station than the destination station. When both origin and destination are flagged, a trip reversal error is implied and the survey is sent to the senior geocoder for investigation.

The logic checks, and subsequent data investigations by the senior geocoders, are the final stage in the geocoding process before the cleaned survey records are exported and presented to the client.

The fact that TARA encapsulates all of the above functionality into a single, integrated database application that is easily installed has made it a highly attractive tool for other large-scale travel surveys. Beyond the LATS surveys, TARA has been adapted for many clients' surveys including those undertaken by the Countryside Agency, Department for Transport, and Local Authorities.

These other travel surveys have required modifications to TARA. The key change made for one survey was to use TARA for both data entry and geocoding in one step. That is, the TARA user interface was used for keying in questionnaire data. As the data were keyed in, the geocoding inherently occurred at the same time. The combination of data entry and geocoding speeded up the process of going from survey to geocoded result considerably. Whilst it is impossible to compare directly the data quality from one survey to another, it was felt that data quality was also improved as the data passed through fewer steps.

Additional enhancements made to TARA were made to the Geographic Information System (GIS). For certain surveys, additional boundaries have been added providing more detailed information about respondents' travel behaviour. For example TARA has been adapted to

display Areas Of Outstanding Natural Beauty and National Parks. Using the integrated GIS functions in TARA, it has been possible to ascertain very quickly respondents' travel behaviour and usage of National Parks. These levels of geocoding would be difficult, if not impossible, to undertake using simple address lookup algorithms.

#### 4. Beyond The Transport Address Resolving Application (TARA)

Over the last 3-4 years, the TARA system has proved to be very successful in producing high quality results for large-scale travel surveys. The automation of many tasks that hitherto have been performed manually has reduced the time and effort involved in post-processing survey data, and through the application of techniques like geocoding and geo-logic checks the accuracy and completeness of the data has much improved.

Even so, there are several enhancements that could be made to survey data collection to improve the quality of the geocoded results. Elements of TARA could be used to enhance this survey process.

Whilst the TARA system undoubtedly adds value to the data through the sophisticated GIS and post-processing routines, its basic purpose is in fixing poor quality data. Collection of precise address information at the roadside, for example, is a difficult task. Inclement weather, the dangers of undertaking roadside (or indeed household surveys) and data collector training all contribute to data quality. The challenge is to get the data at the appropriate level of accuracy when it is collected whilst minimising the impact upon the time taken to complete the survey and the safety of the participants.

To address this requirement a scaled-down version of TARA, called TARA Mobile, has been developed that can be used in a Computer Aided Personal Interview situation (CAPI), effectively replacing paper and pen. With this system, a questionnaire is built on a single desktop computer. The questionnaire designer sets which questions will elicit text-based responses (Yes/No, Comment, Name, *etc.*), which questions will elicit address-based responses, and which questions require the use of a GIS. The questionnaire is then "deployed" (sent to) one or more handheld PocketPCs or PDAs. The handheld devices then allow interviewers to collect questionnaire responses. The handheld(s) are then "docked" back at base, and the questionnaire responses are transferred to the desktop computer or server for further analysis (data can also be transferred via wireless connection where this is set-up). TARA Mobile can be run on a tablet PC or handheld PocketPC/Smartphone. In a personal interview situation, interviewees can then, for example, point to a map to show their location rather than having to describe their location in words.

A key challenge facing transport or land-use planners is assessing the level of use of large open areas such as National Parks or Areas Of Outstanding Natural Beauty where there are few addressable locations. A National Park, for example, may contain only a few locations recognised by Royal Mail (*e.g.* Public Houses, Farms, *etc.*). Assessing usage of a National Park therefore requires a different survey technique. It would be possible to monitor through a traditional travel survey (personal interview or travel diary) which car park the respondent used or where they left the road. However, describing their journey *through* the National Park would be impossible using addressable locations.

A possible solution would be to use Global Positioning System (GPS) devices which are carried by respondents. GPS devices are relatively inexpensive (a consumer device typically

costing between £50 and £100). These GPS devices can store a date-stamped “breadcrumb” trail (geocodes) of the respondent’s location to an internal memory card. When back at base, the data can be transferred from the GPS device to a central database (TARA, for example). The “breadcrumb trail” can then be electronically re-created in a GIS system (or a modified version of TARA) and the “route” plotted for further analysis. A further benefit of using GPS units is that they could be used in conjunction with personal travel diaries. If respondents were asked to keep travel diaries, a GPS device could be used as a “calibration” of respondents’ travel diaries. In other words, the travel diary could be “corrected” based on the results of, for example, one day’s use of the GPS unit.

As mobile devices such as PocketPCs, Smartphones, tablet PCs, and GPS devices become cheaper, more “ruggedized” and easier to program, it is highly likely that they will become increasingly prevalent in travel surveys. Whilst there is a larger up-front cost to the survey in procuring and configuring these devices when compared to a traditional paper-based survey, the benefit is a reduced downstream cost in post-processing data. Additionally, and perhaps more crucially, these devices do not simply automate an existing survey process. Rather, they enable the collection of a far greater range and depth of information about people’s travel behaviour that is not possible with traditional survey techniques.

## About the Author

Andrew Smith is a Managing Consultant at MVA Consultancy. MVA provides advice on transport and other policy areas to central, regional and local government, agencies, developers, operators and financiers. Andrew has a degree in IT and is a PRINCE2 qualified project manager and Chartered IT Professional. Throughout his career, he has worked on many software development projects as a programmer, systems analyst, consultant, or project manager. Andrew has extensive practical experience of a wide range of technologies, including client/server and web-based development tools, relational databases, address processing, and Geographic Information Systems. Andrew is a technical expert in the Oracle relational database management system, applied in particular to solving transport-based problems and has written and delivered training courses in this field.

Andrew has developed MVA Consultancy’s address geocoding business, having been responsible for the delivery of several large systems for clients including Transport for London (TfL), the Department for Transport, and the Countryside Agency. He project-managed the delivery of a system to provide statistics on the effect of the Congesting Charging scheme, and also the migration of a transport network modelling system to an Oracle database for a rail authority in Hong Kong.

Andrew has developed several systems for public authorities including London Buses, West Yorkshire PTE, and Hertfordshire County Council. He has also worked in the area of finance, developing and delivering a treasury management system.

Andrew can be contacted at: MVA Consultancy, Dukes Court, Duke Street, Woking, Surrey, GU21 5BH; Tel. 01483 742 870; e-mail [asmith@mvaconsultancy.com](mailto:asmith@mvaconsultancy.com).

# The Use of Seasonal Adjustment Software within the Office for National Statistics

Fida Hussain, Craig H. McLaren, Nigel Stuttard

## Abstract

The Office for National Statistics (ONS) is responsible for the derivation and production of a large number of time series estimates. The office has recently developed a software system which enables time series functions to be applied consistently across a wide range of time series. The system uses the X-12-ARIMA software package to carry out seasonal adjustment and forecasting. A number of other time series functions, including interpolation, are carried out by the system using SAS®. This paper gives an overview of the software system and highlights the benefits of an integrated approach that allows time series estimates to be derived in a consistent, timely and transparent way. We also consider future directions for the analysis of time series at the ONS.

## Keywords

Seasonal adjustment; forecasting; interpolation

## 1. Introduction

The Office for National Statistics (ONS) produces a wide range of statistics which are used by policy makers, institutions, academics and the public. These outputs include estimates of Gross Domestic Product, unemployment and price and earnings indices. Published data are usually provided in two forms: non-seasonally adjusted estimates and seasonally adjusted estimates. It is important that the seasonal adjustment process is well understood and that published seasonally adjusted estimates are of high quality.

The ONS has previously announced plans to streamline and modernise software systems for the production of official statistics. The modernisation will improve the quality of published seasonally adjusted outputs by building a common platform that supports state of the art time series packages. Currently, a variety of different time series packages and methodologies are used for carrying out the same tasks across the ONS. The approach to modernisation of statistical systems within the ONS has focused on a number of re-engineering projects. This paper looks at the development of the new software system focusing on the time series functions for seasonal adjustment, forecasting and interpolation used within the National Accounts system.

## 2. Current time series system for National Accounts

National Accounts is the central framework for the presentation and measurement of the stocks and flows within the economy. Estimates are published in key publications: the United Kingdom Economic Accounts (published quarterly) and the annual United Kingdom National Accounts (the Blue Book). In addition, monthly estimates of output for production and services are also published. Seasonal adjustment is performed on a regular basis each month or quarter. A review of the seasonal adjustment parameter options is performed on an annual basis to ensure the quality of all time series. A centralised database system is currently used to coordinate and manage seasonal adjustment within National Accounts. This system was developed in 1991 and performs a number of functions, including:

1. Serving as a time series database which contains all National Accounts data;
2. Providing the functionality necessary to produce National Accounts, for example aggregation, benchmarking, chain-linking and balancing;
3. Providing a variety of specialist time series tools, including seasonal adjustment (X-11-ARIMA, Dagum 1988), forecasting based on the Holt-Winters modelling approach (Winters, 1960) and interpolation based on an internally developed splining approach (Baxter, 1991).

The broad steps in the current process for reviewing and implementing seasonal adjustment of National Accounts are:

1. Time Series Analyst (TSA) specialists meet with the client to discuss dataset and any potential issues related to seasonal adjustment.
2. TSA, download time series and associated prior adjustments from the central database.
3. TSA seasonally adjust individual time series using X-12-ARIMA to obtain optimal parameter settings and provide the settings to the client.
4. Client area manually sets up analogous X-11-ARIMA settings in central database.
5. Client area performs a production run using concurrent seasonal adjustment.

A similar process is used for other production areas in the ONS. Generally, X-11-ARIMA is used to produce the seasonally adjusted estimates, although the parameter settings used in the seasonal adjustment process are derived by the Time Series Analysis area using the more advanced seasonal adjustment package X-12-ARIMA (Findley, et al. 1998). These parameters include the time series decomposition model, ARIMA model, and prior adjustments to account for any possible outliers and structural breaks. One limitation of the current production framework is that the production area have to manually apply these settings to X-11-ARIMA. This process can be streamlined by the consistent use of a single seasonal adjustment package. This will reduce the risk of human error as there will be less manual manipulation, for example, in uploading prior adjustment factors for individual time series.

The current database system is due to be decommissioned in 2009. A modernised replacement database and system has now been developed to ensure an internally consistent National Accounts dataset from 1997 onwards. Pre-1997 data will be available at the published level.

### 3. Methodological change: improving time series functions

Improved methodology for time series issues are now available for implementation into ONS production systems. The modernisation program will enable sophisticated methodologies to be implemented so that the quality of the published statistical output is improved.

#### 3.1 Improving methodology for seasonal adjustment

X-12-ARIMA is a leading seasonal adjustment package commonly used by official government agencies for the purposes of seasonal adjustment. For example, X-12-ARIMA is used within the ONS, Bank of England, Statistics Canada, Australian Bureau of Statistics and the United States Census Bureau. Both X-12-ARIMA and TRAMO-SEATS (Gomez and Maravall, 1997) have been endorsed by Eurostat as acceptable seasonal adjustment approaches for official statistics within the European Union.

X-12-ARIMA has a number of features that are improvements over X-11-ARIMA the current production used in the ONS. They include:

1. The use of user-defined regression matrices which allow complicated effects to be estimated and removed. This aids in the identification and estimation of outliers, structural breaks and calendar effects as well as improving the estimation of seasonality.
2. Various methodological improvements, including automatic decomposition and model selection methods, improved statistical diagnostics and more flexible options to determine quality and stability of seasonally adjusted estimates.
3. New functionality including a more flexible benchmarking tool, developed by Statistics Canada (Quenneville, et al. 2004).

See Findley et al., (1998) for more details.

The consistent and integrated use of X-12-ARIMA for all aspects of seasonal adjustment will provide a coherent interface between the Time Series Analysis area that conducts the annual seasonal adjustment review and the client who is responsible for production. Each time series will continue to be reviewed by the Time Series Analysis area annually, and the X-12-ARIMA specification file updated to take into account any changes. Production areas will then use these specification files to seasonally adjust series, individually or in groups, during the production process. Seasonally adjusted aggregates using component time series will be calculated in a separate framework. On occasion the client area may require time series, which do not have a specification file, to be seasonally adjusted. In these circumstances a default specification file will be used. The Time Series Analysis area will retain overall responsibility for the quality assurance of all published seasonally adjusted estimates. Seasonal adjustment reviews will continue to be held on an annual basis, with emerging issues addressed on an as needed basis.

#### 3.2 Improving methodology for forecasts and backcasts

Forecast and backcast estimates are used for a variety of purposes including imputation of missing values, benchmarking, and interpolation (see Section 3.3). In National Accounts, forecasts are needed to impute missing values particularly for the preliminary Gross Domestic Product estimates. In the current time series system, the Holt-Winters methods (Winters,

1960) are usually used to forecast most National Accounts series. Some restricted ARIMA (Autoregressive Integrated Moving Average) models are also available.

The Holt-Winters approach is a specialised case of the more general ARIMA model framework. Cryer and Miller (1991) and Taylor (2003) discuss how the Holt method can be replicated by Box Jenkins ARIMA models. For example, the double exponential smoothing Holt model is equivalent to the ARIMA (0,2,2) model which is given by:

$$y_t = 2y_{t-1} - y_{t-2} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

where  $y_t$  is an observation at time  $t$ ,  $\theta_i$  is a parameter estimated from the time series, and  $\varepsilon_t$  is a white noise process with mean 0 and variance  $\sigma_\varepsilon^2$ . See Sclove (2002) for more details.

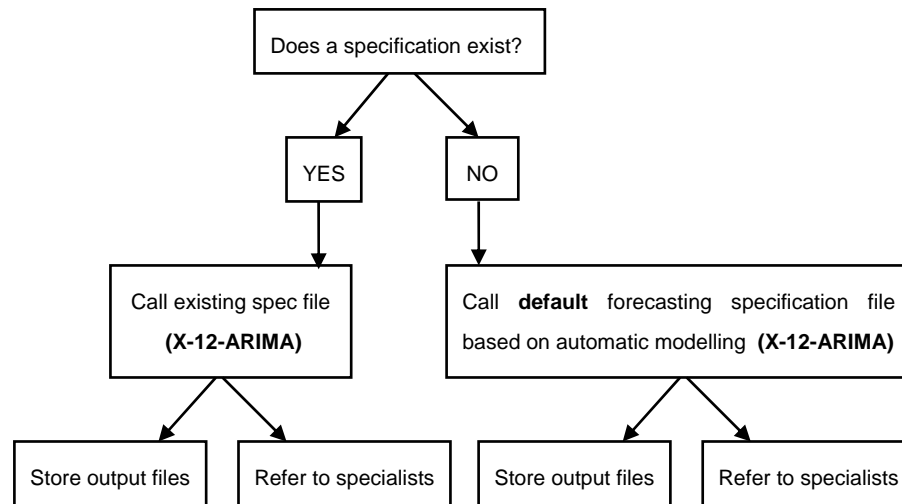
X-12-ARIMA will not generate exactly the same estimates for the parameters as the current central database because the method of parameter estimation is different. However these parameters within X-12-ARIMA, unlike the parameters used within the current system, are statistically optimal having been derived using maximum likelihood estimation as opposed to a generic grid-search technique using initial start up values used in the current system. Simple extrapolation techniques are available to be used for short times series of less than three years in length.

There is potential for significant improvements in using X-12-ARIMA for all forecasting and backcasting. For example:

1. Tailored forecasting models can be generated for important series allowing the flexibility to obtain an adequate ARIMA model for forecasting purposes only.
2. ARIMA modelling within an X-12-ARIMA framework provides important model diagnostics: whether the model is stationary and invertible, standard errors, confidence intervals and autocorrelation measures. These statistics can aid the user in assessing whether the model chosen is adequate and gauge the uncertainty of a chosen ARIMA model. These diagnostics cannot be generated for the existing Holt-Winters approach within the current forecasting system.
3. X-12-ARIMA's ARIMA functionality enables greater flexibility. In addition to a variety of ARIMA models, the current Holt method and Holt-Winters method (for additive series) can be replicated.
4. ARIMA models are used in the seasonal adjustment process, hence there will be a consistent approach in the use of models.

Figure 1 outlines the proposed approach for forecasting. X-12-ARIMA will automatically identify an ARIMA model that best fits the existing data and then will use the model to forecast future data. The initial step in the proposed system will be to assess whether the time series has an X-12-ARIMA specification file. If so, the file will be used and the pre-selected ARIMA model will be used as the forecasting model. This will ensure consistency of model choice. Where no specification file exists, the means of choosing a model will depend on the periodicity of the series. Annual time series, for which no seasonal adjustment specification file will exist, will be forecast using the ARIMA (0,2,2) model which is equivalent to the non-seasonal Holt model. The quarterly or monthly time series will be forecast using a default forecasting specification file which will call the adapted TRAMO (Time series Regression with ARIMA noise, Missing values and Outliers) forecasting routine of X-12-ARIMA and will automatically construct an appropriate ARIMA model.

Backcasts of time series are required as part of the interpolation and benchmarking processes. The model used for backcasting a series is the same as that used for forecasting.



**Figure 1. Proposed forecasting approach**

### 3.3 Improving methodology for interpolation

Interpolation is a technique used to produce estimates for indicators that are needed at a higher frequency than are available from the source data, for example, when quarterly or monthly data are required and only annual data are available. The situations in which interpolation is used fall into two separate classes: a) temporal disaggregation where a total or average covering the whole time period needs to be distributed over a shorter time period or b) where a point in time estimate is known for a particular instant in the time period and estimates are required for other instances during that time period. Temporal disaggregation is typically used for flow series whereas point in time interpolations are required for stock series.

In general, the process of interpolation involves initially forecasting and backcasting the low frequency series and then interpolating by fitting a set of continuous curves to the expanded series. When the input data are the total or average for a period and we are interested in temporal disaggregation, then the areas under the curve over the input intervals are constrained to equal the given total. When the input data relate to a point in time, the curve is constrained to pass through the given data points. High frequency time series are then obtained with the additional step of truncating the series to the length of the original series.

The current approach is to use a custom ONS developed package for temporal disaggregation and point in time data (Baxter, 1991). The proposed approach will use an in-built function available within SAS®. version 9.1.3 for both situations. The use of the SAS®. package has a number of advantages including: ability to deal with both temporal disaggregation as well as interpolation in a single function, being methodologically flexible with a range of sophisticated options, complying with the modular design aspects of the modernised system and ensuring production of high quality outputs.



## 4. Results: Empirical analysis

### 4.1 Evaluating a default X-11 specification file using X-12-ARIMA

Default settings for X-12-ARIMA specification files were derived from past experience of seasonal adjustment. The Time Series Analysis area has gained expertise from undertaking the seasonal adjustment re-analysis of predominantly ONS data series. This has involved building upon past experience with X-11-ARIMA, and by working with beta versions of X-12-ARIMA supplied by the United States Census Bureau. For example, seasonal adjustment settings have been refined for ONS time series by extending the default list of ARIMA models.

We compared three different versions of X-12-ARIMA: version 0.3 Build 135, version 0.3 Build 169 and X-13A-S version 0.3 Build 128. The most recent release of X-12-ARIMA is available for download from: <http://www.census.gov>. For X-13A-S, the SEATS component was not used so as to analyse the model based X-11 functionality. Analyses of the final values for the irregular component from the X-11 process showed no difference between the three software versions indicating that, under the chosen default specification and the use of the X-11 process, that all three versions produce identical results. Table 1, shows the range of ARIMA models selected using the available default automatic modelling (AUTOMDL) routine within X-12-ARIMA. A variety of time series of differing lengths were used. This shows that the commonly used airline model is a reasonable fit to a significant proportion of ONS monthly and quarterly time series data.

Monthly		Quarterly	
Chosen model	Percentage	Chosen model	Percentage
(011)(011) <sub>12</sub>	50.4	(011)(011) <sub>4</sub>	38.7
(100)(011) <sub>12</sub>	11.3	(100)(011) <sub>4</sub>	9.7
(010)(011) <sub>12</sub>	5.0	(010)(011) <sub>4</sub>	9.7
(110)(011) <sub>12</sub>	3.2	(110)(011) <sub>4</sub>	6.5
(210)(011) <sub>12</sub>	2.7	(210)(011) <sub>4</sub>	3.2
Other models	27.5	Other models	32.2

Table 1: Chosen models for seasonal adjustment runs using automatic modelling in X-12-ARIMA

### 4.2 Assessing Forecasting and Backcasting

Our primary objective was to ensure that appropriate forecasts could be generated by using the available automatic modelling (AUTOMDL) routine within X-12-ARIMA. The automatic model routine is closely based on the TRAMO methodology of Gomez and Maravall (1997).

Table 2 gives the results of the application of the default specification file and failure rates of the automatic modelling routine against selected ONS time series. These results show that, for the monthly and quarterly series, AUTOMDL builds a suitable ARIMA model using an automatic process, and successfully fits a model in all but a small proportion of cases. However, it fails to produce an output in over one fifth of cases when dealing with annual

series. This is because of the difficulty of fitting a model with a limited number of observations. An alternative procedure available within X-12-ARIMA is the use of pick model (PICKMDL) which selects an ARIMA model based on a pre-defined set of ARIMA models. Table 3 shows that the AUTOMDL procedure is preferred when compared to the PICKMDL procedure. The comparison is based on an adequacy criteria which assesses the stationarity and invertibility of the ARIMA model and that there is no serial correlation in the irregular component. The comparison is a summary over 200 time series where each individual series has been fitted using the alternative procedures.

As a result of our trials, we recommend the use of AUTOMDL for model fitting for monthly and quarterly series and the use of a fixed model, ARIMA (0,2,2) for annual series.

Njimi and Melard (2007) conducted a study analysing two automatic forecasting packages which are based on TRAMO methodology, TRAMO-SEATS and TSE-AX. The TRAMO methodology is equivalent to the AUTOMDL procedure available within X-12-ARIMA: They looked at nearly 3000 (annual, quarterly and monthly) series and make the point that even though TRAMO-SEATS was not designed primarily as a forecasting software package, it appears as a "very satisfactory forecasting solution". Maravall (Pena et al. 2000) applied TRAMO on 35 series, including 22 simulated series, and found that TRAMO identified "either the same model as the one identified by the time series expert or an also acceptable, sometimes better, model".

Periodicity	Time series used	AUTOMDL results	Recommendation
Monthly	265	3 failed to produce output	Use AUTOMDL
Quarterly	175	6 failed	Use AUTOMDL
Annual	370	80 failed	Use ARIMA (0,2,2)

Table 2: Default specification tests using National Accounts data

		Monthly: AUTOMDL		Quarterly: AUTOMDL	
		Adequate	Not adequate	Adequate	Not adequate
PICKMDL	Adequate	59%	4%	84%	9%
	Not adequate	14%	23%	2%	5%

Table 3: Comparison between AUTOMDL and PICKMDL within X-12-ARIMA for selected time series in terms of adequacy of model fit.

### 4.3 Assessing the seasonal adjustment process

Having established that the differing versions of X-12-ARIMA and the model-based X-11 part of X-13A-S, as outlined in Section 4.1, gave similar results we investigate whether the SEATS seasonal adjustment is at least as good as the model-based X-11 adjustment. Analysis is conducted on default specifications for both X-11 and SEATS so the quality of seasonal adjustment is not assessed on an individual series by series basis. This is on going research.

Revision measures are useful in comparing different seasonal adjustment methods. The basic revision is the difference between the initial seasonal adjustment (concurrent adjustment) and the seasonal adjustment with all the data available at the time of the analysis (i.e. the final adjustment). The percent revision for the seasonal adjustment at time  $t$  is defined as  $R_t = (A_{t|T} - A_{t|t}) / A_{t|t}$  where  $A_{t|t}$  is the seasonal adjustment at time  $t$  for data up to time  $t$  and  $A_{t|T}$  is the seasonal adjustment at time  $t$  for the data up to the end of the series at time  $T$ . Similar revisions can be generated for the month-to-month changes in the seasonally adjusted data, and the trend estimates.

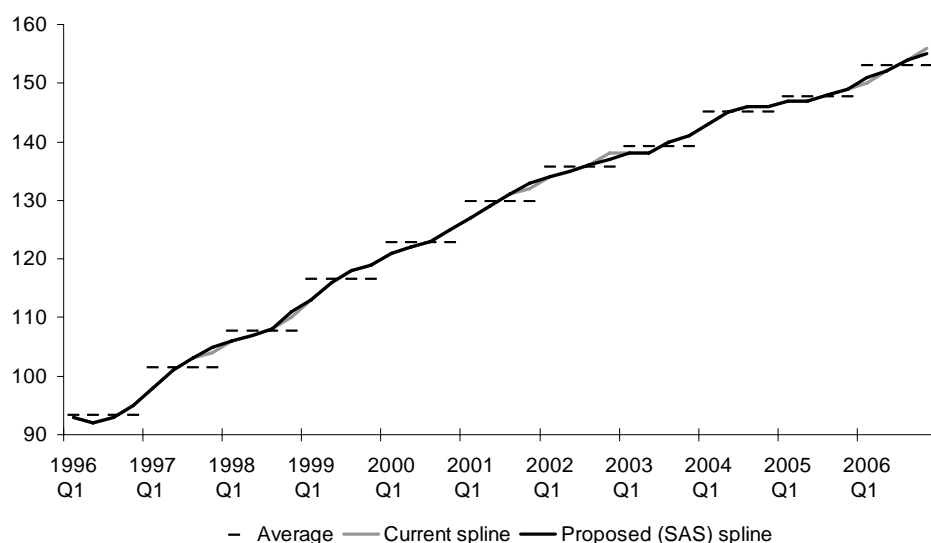
Table 4 highlights the results for a dataset of the Retail Sales Inquiry. The first two diagnostics are the percentage of months flagged for unstable seasonal factors (S %), and month-to-month changes in the seasonally adjusted estimates (M %). The other measures are revisions based and are the average absolute revisions from the initial to the final estimate of the seasonally adjusted data, change in seasonally adjusted data and trend. In terms of stability of seasonal adjustment, SEATS seasonal adjustment estimates are generally smoother than those for X11 estimates. Both SEATS and X11 adjustments are similar with respect to revisions. SEATS adjustments are preferred in terms of providing a more stable seasonal adjustment in terms of sliding span and revisions analysis. Monsell et al. (2003) assessed a single time series, the Retail Sales of Shoe Stores, using similar criteria and found that the SEATS adjustments were 'slightly' more stable than those for the default X-11 adjustment, in terms of revisions and sliding spans diagnostics.

Dataset: MONTHLY start date Jan 1986 (22 Retail Sales Inquiry series)		
Diagnostic	X-11	SEATS
Percentage of unstable seasonal factors	1.81 (2.60)	0.54 (1.02)
Percentage of unstable month-to-month changes in seasonally adjusted estimates	4.55 (4.29)	1.40 (2.56)
Average Absolute Revision: seasonally adjusted	0.45 (0.23)	0.52 (0.28)
Average Absolute Revision: change in seasonally adjusted	0.41 (0.22)	0.34 (0.21)
Average Absolute Revision: trend	0.89 (0.36)	0.89 (0.41)

Table 4: Seasonal adjustment diagnostics for X-11 and SEATS. standard deviation in brackets.

#### 4.4 Assessing Interpolation

Approximately fifty National Accounts series were investigated and, on average, the new proposed method gave improved results, i.e. a smoother curve. Figure 2 shows a comparison of the current and proposed methods for an individual series (NHS Pharmaceutical Services). The dashed horizontal lines show the annual figure averaged over four quarters. The differences between the two approaches are small with an average discrepancy of 0.08%.



**Figure 2: Comparison of Splining techniques (backcast period)**

## 5. Discussion

The proposed modernised infrastructure for time series analysis will provide a consistent and coherent approach for the production of time series estimates. The quality of the final published seasonally adjusted outputs will be improved with the use of improved methodological approaches such as the consistent use of X-12-ARIMA for seasonal adjustment, forecasting and backcasting and the use of SAS for interpolation. The proposed system is planned to be implemented and operational by September 2008. The use of the proposed system will provide a number of challenges. For example:

1. The assessment and implementation of the most appropriate level for seasonal adjustment within National Accounts bearing in mind the need for high quality and the operational needs of National Accounts.
2. Use of default forecasting will need to be closely monitored within the production environment. An inadequate default specification may result in failure to generate forecasts and backcasts.
3. Provision of practical guidelines and training to ensure client areas have an appropriate level of knowledge, understanding and skills to address any time series issues as they occur, particularly in the initial use of the proposed system.

Further developments are planned to continually improve and enhance the ONS approach to seasonal adjustment. There has been a recent detailed review of the seasonal adjustment review process. This has identified areas to focus on which include: development of a dedicated seasonal adjustment analysis system to assist with the management of consistent application of seasonal adjustment parameters, and future consideration and evaluation of the benefits of the X-13A-S approach.

## References

- Baxter, M.A. (1991) Reduction of Annual Data to Monthly or Quarterly Data. ONS
- Box, G.E.P. and Jenkins, G.M. (1970) Time Series Analysis, Forecasting and Control. San Francisco: Holden-Day (rev. edn publ. 1976).
- Cryer, J.D., and Miller, R.B. (1991) Statistics for Business: Data Analysis and Modelling. PWS-Kent (Wadsworth), Boston.
- Dagum, E.B. (1980) The X11ARIMA Seasonal Adjustment Method. Ottawa: Statistics Canada, cat. 12-564E.
- Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C. and Chen B. (1998) New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program. Journal of Business and Economic Statistics, Vol. 16, No. 2., 127-177.
- Gomez, V. & Maravall, A. (1997) Programs TRAMO and SEATS : Instructions for the user (beta version: June 1997), Working Paper 97001, Ministerio de Economia y Hacienda, Dirrection General de Analisis y Programacion Presupuestaria, Madrid.
- Monsell, B.C., Aston, A.D., and Koopman, S.J. (2003) Toward X-13? ASA Proceedings.
- Njimi, H., Melard G. (2006) An empirical comparison between two automatic procedures for time series. IAP Statistics Network systems. Brussels University.
- Office for National Statistics, (2007) "United Kingdom Economic Accounts", No. 56-59, 1350-4401, Palgrave Macmillan
- Office for National Statistics, (2006) "United Kingdom National Accounts – The Blue Book", 1-4039-9398-X, Palgrave Macmillan
- Pena, D., Tiao, G.C., and Tsay, R.S. (2000) A course in Time Series Analysis. Wiley Press pp194
- Quenneville, B., Cholette, G., Huot, K., Chiu, and T.D. Fonzo (2004) Adjustment of seasonally adjusted series to annual totals. Research report, Statistics Canada.
- Sclove, S. (2002) Some remarks on exponential smoothing, University of Illinois, Chicago, Working Paper.
- Scott, S. (2006) Evaluation of X-11 and Model-based Seasonal Adjustment Methods for U.S. Economic Time Series, Conference on Seasonality, Seasonal Adjustment and their Implications for Short-Term Analysis and Forecasting Conference, May 2006, Eurostat, Luxembourg.
- Taylor, J. W. (2003) Short-term Electricity Demand Forecasting Using Double Seasonal Exponential Smoothing. Journal of Operational Research Society, 2003, Vol.54, pp 799-805.
- Winters, P.R. (1960) Forecasting sales by exponentially weighted moving averages. Man. Sci., 6, 324-342.

## About the Authors

**Fida Hussain** is a Methodologist within Time Series Analysis Branch (London). His work relates to seasonal adjustment, forecasting and interpolation issues. He can be contacted at ONS, Room D210, 1 Drummond Gate, London SW1V 2QQ; Tel: 0207 533 5978; Email: fida.hussain@ons.gsi.gov.uk.

**Dr. Craig H. McLaren** is Principle Methodologist within Time Series Analysis Branch (Newport). His work relates to seasonal adjustment, forecasting and interpolation issues. He can be contacted at ONS, Room A012, Government Buildings, Cardiff Road, Newport Gwent, NP10 8XG; Tel: 01633 655644; Email: craig.mcLaren@ons.gsi.gov.uk.

**Nigel Stuttard (corresponding author)** is Principle Methodologist within Time Series Analysis Branch (London). His work relates to seasonal adjustment, forecasting and interpolation issues. He can be contacted at ONS, Room D211, 1 Drummond Gate, London SW1V 2QQ; Tel: 0207 533 6324; Email: nigel.stuttard@ons.gsi.gov.uk.



## New Technologies and Tools for Study Management: Designing, Implementing and Maintaining a Web-Based Data Management System for a Multi-Site Longitudinal Intervention Study

Lauren Courtney, Emily Warmoth, Margaret Rodan, Kathy Katz, Siva  
Subramanian, Michele Kiely

### Abstract

This paper addresses the design, technical implementation and field experiences of using a Web-based data management system (DMS) for the *GirlTalk for Teen Moms Study*, being conducted in the NIH-DC Initiative to Reduce Infant Mortality in Minority Populations in Washington, D.C. (NIH-DC Initiative). The primary objective of this multi-site randomized clinical trial is to evaluate the effectiveness of a community-based intervention for minority teen mothers with the primary purpose of reducing subsequent teen pregnancies. The DMS monitors all activities for screening, recruitment, enrolment, in-depth interviews, randomization, interventions, repeated measures, outcomes, adverse events and follow-up activities for each participant over the course of their two-year enrolment in the study. Field staff from multiple sites use the DMS to manage over one hundred possible activities with a complex series of triggers that lead from one activity to the next.

The steps taken in designing the DMS, its technical implementation and the lessons learned from experiences in the field are presented. Previous studies in the NIH-DC Initiative utilized a PC-based data management system requiring nightly data transmissions to and from each site to synchronize study data. This Web-based DMS is compared with the earlier PC-based systems, and the efficiencies obtained are described, including real-time data access, simplified software management, data security, and anytime/anywhere access for all users. Cost considerations are outlined for the web-based DMS versus the PC-based system in terms of study size, study duration, number of sites and technical infrastructure. Several features of the DMS are discussed, including the role-based menu, events tracking, appointment scheduling, shared calendars, report generation, access to study documents, and uploading of data files to RTI secured servers. Lessons learned are discussed based on our field experiences and user feedback during the first year of implementation.

### Keywords

Web-based data management system, Teenage Pregnancy, Multi-site study, Longitudinal



## 1. Study Background: The NIH DC Initiative to Reduce Infant Mortality

In epidemiologic research effective tools for study management improve data quality, staff/time efficiency, cost effectiveness, and strengthen study outcomes (Unutzer et al, 2002; Winget et al, 2005). The use of a web-based data management system (DMS) is a relatively new technology in epidemiologic research, but provides significant benefits. The Internet is becoming widely available with nearly 70 percent of Americans reporting Internet access in December 2006 (InternetWorldStats, 2007). For the *GirlTalk for Teen Moms Study* (GirlTalk) a web-based DMS was developed to assist field staff from multiple sites in efficiently managing participant activities and data collection for this complex five-year study. This paper discusses the design and implementation of a web-based DMS for a multiple component multi-site longitudinal study. The features of the web-based DMS used currently are compared with a Personal-Computer (PC)-based DMS used previously by this research group.

### Study Background

GirlTalk is being conducted as part of the NIH-DC Initiative to Reduce Infant Mortality in Minority Populations in Washington, D.C. (NIH-DC Initiative). Washington, D.C. has historically reported some of the highest infant mortality rates (IMR) in the United States, with 11.0 infant deaths/1000 live births in 2004, as compared with a national average of 6.6/1000 (Munson et al, 2006). With particularly high rates, African Americans in Washington, D.C. reported 14.5/1000 infant deaths in 2002 (Kochanek et al, 2004). In response to these high mortality rates, beginning in 1993 the National Institute of Child Health and Human Development (NICHD) provided five-year grants to collaborating universities and research institutions in the Washington DC area to evaluate multiple contributing factors to infant mortality and develop effective community-based intervention strategies. The NIH-DC Initiative is currently in the third phase (conducted 2004-2009) that includes GirlTalk and four other NIH-DC Initiative studies. During Phase 2 (1998-2004) Project DC HOPE was conducted, and, like GirlTalk, had similar data management requirements. Previous studies in the NIH-DC Initiative, including DC HOPE, utilized a PC-based DMS requiring nightly data transmissions to and from each site to synchronize study data. In developing study management tools for Phase 3 RTI staff reviewed the DMS used in Phase 2 and developed an innovative DMS using new current generation applications to improve efficiency and data quality.

The primary objective of the GirlTalk Study, a multi-site randomized clinical trial, is to evaluate the effectiveness of a community-based intervention for minority teen mothers with the primary purpose of reducing subsequent teen pregnancies. Various staff from multiple organizations and sites are involved in data collection and implementation of the intervention for this study, each with unique roles and responsibilities. Staff from Georgetown University (GU) manage participant recruitment, enrolment, baseline interviews, counselling interventions and overall participant retention for the follow-up interviews. Staff from Children's National Medical Center (CNMC) conduct baseline and follow-up interviews and manage participant retention. As the data coordinating center (DCC), RTI International staff create data collection instruments, monitor data, and conduct data analysis.

## GirlTalk Study Procedures

Figure 1 provides a brief overview of the GirlTalk procedures. In total, 340 teens will be recruited at staggered intervals and enrolled for a two-year duration. After conducting a screening, staff enter eligible participants (e.g. age 15-19 years, African American or Latina, and primiparous) into the DMS. Teens and their mothers are enrolled as a dyad if a mother or mother-figure is available. Legal guardian written consent is required for teenage participants who are minors (under age 18).

Evaluation measures are collected at baseline and 6-month intervals for the teen and at baseline and 12-month intervals for the mother. Additionally, teens are contacted every three months to verify their pregnancy status through an Early Pregnancy Test (EPT) or self-report.

Randomization occurs after the baseline evaluation data is uploaded and receipt is confirmed by the DMS. Randomization is automatically generated by the DMS, although the request must be initiated by a staff member. If randomized into the intervention group, participants are given cell phones and receive phone counseling sessions over the two-year period. The counseling sessions promote behaviors which may postpone another pregnancy.

**Figure 1: Flowchart of the GirlTalk Study Activities**

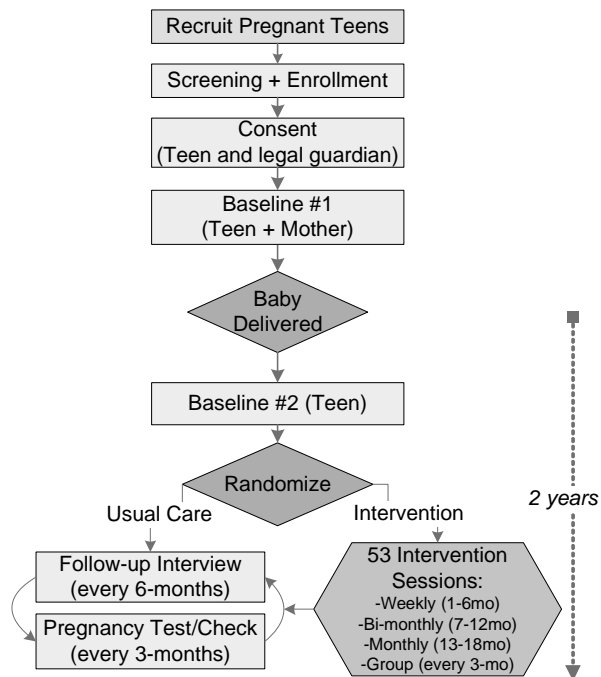


Figure 1 illustrates staff responsibilities for tracking a number of different procedures for each teen at different intervals because of a staggered enrollment. Field staff manage over one hundred possible activities for each participant with a complex series of triggers that lead from one activity to the next.

## 2. Design of the DMS

In order to develop the GirlTalk DMS, field and research staff worked collaboratively with the data management team to define design goals they wished the GirlTalk DMS to accomplish. Staff then completed the design process and documented the program requirements.

### Design Goals for the DMS

In planning the design of the DMS for the GirlTalk Study, staff reviewed the DMS developed for previous DC Initiative studies, specifically Project DC HOPE and developed a list of goals:

- Provide real-time data: The GirlTalk DMS should have “real-time data”, updated continuously and available at any location in order to effectively link the multiple sites in the study.
- Reduce staff time/burden: All data forms should be completed electronically so that forms with common data elements, such as contact information, can be linked and pre-populated to prevent the need to enter the same data in multiple places.
- Increase efficiency: By allowing multiple users to enter and access data in real-time, there is a decreased need for phone calls and email communications between sites to inform each other of activities that impact their responsibilities.
- Increase protocol adherence: Over the course of the two-year study, each of the 340 participants must complete over one hundred activities, including data collection and intervention sessions. Through reports, triggering subsequent events, and providing reminders, the DMS assists staff in monitoring the events for each participant ensuring that study activities are completed as per protocol.
- Increase Participant Retention: DMS reports allow study investigators to monitor participant progress and alert them of difficult-to-reach participants so that retention strategies can be implemented before participants become lost to follow-up.

### The Design Process

Development of the DMS began with the creation of a detailed flow chart of the activities required for each participant throughout the two-year enrollment period. The structure of the DMS was based on the three main components of the study: recruitment, intervention, and evaluation. The following elements of each component were determined collaboratively:

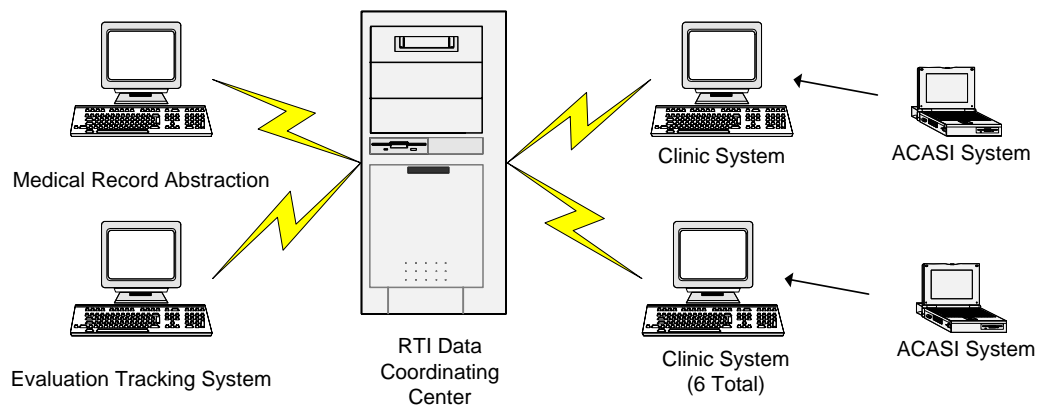
- Whether a data form should be completed on hard copy or electronically;
- Whether a report was needed to track a specific activity for purposes of scheduling or to monitor progress of the activity for purposes of quality control; what key data was required to meet the purpose of the report;
- The window of time allocated to complete each activity. This information was used to develop reports that would alert staff to upcoming activities and commit the necessary resources for completing the activity within the time frame. The schedule of completing individual study activities will also be used to explain analytical anomalies.
- Which study staff should have access to specific information in the DMS, in order to limit access based on each staff’s individual role. For example, to maintain the integrity of the study design, staff conducting the evaluation needed to be blinded from the randomization result.

The requirements documents were created to describe the various elements of the DMS design, and were used to plan for and execute the technical implementation.

### 3. Technical Implementation of the Web-Based DMS

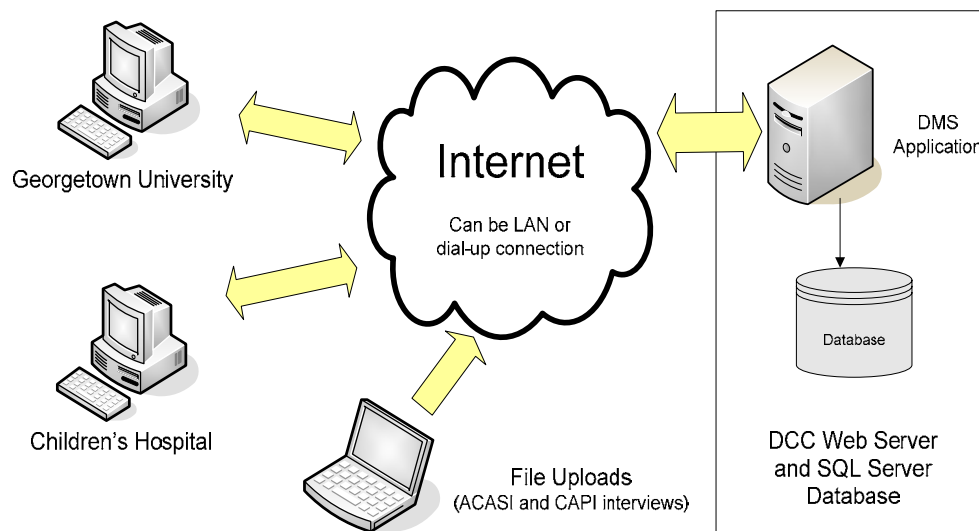
#### Description of PC-based and Web-based Systems

Figure 2 shows a graphic of the PC-based system that was used in the DC HOPE project described in Section 1. In this system, all transmissions were initiated by a central server at the DCC. Each night the central server dialed into the systems at each site and transferred the data. This design required successful nightly data transmissions to and from each site via point-to-point modem transmission. The physical security of the central server PC also had to be addressed.



**Figure 2: PC-Based System (DC HOPE)**

Exhibit 3 shows a graphic of the Web-based system that is being used for the GirlTalk study. Computers at GU and CNMC access the DMS through the Web. Audio Computer Assisted Self Interview (ACASI) and Computer Assisted Personal Interview (CAPI) data are uploaded through the DMS. Staff members usually connect over their university or hospital Local Area Network (LAN). Occasionally users will connect from home using a dial-up connection.



**Exhibit 3:**

### Web-Based System (GirlTalk)

The Web-based model eliminates the problems of delayed data transmission. A central server PC is no longer needed to manage the nightly data transmissions. Data are stored in a secure SQL Server database at the DCC. Since all the data are stored in a central database, data are available to users as soon as they are entered in the DMS or after interview data from a laptop is uploaded. No synchronization of the data is needed.

### Comparison of the DMS systems

The advantages and disadvantages of the PC and Web-based systems are compared in Exhibit 4. These comparisons guided our decision to use a web-based system to increase efficiency of the study.

**Exhibit 4: Comparison of the DMS systems**

	PC-Based System (DC HOPE)	Web-Based System (GirlTalk)
Offline Availability	Internet connection was not needed to access and use the DMS.	Internet connection is required to access and use the DMS.
Direct Remote Access	Modem connections at each site allowed developers to dial directly into the computers to troubleshoot problems and provide user support remotely.	Developers cannot dial in directly to users' computers to provide support. However, Fog Creek Copilot <sup>SM</sup> or a similar Web-based service can be used to troubleshoot users' problems by connecting to their computers via the Internet.
Learning Curve for Programmers	The application development tool (Microsoft Access) and the Windows environment were familiar to the programming team at the time (2000).	Developers needed to learn the chosen Web development technology (ASP.NET v 1.1). However, by building the GirlTalk DMS with widely used current generation tools (.NET and SQL Server) the "shelf life" of the system will be extended.
Software Updates	Whenever the software was modified, to fix bugs or add new features, updates had to be transmitted to each PC and installed. On-site technical support had to be provided for locally installed software.	Updating the DMS software is simplified since the application files are housed and managed at the DCC. When the application must be changed, the programmer updates the files on the DCC's Web server. Users immediately see the changes in their Web browser. Since there is no need to update software installed on individual users' computers, this speeds the correction of bugs and implementation of changes in study procedures.
DMS Availability	The DMS could only be used on specific PCs where the software had been installed.	Users can access the DMS from any computer that has an Internet connection and the Internet Explorer browser.
Synchronization of Data	Synchronization of the data between the sites and the DCC did not take place in real time. The	Data are instantly available as soon as they are entered in the DMS.

	PC-Based System (DC HOPE)	Web-Based System (GirlTalk)
	nightly data transmissions resulted in at least a one day delay.	
Physical Security of the Data	All study data, including participant identifiers, were stored on PCs at each site. Thus, the physical security of the PCs was a concern.	Users enter data directly into the Web application, and data is stored in a secure database at the DCC. When the data are travelling across the Internet, they are encrypted using Secure Socket Layer technology. Only ACASI and CAPI files are stored on PCs, and these files contain no identifying information.
Quality of Internet Connection	Telephone connections were unreliable and slow, usually less than a 56 Kbps connection.	Staff connect over their university or hospital Local Area Network (LAN), which is fast and reliable.
Timeliness of Data Entry	Users often did not enter data in a timely manner, since the DCC could not view the data until a transmission took place.	Because data can be entered at a user's location at any time, users tend to enter data soon after an event has occurred. This improves the timeliness of data and reduces errors.
Richness of User Interface	Applications that run on PCs (e.g. MS Word or Outlook) have a rich and "smooth" interface, allowing users to navigate through many tasks on a complex screen. Since the application runs locally, responsiveness is quick.	For Web-based applications the user interface toolset is smaller and more limited. With Web pages the user typically performs one simple task per screen. Response time can be slower. When a user performs an action on a web page, such as clicking a button, it often involves a trip to the server and back.
Data monitoring, sharing and coordinating	Data were not shared between sites. Data were received and aggregated at the DCC, where monitoring took place after a delay.	Data monitoring, sharing, and coordinating between sites is easy and quick. Study investigators can log in to the DMS and monitor the study in real-time.
User Training	Training sessions had to be conducted on-site. Users had to be taught the functionality of the DMS screens.	Since the sites already have Internet connections, training can be done remotely. Also, many users are already familiar with Web browsers which may reduce training requirements.
Hardware and Support	Studies often purchase identical equipment for staff to simplify the support of hardware and software. Studies must consider whether purchased equipment fits within the limitations of the particular computing environment at the site.	With a Web-based system, staff may use their own computers and receive general maintenance and support from their IT staff. These factors save the project time and money.

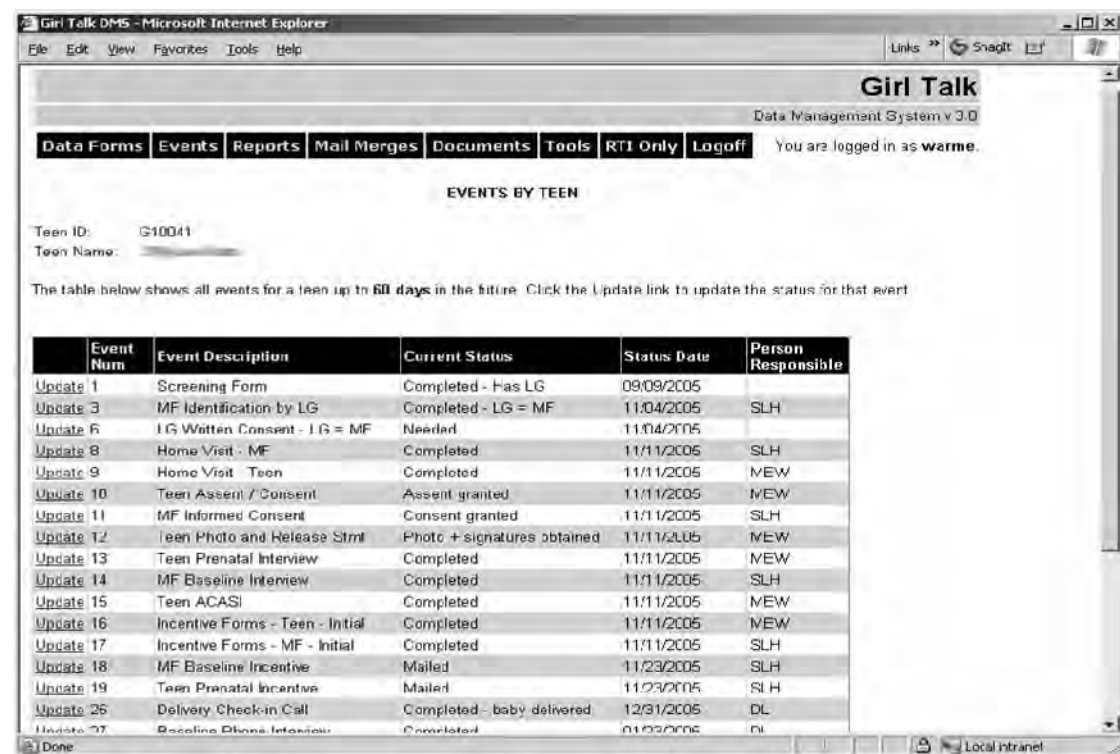
#### 4. Features of the web-based DMS

**Role-based Menu:** Access to DMS features is controlled through roles assigned to menu items. Each menu item has certain roles assigned to it, such as the “counselor” role or the “interviewer” role. The advantage is that the DCC can develop and maintain one application for all users. Figure A in the Appendix shows an image of the role-based menu.

When the counselors at GU log in, they see the menu shown in the top image in Figure A. They have access to all of the options under Reports. When the interviewers at the CNMC log in, they see the menu in the bottom image with fewer options under Reports. Thus, the role-based menu can be used to hide information based on a person’s role in the study, for example, staff conducting the evaluation would be blinded from the randomization result

**Events Tracking:** The Events feature allows staff to track all of the study activities and events for each teen. The Events grid lists a Description of the Event, the Event Status, the date the Status was updated, and the Person Responsible for that status. Exhibit 5 shows an image of the Events feature.

##### Exhibit 5: Events Feature in the Web-based DMS



The screenshot shows a web browser window titled "Girl Talk DMS - Microsoft Internet Explorer". The page header includes "Girl Talk Data Management System v3.0" and a navigation menu with links: Data Forms, Events, Reports, Mail Merges, Documents, Tools, RTI Only, and Logoff. The user is logged in as "warne". The main section is titled "EVENTS BY TEEN" and displays information for Teen ID: G10041 and Teen Name: [redacted]. A note states: "The table below shows all events for a teen up to 60 days in the future. Click the Update link to update the status for that event".

	Event Num	Event Description	Current Status	Status Date	Person Responsible
<a href="#">Update</a>	1	Screening Form	Completed - Has LG	09/09/2005	
<a href="#">Update</a>	3	MF Identification by LG	Completed - LG = MF	11/04/2005	SLH
<a href="#">Update</a>	6	LG Written Consent - LG = MF	Needed	11/04/2005	
<a href="#">Update</a>	8	Home Visit - MF	Completed	11/11/2005	SLH
<a href="#">Update</a>	9	Home Visit - Teen	Completed	11/11/2005	MEW
<a href="#">Update</a>	10	Teen Assent / Consent	Assent granted	11/11/2005	MEW
<a href="#">Update</a>	11	MF Informed Consent	Consent granted	11/11/2005	SLH
<a href="#">Update</a>	12	Teen Photo and Release Stmt	Photo + signatures obtained	11/11/2005	MEW
<a href="#">Update</a>	13	Teen Prenatal Interview	Completed	11/11/2005	MEW
<a href="#">Update</a>	14	MF Baseline Interview	Completed	11/11/2005	SLH
<a href="#">Update</a>	15	Teen ACASI	Completed	11/11/2005	MEW
<a href="#">Update</a>	16	Incentive Forms - Teen - Initial	Completed	11/11/2005	MEW
<a href="#">Update</a>	17	Incentive Forms - MF - Initial	Completed	11/11/2005	SLH
<a href="#">Update</a>	18	MF Baseline Incentive	Mailed	11/23/2005	SLH
<a href="#">Update</a>	19	Teen Prenatal Incentive	Mailed	11/23/2005	SLH
<a href="#">Update</a>	25	Delivery Check-in Call	Completed - baby delivered	12/31/2005	DL
<a href="#">Update</a>	27	Baseline Phone Interview	Completed	01/03/2006	DL

All users access the same Events feature, but the DMS determines the subset of events to display based on the user’s study site. Therefore, users only see events that are necessary to perform their jobs.

The Events feature can be thought of as a detailed flow chart for the study. Each event is triggered by a previous event, so that the tracking system is as automated as possible. The user does not create any events manually, but is responsible for updating the status of events. Every activity the user completes is recorded as a status change in the Events feature for easy reference and consistency.

The Events feature utilizes a table-driven design, where events and their possible status codes are stored in tables in the study database. Every time a user updates the status of an event, a

SQL Server trigger is fired which checks the event tables to determine if new event(s) need to be created or if the status needs to be changed on an existing event.

The Event process is a central piece of the DMS. Most of the reports use a certain Event and Status to determine which records to display (for example, "Pregnancy Tests" that are "Needed"). In addition to customized reports, the DMS contains an interactive report known as the "Create Your Own Event Report," which allows users to generate a report dynamically for any event and status code to which they have access. This feature substantially reduces the need for programmer-developed custom reports.

**Appointment Calendar:** Each location has its own appointment calendar. The one shown in Exhibit B is for GU. Appointments can be filtered by staff member. When the user clicks the link for an appointment, a small window pops up with details about the appointment. The pop-up window includes a MapQuest link to the appointment location, so that the users can view a map of the location of the appointment.

Appointments are also displayed for other protocols in the DC Initiative. For example, in addition to conducting evaluation interviews for GirlTalk, staff at CNMC also conduct interviews for two other DCI studies. Appointments from the other studies, which are displayed on the GirlTalk calendar as read-only, allow staff to easily identify scheduling conflicts.

**Report Generation:** Reports are generated in the DMS using real-time data from all sites. The DMS report, "Pregnancy Test Reminders and Self-Check Calls", shown in Exhibit C-Appendix, is an example of how information from different components feeds into one report. The staff at CNMC call the teens about their pregnancy tests, but the counselors at GU are responsible for keeping the teen's primary phone number up-to-date. When the counselors update the phone number, the change is instantly reflected on this report. Thus, from this report, the CNMC staff have the latest contact information for the teen.

**Study Documents:** Study documents, such as consent forms, showcards and questionnaires, are stored centrally in the DMS to ensure that users always access the most recent version of a document. Exhibit D in the Appendix shows the Documents feature.

**Uploading Interview Data:** The counselors and interviewers use the DMS to upload ACASI and CAPI data files to the server at the DCC, thus eliminating the need for a separate transmission system to send the data to the DCC. After the data files are uploaded, an automated process moves the interview records into the central study database. At this point selected ACASI and CAPI data items can be viewed on reports in the DMS. The DMS also contains "Data Uploads" reports showing which interviews have been uploaded for each teen. Exhibit E in the Appendix shows an image of the "Upload Files" page.

**Mail Merges:** Because mailings are frequently sent to each participant during the course of her enrollment, necessary cover letters with pre-populated name, address and date are generated by the DMS. When the Events section indicates that a letter is needed, the DMS creates the letter which is printed by staff.

**Participant Notes:** A large free-form text box is available for staff from all sites to add notes for any participant. Comments are shared between sites, serving as a place to exchange information on each participant.



## 5. Discussion of Lessons Learned

In April 2005 development of the DMS began, with the first DMS staff training conducted in July 2005. The DMS was launched in three stages: Stage 1 in September 2005 contained features for Recruitment. Stage 2 in November 2005 contained features for Evaluation. Stage 3 in February 2006 contained features for the Intervention. We have released a total of 13 versions of the DMS since its inception. As of July 1, 2007, 22 DMS users have logged more than 6750 sessions to record study activities and data. A total of 221 participants have been screened, and 159 have been randomized. Over 7260 events have been generated for study participants, with an additional 4070 events scheduled as due in the future. In the next sections we discuss the lessons learned from our experiences in developing, implementing and maintaining the DMS.

### Design Lessons Learned

Efficient data management that fulfills client and user expectations is necessary, especially for research studies which often have budget limitations. Even with the most complex studies, it is best to maintain a simple system.

In order to reduce labor and programming time, much of the study planning must be completed before the DMS programming begins. Advance preparations will reduce labor costs as well as ensure that an organized framework is created for the DMS. All required data entry forms should be conceptualized at the beginning. This will reduce the quantity of forms needed and enable staff to create a more concise structure. In addition, all study procedures to be monitored by the DMS events tracking system should be finalized in advance to keep the structure simple and reduce the need for re-programming. Any modification to the event tracking added at a later date may cause complications because of their impact on other events.

Before the DMS is put into the field, a pilot test is highly recommended. Because of time constraints the GirlTalk DMS was not piloted. When the study experienced some procedural changes in the first few months of the study, these impacted the DMS which required additional reprogramming.

While the sophisticated web-based DMS contains enhanced features which allow for some automatic triggers, users tend to expect the DMS to conduct all processes automatically, which is not the case. Users must ensure that all data entered into the DMS is correct--if data input into the DMS are incorrect, the DMS will also be incorrect. Because multiple users access data in the DMS, it is especially critical to double-check accuracy.

Although the DMS eliminates much need for continual communication between sites, the GirlTalk Study recognized the importance of holding weekly or bi-weekly multi-site staff meetings. The weekly meetings were established to discuss common issues faced by all sites, in addition to addressing potential errors or questionable data in the DMS.

### Implementation Lessons Learned

A phased approach to implementing a system works well when the development time frame is short, as it was for GirlTalk. The study team determined which features users initially needed and programmed those first. Then the DMS was updated incrementally as the study progressed and other features were needed.

A flexible DMS design needs to be balanced with features that are automated and controlled by the DMS. While users often seek flexibility, in order to maintain data quality it is important that the DMS holds some rigidity. For example, because randomization is critical it must be controlled by the DMS. Before allowing randomization to be completed, the DMS confirms that all required baseline data has been received by the DMS. In addition, the DMS prevents randomization from occurring more than once per teen.

Another lesson learned was to avoid re-inventing the wheel. If a software tool with the desired functionality already exists, study teams should consider using it. For the GirlTalk project, the users requested complex appointment scheduling features. RTI recommended that staff use Microsoft Outlook for a complex level of functionality, and use the DMS for more basic scheduling functions. Ultimately the DMS users decided to use the basic scheduling features programmed in the DMS.

Finally, it is important to consider all requests from users and take advantage of their input to build a better system. Users originally brought forth the idea to add a calendar to the DMS which was originally thought to be a complex enhancement requiring resources that were limited. However, with brief investigation, the calendar feature was found to be easy and quick to implement.

### Cost Considerations for a Web-Based System

The initial design and development of a Web-based DMS can be expensive; however, the following are some factors to consider in determining whether or not the cost can be justified:

- Size of study – The size of the study can refer to the number of cases, number of study activities, and/or amount of data being collected. The larger the study, the lower the cost per case or activity will be.
- Duration of study – If the study runs over a long period of time, such as several years, then the investment per year will be reduced.
- Multiple sites – A Web-based DMS is especially suited for multi-site studies because it enhances site-to-site communication by providing users from all sites with real-time data.
- Infrastructure – There are technical support costs associated with maintaining the Web server, database server, and Internet connections. These costs need to be weighed against the costs associated with maintaining a PC-based system in the field. Project leaders need to consider who will be doing the technical support and what it will cost in both the Web- and PC-based scenarios.
- Reusability – Once a Web-based system is developed for one study, it may be possible to reuse it for other studies. The potential opportunity to recycle the DMS infrastructure across different studies may justify the initial investment in developing the system.

## 6. Conclusions and Outcomes

The Web-based DMS creates an efficient and effective management tool for complex multi-site studies. Customized reports allow managers to monitor study progress and conduct regular quality control reviews to ensure data quality. Providing staff with real-time access to

all data reduces the volume of communication required between sites. However, regular communication between sites and among staff continues to be a necessary aspect of study management, allowing staff to confirm or question data inputs. It is important to remember that the accuracy of the system depends on the accuracy of the data that is entered by the user.

Using a Web-based DMS alleviates many of the problems encountered with a PC-based system. For example, transmission difficulties are reduced. The real-time access to data eliminates issues of sharing data between sites.

The Web environment allows for new features that are difficult or impossible to implement with stand-alone PC systems. For example, centralized events tracking, appointment scheduling, and document storage are not possible in the conventional PC-based DMS.

The Web-based DMS is a cost-effective investment for studies which require complex management and a large sample size or lengthy longitudinal studies. Once developed, the DMS infrastructure and portions of the code can be re-used to create new DMS applications for additional studies.

## 7. Future Directions

Future objectives for the Web-based DMS include building new tools to speed the development of DMS applications. For example, RTI would like to create a rapid development tool for the data entry forms. This tool would utilize a codebook containing information about the data items on a form and generate the data entry forms automatically.

A second enhancement is to manage the event specifications within the DMS instead of in Microsoft Excel. Currently, all event specifications are created, managed, and maintained in an Excel spreadsheet and loaded manually into the event tables in the study database. A new feature could be added allowing RTI staff to create events and status codes, associate certain status codes with each event, and specify which events trigger other events. This feature would directly modify the event tables in the study database, thereby eliminating the burden of maintaining the event specifications in two different formats and keeping them in sync.

At this time RTI is migrating the GirlTalk DMS and the other DCI DMS applications to ASP.NET version 2.0. ASP.NET 2.0 offers many new features that developers had to custom build in 1.x. The Web-based DMS will evolve to take advantage of the latest capabilities. For example, the new Site Navigation piece of ASP.NET 2.0 allows a developer to define the navigational structure of the website. This data can then be used by a variety of Web controls, to display site maps, breadcrumbs, treeviews, or menus that highlight the site's navigation and show the user's location in the site.

Also, the new ASP.NET AJAX extensions allow the developer to build web applications in ASP.NET 2.0 that can update data on the web page without a complete reload of the page (a "round trip" to the server). The user may not even perceive that a page update is occurring and can continue interacting with the page while the update is happening in the background. This capability will enhance the richness and responsiveness of the DMS application.

## Acknowledgements

The NIH DC Initiative was developed with funding from the National Institute of Child Health and Human Development. The authors would like to acknowledge the support and assistance of Jutta Thornberry, Nabil El-Khorazaty, David Forvendel and Margo Brinkley from RTI International, as well as the GirlTalk staff at Georgetown University and Children's National Medical Center.

## References

- Internet Usage Statistics for the Americas. (March 10, 2007). Retrieved April 18, 2007, from <http://internetworldstats.com/stats2.htm>
- Kochanek KD, Murphy SL, Anderson RN, Scott C. (2004). Deaths: Final data for 2002. National vital statistics reports; vol 53 no 5. Hyattsville, MD: National Center for Health Statistics.
- Munson ML, Sutton PD. (2006). Births, marriages, divorces, and deaths: Provisional data for 2005. National vital statistics reports; vol 54 no 20. Hyattsville, MD: National Center for Health Statistics.
- Unutzer, J., Y. Choi, et al. (2002). "A web-based data management system to improve care for depression in a multicenter clinical trial." *Psychiatr Serv* 53(6): 671-3, 678.
- Winget, M., H. Kincaid, et al. (2005). "A web-based system for managing and co-ordinating multiple multisite studies." *Clinical Trials* 2(1): 42-9.

## About the Authors

Lauren Courtney, MPH, is a Research Epidemiologist at RTI International in Washington, DC. Ms. Courtney is responsible for managing survey and biospecimen data collection activities for multiple domestic and international studies, with a focus on reproductive health and HIV/AIDS. She is skilled in designing study protocols, training field staff, monitoring data collection activities, and analyzing data. For the GirlTalk study she monitors data collection activities and participated in design of the DMS. She can be contacted at RTI International, 701 13th St, NW Suite 750 Washington, DC 20005 USA; tel. (202) 728-2498; fax (202) 728-2095; e-mail [lcourtney@rti.org](mailto:lcourtney@rti.org).

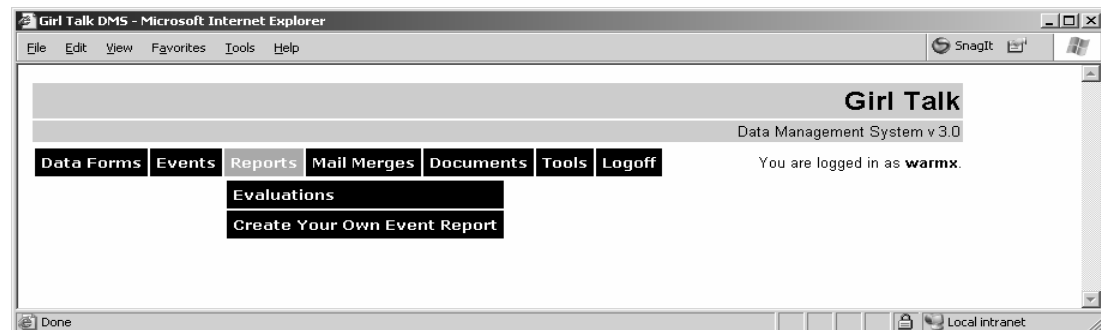
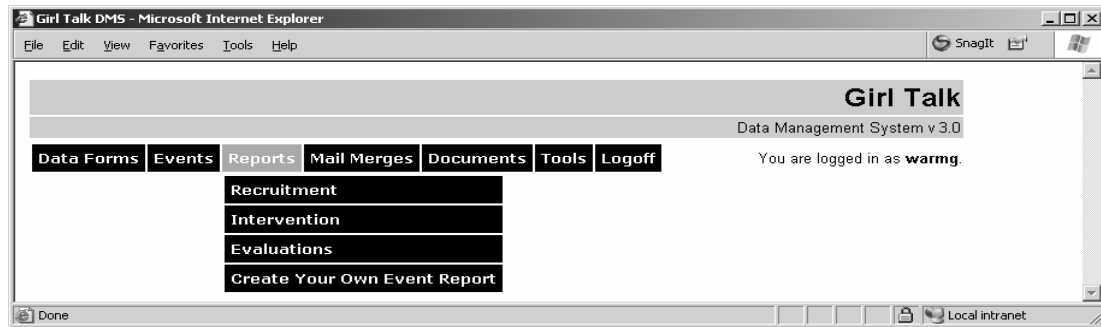
Emily Warmoth, MSIS, is a Research Programmer/Analyst in the Research Computing Division at RTI International. Ms. Warmoth is responsible for designing, programming, and maintaining data management systems for multi-site studies. She has worked on a range of data management systems developed for mobile, Windows, and Web platforms. She can be contacted at RTI International, PO Box 12194, Research Triangle Park, NC 27709 USA; tel. (919) 316-3750; fax (919) 541-6178; e-mail [ewarmoth@rti.org](mailto:ewarmoth@rti.org).

Dr. Margaret Rodan, ScD is the Project Manager for the GirlTalk study with Georgetown University. Dr. Kathy Katz, PhD, is a Clinical Psychologist for the GirlTalk study with Georgetown University. Siva Subramanian, MD, a Professor and Chief of Neonatology at Georgetown University Hospital, serves as principle investigator for the GirlTalk study. Michele Kiely, DrPH, serves as the project officer of the DC Initiative to Reduce Infant

Mortality in Minority Populations for the National Institute of Child Health and Human Development.

## Appendix

### Exhibit A: Role Based Menu



### Exhibit B: Appointment Calendar

The screenshot shows the 'Girl Talk' Data Management System v 3.8 interface. The user is logged in as 'courl'. The page displays the 'APPOINTMENT CALENDAR - Georgetown'. Below the title, it states: 'This calendar displays all appointments that have been entered in the DMS using the "Schedule Appointment" form.' A filter dropdown is set to 'DAN'.

June 2007						
May	Sun	Mon	Tue	Wed	Thu	Fri
27		28	29 3:30 PM (DAN) Session 31 (G10298)	30 10:00 AM (DAN) HV-MF&Teen (G12062)	31 12:00 PM (DAN) HV-MF&Teen (G12146)	1
3	4 11:00 AM (DAN) Session 17 (G11171)	5 11:30 AM (DAN) Session 8 (G11593)	6 3:30 PM (DAN) Session 15 (G11502)	7 6:30 PM (DAN) Session 12 (G11080)	8 12:00 PM (DAN) Session 1 (G12088)	9
10	11 11:00 AM (DAN) Session 18 (G11171)  4:30 PM (DAN) Session 5 (G11965)	12 2:00 PM (DAN) HV-MF (G12153)  2:00 PM (DAN) Session 4 (G11890)	13 10:30 AM (DAN) Session 17 (G11114)  2:00 PM (DAN) HV-MF (G12120)	14 9:30 AM (DAN) Session 6 (G11643)  2:00 PM (DAN) Session 14 (G11221)	15 10:00 AM (DAN) HV-MF (G12161)  11:00 AM (DAN) Session 32 (G10298)  12:00 PM (DAN) Session 2 (G12088)	16

### Exhibit C: Report Generation

**Girl Talk DMS - Microsoft Internet Explorer**

File Edit View Favorites Tools Help

**PREGNANCY TEST REMINDERS & SELF-CHECK CALLS**

**Pregnancy Test Reminder Calls Needed**  
Below are the Reminder Calls needed in the next 4 weeks.  
Pregnancy Test must be completed within **28** days of the Status Date listed below.

Check off	TeenID	Teen Name	Primary Phone	Event	Status Date
	G10082			Pregnancy Test Reminder - 6 Mo	05/06/2006

**Pregnancy Self-Check Calls Needed**  
Below are the Self-Check Calls needed in the next 4 weeks.  
Pregnancy Self-Check must be completed within **28** days of the Status Date listed below.

Check off	TeenID	Teen Name	Primary Phone	Event	Status Date
	G10165			Pregnancy Self-Check Call - 3 Mo	03/22/2006
	G10223			Pregnancy Self-Check Call - 3 Mo	04/27/2006
	G10272			Pregnancy Self-Check Call - 3 Mo	05/01/2006

### Exhibit D: Study Documents

**Girl Talk DMS - Microsoft Internet Explorer**

File Edit View Favorites Tools Help

**Girl Talk**  
Data Management System v 3.8

**Data Forms** **Events** **Reports** **Mail Merges** **Documents** **Tools** **RTI Only** **Logoff** You are logged in as **warmr**.

Childrens	Questionnaires	Teen Baseline Quex
Georgetown	DMS Users Guide for Childrens	Teen Baseline Infant Death Version
	Delivery Confirmation Call	Teen 6mo Quex
	Self-Check Call Script	Teen 12mo Quex
	Teen Scheduling Interview Scri	MF 12mo Quex
	MF Scheduling Interview Script	
	Positive Pregnant Form	
	GirlTalk Telephone Interview Contact Log	
	Adverse Event Form	

Local intranet

### Exhibit E: Interview Data File Upload

**Girl Talk DMS - Microsoft Internet Explorer**

File Edit View Favorites Tools Help

**Girl Talk**  
Data Management System v 3.1

**Data Forms** **Events** **Reports** **Mail Merges** **Documents** **Tools** **RTI Only** **Logoff** You are logged in as **warmr**.

**UPLOAD FILES**

Select Source Computer:

Done Local intranet

# Knitting Patterns: for interview and analysis

Stephen Jenkins

## Abstract

This paper describes an implementation of lexical patterns, based on Regular Expressions and developed for use with Snap survey software. The intention is to make electronic interviews more amenable to respondents whilst providing researchers with in-depth response interrogation tools.

The paper is of interest to all those conducting online and offline research using direct respondent input into electronic questionnaires, and those writing software for such applications.

## Keywords

Survey Metadata, Regular Expressions, Data Validation

## 1. Motivation

The past five or more years has seen a rapid growth in the use of electronic questionnaires such as those intended for web, kiosk and laptop-based interviewing. Key benefits cited by proponents of the technology are that interviews are faster and cheaper because the manual data entry step involved in paper questionnaires is removed.

That simple observation ignores the fact that the process of keying data from paper questionnaires often does not just involve the direct transcription of respondent replies. With open-ended questions in particular there is often a requirement to normalise the replies during keying. The process involves application of one or more transformations of the respondent data in order to perform operations such as:

- remove currency symbols (or other symbols defining units)
- type conversions (e.g. convert a height in feet and inches into metres)
- change worded forms of numbers into numerals (e.g. convert a response of two into 2)
- interpret formatting within numbers (e.g. 1k becomes 1000 and 1,234 becomes 1234)

It is our assertion that if alternate methods of expression such as these are disallowed to respondents of electronic questionnaires by overly restrictive formatting requirements (as they frequently are), then respondents get put off completing. Use of this technology could thus contribute to increased response rates.



## 2. Looking for solutions

As has been alluded to above, the problem of accepting non-conforming respondent replies to questions exists even in the world of paper questionnaires. Aside from the data-entry-time transformation of responses into a standard form mentioned previously, other techniques are used to try to guide the respondent towards giving a response in an acceptable format. These standard solutions generally involve either printing constant parts of the response in the hope (not always realised) that the respondent will not repeat them, or splitting the single input into parts, often separated by constant elements.

For example, when requesting currency amounts, one might see:

£

or:

£  .

Sometimes the elements are labelled, so for date inputs one might see:

/  /   
day month year

None of these is foolproof of course. It is not unusual to see:

£

given as a response on a paper questionnaire.

For electronic questionnaires we are working on the premise that a questionnaire designer would want to be as friendly as they can be towards respondents. They would thus want to be in a position where 1.5k is an acceptable answer to an appropriate quantity question. In general, we can consider that a questionnaire designer has three choices when dealing with such a situation:

1. Constrain the input to the exact format anticipated and perform calculations and range checks etc. on the fly during the interview. In our view these formatting restrictions are unacceptable as a blanket rule, for the reasons stated.
2. Allow any input at interview time and clean the problem up when interview data is collated. This is fine but the opportunity to get the respondent to correct inadvertent typographical or magnitude errors will usually be lost. Furthermore, the response to such a question (whether immediately valid or not) cannot be used in routing or any later calculations.
3. Impose some flexible interpretation mechanism, specialised to the question, that allows appropriate flexible input forms but is still able to expose the intended value to the underlying validation, routing and calculation processes in order to continue driving the interview.

Our solution to this problem is based on Patterns as exemplified in Regular Expressions. Basing the system on a ubiquitous technology means that formatting constraints are more

easily represented in different systems thus allowing, for example, rewriting as JavaScript Regular Expressions for use in web/HTML surveys.

### 3. Regular Expressions

Regular Expressions provide a way to match text with patterns. In general they provide a powerful way to find and replace strings. They are typically used in two ways:

- To provide text-file editing facilities (implementing the common find-and-replace operation in text editors)
- To validate and interpret user input in windows and web-forms.

It is this latter use that we are most interested in for validating respondent input into electronic questionnaires.

#### An example

Consider the problem of validating the response to a date question such as *When did you enrol?*. We want to validate respondent replies such as the following:

1/9/07                  01/09/07                  01.9.07                  1.09.07

That is, we assume that the date will be of the form dd/mm/yy where the dd part is one or two digits, the mm part is one or two digits and the yy part is exactly 2 digits. For flexibility, we want to recognise either a dot or a slash as a separator.

The (or more accurately, a) regular expression for such a pattern is:

`\d{1,2} (\/|\.) \d{1,2} (\/|\.) \d\d`

The construction to `\d{1,2}` specifies "between 1 and 2 digits", the `\/|\.` part specifies "either a slash or a dot", and the construction `\d\d` specifies "two digits". So the whole pattern reads as "1 or 2 digits followed by a slash or a dot, followed by 1 or 2 digits, followed by a slash or a dot, followed by 2 digits."

As complex as it is, the regular expression cited is only an approximate specification for a date. Whilst it allows inputs such as 19/09/07, which could later be interpreted as 19<sup>th</sup> September 2007, it would also allow 58/93/07 (the 58<sup>th</sup> day of the 93<sup>rd</sup> month?) and 19/09.07 (with its inconsistent use of field separators.) Although it would be possible to extend beyond the simple pattern shown to reduce some of these issues, it would not sensibly be feasible to extend it to the level of validating the day, month and year values to ensure that only valid dates were passed (for example to reject 29/02/07). Those value constraints normally require the use of external code: the pattern identifies the individual parts and the outer code validates the parts in conjunction with one another.

So, whilst regular expression patterns are very expressive and powerful in the right hands, and given the right application, they are very complex to learn and to apply successfully for any requirements beyond the most basic in our field of interest.

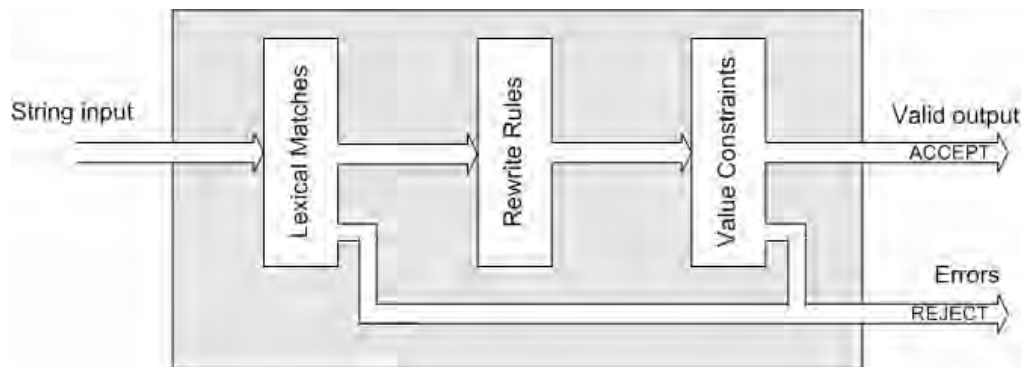
Furthermore, if provided as the sole means of allowable format specification, they require a user (in our case the user of survey questionnaire design or analysis software) to learn another language which, because of its level of abstraction, removes them from the immediate world of survey data manipulation.

#### 4. The essence of a pattern

We can consider the validation process required of a pattern to be comprised of three steps:

1. Attempt one or more lexical matches
2. If successful, possibly rewrite the input in a different form
3. Check that the result is an acceptable value

The steps of the validation process are illustrated in the following figure and described below.



**Figure 1 – The steps in pattern validation**

The **Lexical Matches** specify the allowable syntax of the input. Note the lexical match candidates can either be literal values or more patterns (and thus in that way a pattern may be defined recursively.)

If one of the lexical matches finds the input acceptable, then the **Rewrite Rules** are invoked as the next step. These can either:

- Take components of the input and rewrite them in a standard way. For example, to accept 20.3.07 or 20/3/07 as a date but to output them in a standard form as 20/03/07.
- Convert all letters to upper- or lower-case (this is actually a variant of the above but identified separately as it is such a common requirement).
- Write something lexically different to, but semantically the same as, the validated input. For example, a pattern that converts worded numbers into digits would include a pattern that recognised `two` as input but which writes out `2` as the result.

The final step is to apply **Value Constraints**. Typically these are comprised of a single range of allowable constant values. Value constraints may have been applied already as a side-effect of the available Lexical Matches, for example suppose a pattern was to be set up to allow the respondent to enter integer values of 1 to 3 only: if the Lexical Matches section allowed 1 or 2 or 3 then the Value Constraints could be empty. If the Lexical Matches allowed 0 through 9 as valid input then a Value Constraint of 1 to 3 should be introduced to provide a final check on the value.

#### 5. Patterns as implemented

An overall goal was to provide the user with a way of creating patterns to validate input that are at least approaching the power of regular expressions, and which may be translated into regular expressions for publication to electronic forms such as HTML.

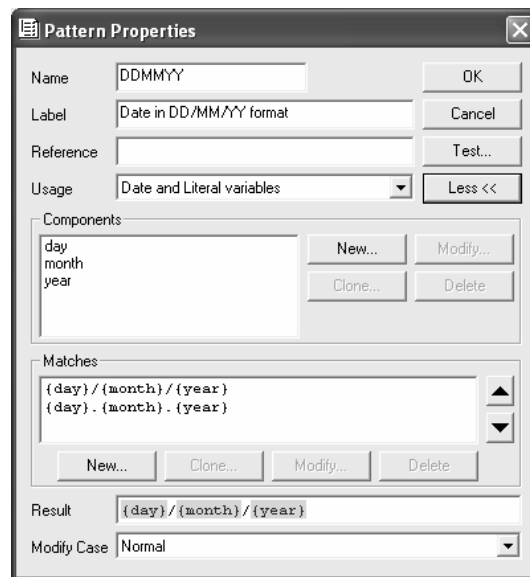
The basic idea introduced was to base the description of the allowable lexical matches on parameterised strings. Snap users are familiar with the use of these to provide text substitution (sometimes known as response-piping) at interview time. For example, they might include a question which includes the following sequence:

Q1 Which of these fruits is your favourite?: Apples / Pears / Bananas

Q2 You said your favourite was {Q1}, why is that? \_\_\_\_\_

Then at interview time, the text of Q2 has the response from Q1 inserted to specialise the question to the individual respondent.

We used a similar technique to enable the user to specify lexical matches. Consider the example pattern below which validates dates (it does essentially the same job as the sample regular expression pattern for parsing dates described in section 3.)



**Figure 2 – Pattern properties dialog**

Amongst other things it contains three **Components**: `day`, `month`, and `year` which are themselves patterns. The lexical matches (shown in the **Matches** section of the dialog) are then described in terms of the components and literal values. Any (respondent) input passing one of the lexical matches is considered valid and is then processed by the **Result** section. In the example shown, the components are again used to standardise the result (and in this case no case conversion is applied.)

It is now quite easy to allow other lexical match formats. For example, to allow dashes as a separator, the user simply needs to create a new match of `{day}-{month}-{year}`. As an alternative, it would be possible to specify the separators as a list of alternatives and associated with a `separator` component. If that component is set to consistent then occurrences of it found in one string must be identical (and hence only slashes, or only dots, or only dashes would be allowed as the separators in a valid date response.) However, working in the way shown helps to keep the pattern easier to compose, understand and test.

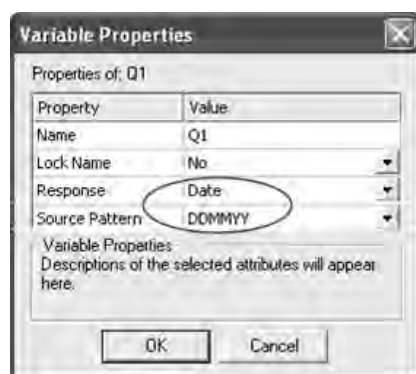
## 6. Patterns and in-built response types

When a user creates an open-ended question in Snap, they select a response type from the list of in-built types (shown in Table 1, below.) For example, the in-built Quantity type allows positive or negative decimal numbers with or without a decimal point. The in-built Date type allows the usual dd/mm/yy forms with 2 or 4 digit years and a variety of separators. It also allows for dates to be expressed in worded form, for example 1st September 2007 would be considered an acceptable response.

Type	Description
Quantity	A numeric value, positive or negative, decimal or integer
Date	A date based on the Gregorian calendar
Time	A time of day (from 0:00 to 23:59:59)
Literal	Any other type of open response

**Table 1 – In-built open-ended response types**

When the user is entering questions and assigning each a response type (Figure 3), they are in fact associating in-built, patterns (which the user doesn't see) that implement the basic validation for the type. Patterns created by the user are associated with one of the in-built types<sup>1</sup> (see Figure 4.) The program uses this association to present a list of possible patterns available for the question (variable) being edited.



**Figure 3 – Variable properties dialog**



**Figure 4 – Pattern properties dialog**

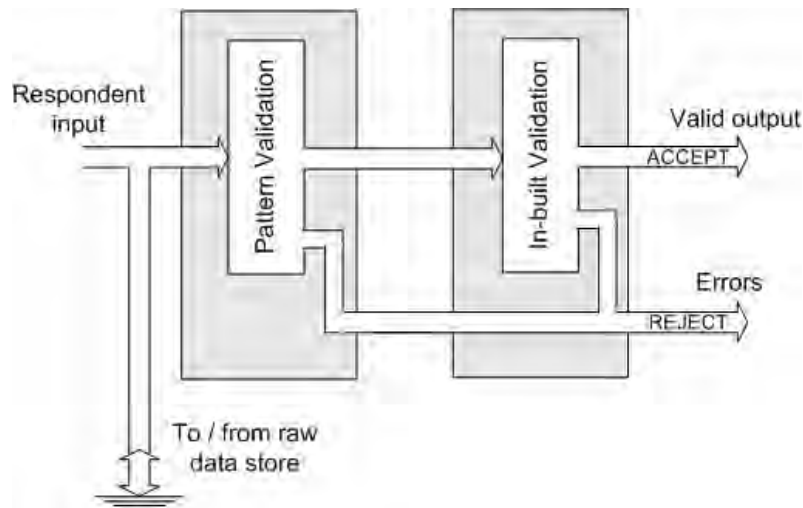
### Processing data through patterns

As has been discussed, the fundamental purpose of a pattern is to validate respondent input. That input could be coming from one of two places:

<sup>1</sup> Note that since any respondent data can be represented as a text string, any pattern intended for Quantity, Date or Time response data may also be interpreted as Literal response data.

- direct from the respondent – as would be the case where the pattern is validating data during an interview.
- from a data store – as would be the case for data already keyed from paper questionnaires or stored from interviews conducted before.

In both cases, the pattern works in the same way: it takes input (respondent) data, passes it through its own lexical matches, rewrite rules and value constraints then, if acceptable, passes it on to the appropriate in-built validation associated with the response type of the variable. The process is illustrated in Figure 5.



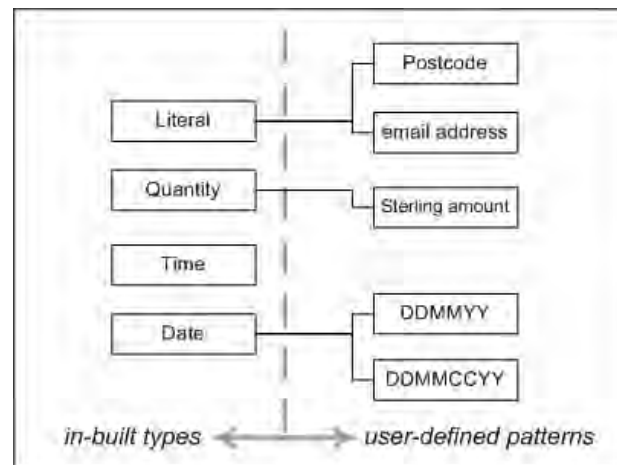
**Figure 5 – Processing respondent input through pattern and in-built validation**

Since the processing done by patterns and the in-built validation can be expensive in terms of CPU time, the results are cached by the program and held in an inverted datastore (that is, processed results are held by question whereas raw data is held by respondent.). Any changes to the raw data results in the automatic re-processing of the data. The result is that analysis tables and charts are always accurate and fast to build.

An important point to note is that the raw data records the exact respondent input, not transformations of it. Thus there is always a direct and documented audit trail from respondent input to analysis results. In addition, the original data is still available to be processed by other variable / pattern combinations to provide different interpretations of individual data entries if deemed appropriate and useful. This contrasts with the traditional method when keying data from paper questionnaires using data entry operators to "clean" the data, where the audit trail between the paper copy and the electronic copy of respondent input is lost.

## 7. Patterns as types

By connecting a pattern to an in-built response type, the user is essentially *specialising* those types.



**Figure 6 – Specialising in-built response types**

Those specialisations are of three forms: those that apply format restrictions; those that apply format extensions; and those that apply value restrictions. Sometimes these are applied singly, sometimes in combination. For examples, consider the following:

- Format restrictions – for example, an **integer** pattern based on the in-built **quantity** pattern would restrict recognition to values with no decimal point allowed.
- Format extensions – for example, a **worded number** pattern based on the in-built **quantity** pattern (allowing respondents to enter `two` to mean 2, for example.)
- Value restrictions – for example, a **human age** pattern based on quantity

As we have seen (Figure 5), we can view the pattern as a small processing machine which, when given an acceptable input string, emits a value of the connected in-built response type at the back end. Thus, an acceptable input to the pattern becomes an acceptable input to an in-built response type. The user has created an IS\_A relationship between the pattern and the associated in-built type. That is, for example, a **worded number** (as outlined above) IS\_A **quantity**.

Now, the in-built types have in-built operations available. For example one quantity can be added to another to derive a quantity representing the sum of the other two. The nature of IS\_A relationship is that all operations available on the in-built type are available for the specialised type represented by the pattern. Thus the in-built operators, operating on the in-built types can be considered to be inherited by the patterns. So, for example, the response to a question built using the worded integer pattern can be added to, or subtracted from, the response to another suitable question.

So, by composing patterns in the way described, the user is extending the type-space of the underlying program.

### New operators for free

One useful side-effect of the use of named components to construct a pattern is that those components become available for interrogation. For example, a UK postcode pattern might be created from four components each representing the standard Royal Mail terminology (Table 2.)

The final match pattern based on these components would then be:

```
{area}{district} {sector}{unit}
```

Component	Lexical match	Meaning
area	one or two letters	Sorting office
district	one or two digits, or one digit and one letter	Delivery office
sector	one digit	Local area
unit	two letters	Postal route

Table 2 – UK Postcode components

Now, if a question, say Q5, was assigned this pattern, it would be possible to interrogate the *components* of a response. So the expression Q5 `area` gives up the area part of a complete postcode response, Q5 `district` gives up the district part, and so on.

The components can be considered to be unary operators, each of which divulges a part of a complete response. Thus patterns composed in this way can be considered to be types in their own right describing, as they do, allowable data values *and operations on them*.

### Response box formatters for free

A second useful side-effect where named components are used is the possibility of drafting complex formatted input boxes from the pattern specification. For example, given the earlier DDMYY pattern, but allowing `{day}/{month}/{year}` as the only possible match, it becomes feasible to publish response boxes questionnaires which look like this for a paper questionnaire:

or like this for a web questionnaire:

This formatting possibility comes absolutely for free in that the user is not required to participate at all – all of the required information is already present in the definition of the pattern.

### Further work

There are a number of extensions we are currently exploring:

- Patterns are currently specialisations of in-built types only. A simple extension would be to allow users to create patterns as specialisations of other patterns (which themselves would ultimately be specialisations of the in-built types). This essentially enables the user to build unbounded trees of types.
- Patterns are currently implemented for open questions but it would be appropriate to consider their use in closed (single- and multiple-response) questions as well.



- It should be possible to allow the user to select *one* of the possibly many lexical matches to use as a default format for reporting response data. This would allow, for example, a pattern to be built that accepts dates in a variety of formats but always reports them in dd/mm/ccyy format.

## 8. Conclusions

The original motivation for this work was an observation that responses to paper questionnaires give an insight into the ways that respondents perceive questions and the way that they think they are expected to respond (in terms of formatting their responses). That was allied to a personal view that many forms-based web applications were (and are) appallingly written in the sense that they apply onerous restrictions to formatting with little guidance as to the expected format of particular data fields.

The solution uses the well-researched and established technique of regular expression patterns for parsing input. An established (albeit a canonical rather than formal) standard as the basis means that translations for external platforms are possible. The main drawback for end-user applications is the arcane syntax of regular expressions. This problem has been overcome in two ways: firstly by building in patterns for common requirements thus removing the necessity for users to get involved in the low-level at all; and secondly by representing patterns as rich objects with identified properties which when assigned appropriate values behave in the same way as their tersely-specified relatives.

The resultant solution has produced some serendipitous side-products that enable patterns to be perceived as genuine response-type builders complete with values constraints and appropriate operations such that they are useful not only at data collection time but equally so at analysis time as well.

## References

"Regular Expressions", wikipedia article, [http://en.wikipedia.org/wiki/Regular\\_expression](http://en.wikipedia.org/wiki/Regular_expression)

"Postcodes and addresses explained", Royal Mail website,  
<http://www.royalmail.com/portal/rm/content1?catId=400120&mediaId=9200078>

"Mastering Regular Expressions", Friedl, J, O'Reilly and Associates, 2006

## About the Author

Dr Stephen Jenkins is Technical Director of Snap Surveys Ltd. and is responsible for the direction and development of Snap survey software. He has over 30 years' experience in the design, development and use of survey software. He is a founding member of The Triple-S Group and an author of the Triple-S survey interchange standard. He can be contacted at Snap Surveys Ltd., Mead Court, Thornbury, Bristol BS35 3UW, UK, or by email [sjenkins@snapsurveys.com](mailto:sjenkins@snapsurveys.com).

# Software Licence Agreements: Just what are you agreeing to when you press the “Accept” button?

Paul Sampson & Peter Wills

## Abstract

Software users merrily click away at the 'Accept' buttons as they install new software and are presented with the terms of the Software Licence Agreement, but few understand why they do it or what they are agreeing to in these legally binding agreements. This paper reviews the current situation and the pitfalls that exist both for the software users and the software authors.

## Keywords

EULA; software license agreements; warranty; copyright infringement

## 1. How have we arrived at the current situation?

Most software is intended for a mass market, so the publisher retains the legal rights in it and gives each user permission to use it; in legal jargon, the publisher licenses it to each user. That permission comes as part of something bigger - the End User Licence Agreement (EULA).

Microsoft XP<sup>1</sup>: *Reservation Of Rights And Ownership. Microsoft reserves all rights not expressly granted to you in this licence agreement... The Software is licensed, not sold.*

But the same reasoning could be applied to books, and yet no one visits a bookshop to license a book. Book tokens and till receipts do not come with licence agreements printed on the back. The customer simply buys his book like a packet of cornflakes. (We are referring here to traditional books printed on paper, not e-books, which are more like software.) So why are software users presented with licence agreements?

The difference flows from an intersection of copyright law and technology. It is an infringement of the copyright in any copyright work, whether a book or a piece of software, to make a copy of it without the copyright owner's permission. But while a book can be read without making a copy of it (ignoring any copy that may be made in the reader's brain), software cannot be put to any practical use without making at least one copy of it on the user's computer. So while a book publisher cannot control who reads a legitimate copy of a book, a software publisher can control who may use a legitimate copy of a program after it has been sold, and how they may use it.

<sup>1</sup> The terms quoted from licence agreements as illustrations may not be the current version.

Some argue that this difference is purely fortuitous for software publishers, and that such post-sale control is not legitimate, but the position is now long established and it is far too late for such debates. This paper discusses today's reality rather than any Utopia.

## 2. Copyright infringement – Is it anything to worry about?

Copyright infringement is a serious matter. Firstly, it is a criminal offence on a par with theft. Prosecutions may be brought by the authorities or by the publisher. An individual can be prosecuted as well as the company he works for.

Secondly, the publisher is entitled to compensation. The starting point when assessing this compensation is the licence fees he ought to have paid. This may be calculated on the basis of the full individual licence fee, as the sizeable reductions which may be available when buying multiple copies of a program often only apply to licences which are bought before the software is used. This figure may then be increased to punish the user.

Finally, the user will have to pay the publisher's legal costs of taking him to court, which in many cases will equal or exceed the compensation.

Unlicensed users face a significant risk of being caught, especially when the software itself tips off the publisher, as and when that computer is connected to the internet. The major software companies fund organisations like the Business Software Alliance, whose purpose in life is to identify infringers, bring prosecutions and obtain money.

Criminal prosecutions are relatively rare in practice, but this may be because the prospect of a prosecution often encourages an infringer to settle the publisher's financial claim on generous terms.

## 3. So what's wrong with licence agreements?

As the core purpose of a licence agreement is to give the user permission to use what he has paid for, it is worth examining what publishers actually put in their licence agreements.

Firstly, software publishers seldom give a simple permission. Instead, they permit use of the software in very specific ways, leaving any other use an infringement of copyright. Then they attach all manner of conditions which have little to do with the core licence but which the user had better observe as any failure to do so almost invariably brings the licence to an end, leaving the user with an empty wallet and software he cannot use.

So what sort of restrictions and conditions can be found in licence agreements, and are they reasonable?

Let us return for a moment to the analogy of buying a book, and the everyday transaction of a man who buys a business related book in London, where he works. He plans to start reading it at home and to finish it on a forthcoming business trip to America. Upon his return he will give it to his colleague to read before either putting it in the company library, handing it in at an Oxfam shop, or selling it to a second-hand book shop.

Now imagine that after he has paid for the book at the till, he is presented with a very long receipt. On the back, under the heading 'licence agreement', it says that:

1. He will be bound by all this small print if he merely opens the book.

2. The book's content may not match the description on the cover, the print might be unreadable and some pages may be missing, but he cannot do anything about it even though he could not reasonably be expected to have discovered any of these things until after he had opened the book.
3. He must pay extra if he wants to read the book while wearing reading glasses, as these boost the performance of his eyes.
4. He must pay extra when he takes the book abroad.
5. He is not allowed to lend it to his colleague or to sell it or give it away when he has finished with it, and he will have to pay for a second copy if someone else reads it.
6. He must keep a written record of where it is at all times.
7. If he subsequently buys a second edition, after the publisher has corrected the errors and typos in the first, he must burn the first.
8. The publisher is entitled to enter his office and home to see if he has complied with all this.
9. Until the publisher has corrected the errors, and removed any bits which have been plagiarised following complaints from other publishers, it is entitled to tear pages out of his book.
10. All this applies to any other books from that publisher which he might read – even if it is someone else's book in someone else's house.
11. The publisher can change these terms whenever it wishes, without even telling him.
12. The publisher's decision is always final.
13. The law of California applies.

This is so ludicrous as to be almost unimaginable, yet this is just what happens when people buy software and click those 'Accept' buttons to agree to the licence agreement.

#### 4. The word “Grant” appears in many licence agreement....

This sounds formal, almost official. It sounds as if the user is being given something special and valuable which he did not have before. In practice, it usually conceals the fact that he is being deprived of something which he probably thought he had paid for.

For example, the licence is often limited to specific individuals (e.g. the purchaser) or specific equipment, which prevents the user sharing the software with a colleague:

**Autodesk:** *Autodesk grants You a non-sublicensable, non-exclusive, non-transferable, limited license to use copies of the Software ...*

**Skype:** *Subject to the terms of this Agreement, Skype hereby grants You a limited, personal, non-commercial, non-exclusive, non-sublicensable, non-assignable, free of charge license to download, install and use the Skype Software on Your computer, phone or PDA.*

**Microsoft XP:** *You may not rent, lease, lend or provide commercial hosting services with the Software.*

The user may only be allowed to use the software where he bought it:

**Autodesk:** *Autodesk grants You a non-sublicensable, non-exclusive, non-transferable, limited license to use copies of the Software in the jurisdiction in which you acquire the Software ...*

So he cannot use the software while in America on business if he buys it in England.

Buying an upgrade may prevent the user from using the software he already has, even if it is not satisfactory:

**Microsoft XP:** *After upgrading, you may no longer use the software that formed the basis for your upgrade eligibility.*

Although the licence might not be as extensive as the user might wish in many ways, in others it may be more extensive than the user realises. In particular, it might cover more software than the user thought he was buying. In one well known instance, the product licence agreement covered code which monitored the user's web browsing activity.

In this example:

**Skype:** *This End User License Agreement constitutes a valid and binding agreement between Skype Software S.a.r.l and You, as a user, for the use of the Skype Software ... You hereby agree and acknowledge that this Agreement covers all Your use of Skype Software, whether it be from this installation or from any other terminals where Skype Software has been installed, by You or by third parties.*

The user agrees that he is subject to all the conditions, restrictions and obligations in the licence agreement not just in connection with his own Skype terminal but also when he uses someone else's. This is equivalent to saying that the small print on the back of a till receipt for a book apply to any of the publisher's books which the customer might read, even those he borrows from a friend.

Many licence agreements include reverse licences - i.e. the customer gives the publisher permission to do things without warning him:

- Permission to automatically update the software

**Microsoft XP:** *The Software contains components that enable and facilitate the use of certain Internet-based services. You acknowledge and agree that Microsoft may automatically check the version of the Software and/or its components that you are utilizing and may provide upgrades or fixes to the Software that will be automatically downloaded to your Workstation Computer.*

- Permission to extract information from the customer's computer

**Microsoft XP:** *You agree that Microsoft and its affiliates may collect and use technical information gathered as part of the product support services ... Microsoft may use this information solely to improve our products or to provide customized services or technologies to you and will not disclose this information in a form that personally identifies you.*

This allows Microsoft and others to gather information about the user and to distribute that information to other unnamed persons. It does not say that either Microsoft or its affiliates

will not collect information in a form that personally identifies the user, or that they will not use it in that form. It does not even say that Microsoft's affiliates will not pass on information in a form that personally identifies the user.

- Permission to enter the customer's premises:

**Macromedia:** *You agree that Macromedia may audit your use of the Software for compliance with these terms at any time, upon reasonable notice. In the event that such audit reveals any use of the Software by you other than in full compliance with the terms of this Agreement, you shall reimburse Macromedia for all reasonable expenses related to such audit in addition to any other liabilities you may incur as a result of such non-compliance.*

**Autodesk:** *To ensure compliance with this Agreement, You agree that upon reasonable notice, Autodesk or Autodesk's authorized representative shall have the right to inspect and audit Your Installation, Access and use of the Software. Any such inspection or audit shall be conducted during regular business hours at Your facilities or electronically.*

This should put paid to home workers working in their pyjamas!

## 5. Just what does the word "Warranty" mean?

Anyone who buys anything from anyone usually expects certain assurances; that the seller has the right to sell it, that the user will be able to use it, that the item supplied is as the seller described, and that it does what the seller claimed and does it reasonably well. Lawyers call these assurances 'warranties', and the courts will order the seller to compensate the buyer for any breaches of warranty. These particular warranties sound reasonable; so much so that Parliament passed laws which add them to many contracts if they have not been spelt out. But they are obviously too much for some publishers:

**Skype:** *No warranties. The Skype software is provided "as is", with no warranties whatsoever; skype does not, either expressed, implied or statutory, make any warranties, claims or representations with respect to the skype software, including, without limitation, warranties of quality, performance, non-infringement, merchantability, or fitness for use or a particular purpose.*

**Remedy.** *Your only right or remedy with respect to any problems or dissatisfaction with the skype software is to deinstall and cease use of such skype software.*

If this applied to a book, the content may not match the cover, the print may be unreadable, some pages may be missing, and what is there and readable might be an illegal copy of another publisher's book, but all the customer can do is stop reading it and throw it away. He should not expect an exchange or a refund.

**Autodesk:** *Limited Warranty. Autodesk warrants that, as of the date on which the Software is delivered by Autodesk and for ninety (90) days thereafter, the Software will provide the features and functions generally described in the User Documentation and that the media on which the*

*Software is furnished, if any, will be free from defects in materials and workmanship.*

Most people would expect a book to last for many years without the story changing or the print fading away or falling off the page, but software users should have lower expectations.

As well as giving very limited warranties, or even none at all, some licence agreements include reverse warranties. In other words, the user becomes obliged to compensate the publisher.

## 6. Special Conditions may apply.....

These may go beyond preventing the user from abusing the publisher's rights. In the well known instance mentioned above, where the product licence agreement allowed the publisher to include code which monitored the user's web browsing activity, a condition prohibited the user from using anti-spyware software to remove this code.

Publishers enforce these conditions by giving themselves the right to terminate the user's licence if he breaches a condition:

**Microsoft XP:** *Without prejudice to any other rights, Microsoft may terminate this agreement if you fail to comply with the terms and conditions of this agreement. In such event, you must destroy all copies of the Software and all of its component parts.*

The publisher's decision may be final:

**Skype:** *...You acknowledge and agree that Skype, in its sole discretion, may modify or discontinue or suspend Your ability to use any version of the Skype Software, and/or disable any Skype Software You may already have accessed or installed without any notice to You ... where You, at Skype's discretion, are in breach of this Agreement ....*

Some rather extreme licence agreements turn all use of the software after a breach of the licence agreement into a criminal offence – even if the breach is trivial and inadvertent, and even if the user is unaware of it:

**Autodesk:** *Autodesk's license grant is conditioned on Your continuous compliance with all license limitations and restrictions described in this Agreement. If You violate any of these limitations or restrictions, the license grant will automatically and immediately expire.*

**Mozilla Firefox:** *If you breach this Agreement your right to use the Product will terminate immediately and without notice, but all provisions of this Agreement except the License Grant (Paragraph 1) will survive termination and continue in effect. Upon termination, you must destroy all copies of the Product.*

## 7. Just which licence agreement are you signing up for?

As if grappling with the complexities of one licence agreement were not enough, a user may be faced with several:

**Redhat Linux:** *Red Hat Enterprise Linux is a modular operating system consisting of hundreds of software components. The end user license agreement for each component is located in the component's source code.*

Some publishers have included an evolutionary term in their licence agreement, so that the entire licence agreement is automatically superseded by any new version of it which the publisher displays on its website at any time in the future, whether or not the user has ever seen the new licence agreement or is even aware of its existence.

**Skype:** *Furthermore, by installing and continuing to use the Skype Software You agree to be bound by the terms of this Agreement and any new versions hereof... Skype reserves the right to modify this Agreement at any time ... by publishing the revised Agreement on the Skype Website. The revised Agreement shall become effective within thirty (30) days of such publishing ...*

So the publisher can change the rules whenever it chooses.

The courts would undoubtedly find this practice objectionable, but it is not clear that they would refuse to allow the publisher to rely on the new licence agreement terms.

## 8. Can publishers always rely on their own licence agreements?

A licence agreement's terms do not have any legal effect unless they form part of a legal agreement between the publisher and the user. At the risk of stating the obvious, this raises the question of which law applies:

**Macromedia:** *This licence agreement shall be governed by the internal laws of the State of California, without giving effect to principles of conflict of laws. You hereby consent to the exclusive jurisdiction and venue of the state courts sitting in San Francisco County, California or the federal courts in the Northern District of California to resolve any disputes arising under this licence agreement. In each case this licence agreement shall be construed and enforced without regard to the United Nations Convention on the International Sale of Goods.*

This is equivalent to a customer buying a selection of business books from a high street bookshop only to find that his right to read each book is subject to the law of a different country. For simplicity, this paper assumes that English law applies.

Most publishers use a software installation process which requires the user to see the licence agreement and to click an 'accept' button before the software will install or activate, but some go further:

**Microsoft XP:** *You agree to be bound by the terms of this licence agreement by installing, copying, or otherwise using the software.*

The bookshop equivalent is being given a till receipt bearing small print which says that the purchaser is bound by that small print if he opens the book.

The problem for the bookseller would be how to ensure that anyone else coming into possession of the book would also be bound by the small print on the till receipt. Software publishers – or rather their clever lawyers – have the answer:



**Microsoft XP:** *The initial user of the Software may make a one-time permanent transfer of this licence agreement and Software to another end user, provided the initial user retains no copies of the Software ... Prior to the transfer, the end user receiving the Software must agree to all the licence agreement terms.*

This is to ensure that the licence agreement goes wherever the software goes, even if the new owner of the computer has not had to click the 'Accept' button. If he tries denying that he is bound by the licence agreement – which he probably could under contract law – then his licence evaporates and his further use of the software is an infringement of copyright. That should stop him taking the point.

A licence agreement is usually introduced after the software has been supplied and paid for, and under the normal principles of contract law this is too late for any part of the licence agreement to have any legal effect. But the position is complicated by the copyright point; a purchaser may be free to keep the copy he has bought unaffected by anything in the licence agreement, but he needs a licence from the publisher before he can use it and that licence, which is in the licence agreement, is subject to all the licence agreement's other terms:

**Microsoft XP: GRANT OF LICENSE.** *Microsoft grants you the following rights provided that you comply with all terms and conditions of this licence agreement.*

So the user would be unwise to say the licence agreement is of no effect at all. But a court which disliked some of the licence agreement's more onerous terms could give effect to this as meaning all those terms which form part of the legal agreement, rather than all those printed in the licence agreement. This then raises the question, 'Which of the licence agreement's terms form part of the legal agreement?'

If a user physically signs a printed licence agreement, almost any term in it will then be part of the agreement. In the more usual situation, where the user clicks an 'Accept' button, whether any particular term forms part of the agreement will depend on various factors, including the nature of the particular term and whether the publisher has taken sufficient steps to alert the user to the term. A draconian term hidden away at the end of six pages of small print will almost certainly not be part of the agreement. It is not clear whether the increasingly frequent practice of requiring the user to complete his name before clicking the 'Accept' button constitutes a signature for these purposes.

Even if a particular term in a licence agreement is part of the user's legal agreement with the publisher, it does not automatically follow that the publisher can enforce it: legislation like the Unfair Contract Terms Act 1977 and the Unfair Terms in Consumer Contracts Regulations 1999 may prevent it from doing so. Terms which may be affected include those which allow the publisher to supply something different to what the user reasonably expected; restrict certain statutory warranties; restrict the publisher's liability to compensate the user; oblige the user to compensate the publisher.

Consumers receive special protection, but even Microsoft receives protection against other publishers' licence agreements. For commercial users, the reasonableness of the term is often the key issue: was it reasonable for the publisher to include that term in that particular agreement with that particular user.

The concept of reasonableness is very wide, and the court is specifically required to take account of all the relevant circumstances. Nevertheless, the courts have developed a reasonably consistent approach to this issue when it comes before them and some factors have begun to emerge as especially important. For example, it is very unlikely that a publisher would be able to rely on a complete exclusion of its liability to compensate a user when it is a substantial organisation or it could obtain insurance to cover its liabilities.

It is up to a publisher to ensure that it stays within the bounds of reasonableness: a court will not substitute a more reasonable term for an unreasonable one. So a sensible publisher should voluntarily accept a higher level of liability which it is more likely to be able to rely on rather than impose too low a level which it cannot rely on. This is an important point which many publishers do not appear to appreciate, but it is not easy to achieve the right balance and, perhaps more than anything else in software licensing, this is an issue on which specialist legal advice is essential.

## 9. Conclusion

The very existence of licence agreements is the result of the odd interaction between computer technology and copyright law. They developed when computers were generally seen as something new and special, which might explain why users allowed publishers to include the terms we see today. There were very few specialist lawyers then, so licence agreements tended to evolve without a great deal of legal analysis and negotiation by users. By the time all this changed, and disputes started to come before the courts, many of the terms in today's licence agreements had become commonplace. In itself this meant that they were more likely to survive the courts' assessment of their reasonableness, but more recently the courts have shown themselves willing to strike terms down as unreasonable, and intelligent publishers should react to this by revising their licence agreements.

Users who find themselves in dispute with a publisher may be able to avoid some of the worse excesses of licence agreements, but this may be an expensive exercise with no guarantee of the right outcome. A user should certainly not see this as an alternative to reviewing licence agreements before accepting them. Admittedly there is probably very little scope for negotiating changes to the licence agreement for most off-the-shelf software, but it is possible with more expensive packages. As for the rest, there are always two options in the 'take-it-or-leave-it' situation, and users should seriously consider exercising the second option on occasions.

## About the Authors

**Paul Sampson** leads the technology law team at the law firm Veale Wasbrough. He has specialised in IT contracts and disputes for 16 years. He is a chartered engineer as well as a solicitor, and is recognised for his experience on complex projects. He can be contacted at Veale Wasbrough, Orchard Court, Orchard Lane, Bristol BS1 5WS, tel:+44 117 925 2020 Fax: +44 117 925 2025; e-mail [psampson@vwl.co.uk](mailto:psampson@vwl.co.uk).

**Peter Wills** is the Chief Executive and Chairman of Snap Surveys Ltd, a company set up over 25 years ago. He has been involved with both the development and marketing of Snap software since the beginning and has recently been responsible for the lengthy process of

producing a new software license for Snap. He can be contacted at Snap Surveys Ltd, 5 Mead Court, Thornbury, Bristol BS35 3UW, tel. +44 1454 280800; fax +44 1454 281216; e-mail [pwills@snapsurveys.com](mailto:pwills@snapsurveys.com).