Softly, Softly Catch the Monkey: innovative methodologies to measure socially sensitive and complex issues in evaluation research

Anne Sharp and Catherine Eddy
Marketing Science Centre
University of South Australia
GPO Box 2471
Adelaide
South Australia 5001

All contact should be directed to:
Anne.sharp@msc.unisa.edu.au
Tel: +64 8 8302 0657
Fax: +64 8 8302 0123

Abstract

Many Government programme evaluations require information to be captured that is hard to measure, of a sensitive nature and difficult for the respondent to articulate. Topics such as household financial decision making, inter-family views on unemployment or equity, and the impact of problem behaviours, such as gambling, on other family members, are all examples of where such issues are faced.

This paper suggests research designs and methodologies to assist in overcoming such problems in evaluation research. Our discussion is illustrated by three evaluation case studies, undertaken on behalf of a State Government regulatory body and a Federal Government department providing family assistance payments, where we have implemented such approaches. The research design suggestions focus on increasing intersubjective certifiability through the use of triangulated respondent groups, as well as varying the research team composition at different stages of the research. Methodological suggestions are for multi-faceted research processes, run in parallel and in sequence, to uncover topics on which findings vary or to find information ‘hidden’ in other approaches. Methods for improving recruitment and retention of respondents are also made. We conclude by critically evaluating the outcomes of having applied these new approaches and discuss the implications of gaining different or new information from having adopted such innovative approaches.
Introduction

Evaluation research often calls for measurement of constructs that are more complex and difficult for respondents to articulate than in the case of other more ‘conventional’ research. This is because the issues being evaluated are often to do with social dynamics in a family, or are sensitive in nature, such as respondents frequency of undertaking a pap smear when trying to evaluate a cervical cancer programme. In such circumstances, there is increased reliance on the researcher to interpret what the respondent is trying to articulate, to be able to observe body language, or other such subtle dynamics, in order to arrive at research findings. The researcher’s skills in research design are critical as some of the sensitive or hidden issues (eg a youth’s degree of shared values with his/her parents) can not be fully explored with all research approaches.

Because of these factors, we need to be extremely careful to avoid introducing bias into the research either through the researchers, respondents we select, or the research design. While the topic of bias introduced by respondents has received attention in the literature (Interviewer 1998) by contrast, researchers and research design have received far less attention (Herschell 1999). Yet these issues are all equally important (Welch 1985). In this paper we illustrate the benefits for evaluation research of integrating intersubjective certifiability more deeply into research practice. In particular, we focus on these concepts in the context of research design and methodology, as well as respondent selection in the evaluation process.

Intersubjective Certifiability, Triangulation and Hidden Behaviour

One characteristic that distinguishes the scientific process of knowledge generation from that of other means such as intuition or guesswork, is intersubjective certifiability (Hunt 1991, Hunt 1976). When something is said to be intersubjectively certifiable, we mean that two or more researchers or processes will draw similar results and interpretations after exposure to the same material. As researchers and good marketing scientists, we strive for objectivity in that our statements are capable of public tests with results that do not vary essentially with the tester. Improvements in research practice to achieve such objectives have been called for by numerous authors (Gabriel 1990, Herschell 1999).

This concept of intersubjective certifiability is most often implemented through researcher, respondent, or methodological triangulation. That is, multiple researchers, respondents or methodologies are used and the results from the various people and/or approaches are compared (Wallendorf and Brucks 1993). Where the same results are arrived at then the research is said to be intersubjectively certifiable. Where it is not, then further work needs to be undertaken to determine reasons for the disparity.

This idea of intersubjective certifiability is extremely useful in instances where the research issues can be articulated and easily measured. However, in programme evaluation research, there are sometimes areas where respondent’s behaviours and attitudes are hidden. For example, when interviewing a couple in a partnership, feelings of equity and role sharing in the relationship may not be as fully disclosed as when the couple are interviewed separately, with no feedback to the other partner as to their responses. This variation in what the respondent is willing, prepared, or can remember to say, impacts upon our findings. Potential for such hidden behaviour must also be taken into account when determining the respondent groups, research designs and methodologies to be employed in programme evaluations. Some groups of interest may well be unsuited to different methodologies than others. For example, indigenous populations living in traditional settings may find a focus group methodology very intrusive and a face to face methodology may be more appropriate. However in exploring another subgroup (eg. Job seekers) it may assist the researcher by having more than one respondent to raise issues of interest and to
assess corroboration or otherwise from different members of the group. Having said this, sometimes a group may appear to be similar when in fact there are differences that are not immediately obvious and may arise in the context of the focus group. Their willingness or preparedness to disclose information may be largely determined by the research setting.

The Evaluation Case Studies

The first case involved a longitudinal evaluation of a Federal Government family assistance programme. This programme provided financial payments to the main carer in two parent households with dependent children, where the main carer was not engaged in full time work. The programme objective was to provide recognition for the role of caring through the payment and to increase equity within the family through this. It also sought to increase choice for families as to the balance between the role of caring and engaging in paid work. The evaluation required information to be obtained from both partners in a family about issues such as financial management in the household, perceptions of role swapping and role sharing, as well as perceived barriers to achieving the respondent’s desired level of workforce participation. The evaluation took place over three years, with the same 1300 respondents, where possible, being recontacted in each of the research phases. The first phase of the evaluation had a significant qualitative component where respondents were interviewed in their home. This was followed by telephone interviews in subsequent phases.

The second evaluation also involved a Federal Government family assistance evaluation, but in a cross sectional context. The research sought to determine how families choose between various family assistance payment options that span the taxation and welfare systems. The likelihood of future behaviour was measured and also the responses to various penalty and reward scenarios for the different options from which respondents could choose. This research required an overall measure of family decision-making and so both partners (if applicable) in the family were interviewed. Again, the evaluation had a significant indepth interview component before conducting telephone interviews with families nationally.

The third case study involved another longitudinal evaluation, again over three years. The evaluation was conducted for a State Government regulatory body. The objective was not to evaluate any programme or intervention, but rather to understand how respondents may form perceptions of gambling and an understanding of the gambling industry, and how their perceptions and behaviours change over time. By its nature, this research involved many sensitive and, at times, hard to measure issues such as problem gambling, expenditure levels on gambling and perceived social, economic and other impacts of gambling. This evaluation also had a significant qualitative component. It consisted of focus groups of like respondents, split by age, gambling frequency and the respondents stated attitude to gambling (positive or negative).

These case studies all posed the challenges previously discussed and we use them in the following discussion to illustrate approaches to dealing with these issues.

Respondents, Research Design and Methodology

Methods for both improving intersubjective certifiability and maximising findings accessible to the researcher are discussed under three broad headings; the researchers, the research design, and the respondents.

The Researchers

The significance of who is doing the research often seems to be overlooked in determining likely sources of
research bias. It is seen as an invisible, unquantifiable and unmeasurable aspect of the research. Typically, it has been assumed that the more experienced the researcher is about the topic at hand, the better the research outcome. Evaluation tenders are often assessed with a component given to expertise and experience in the topic area. But, this might not be ideal. A researcher may become jaded or even presuppose the opinions of research respondents. Even when they attempt to be neutral and not bias or influence the respondent, their experience and intuition are crucial in interpreting the evidence and making sense of what is discovered. Their interpretation is, to an extent, dependent on whatever experience the researcher brings to bear, and so it may be idiosyncratic (Colwell 1990, Peter and Olson 1983). While the inter-rater reliability research stream offers some insights into the reliability of human observers’ judgements, this literature is more about measuring such error rather than controlling for it. The exploratory work that has been done, however, suggests that different researchers may produce quite different research findings (Twyman 1996)

The use of multiple researchers who hold varying perspective on the topic under investigation should assist in increasing the reliability of findings. We have found that using multiple researchers, especially in the early stages of the research, promotes richness in developing the key topics for discussion and in the outcomes of the research process. For example, in the family assistance work, we have been careful to use a research team that consists of single people, partnered people with no children, and also partnered researchers with children. It follows logically that intersubjective certifiability is increased with the use of several, rather than a single, researcher. By disclosing personal experience and information at the beginning of the research process, other researchers in the team can identify biases that might occur. For example, in the two family assistance case studies, the fact that some of the research team were current recipients of the payments being researched and evaluated meant there was a whole area of experience that they brought to the research that the other research team members lacked. This experience was extremely useful in conducting the interviews and designing the research instrument, as they were familiar with terminologies and the experience of being a payment recipient. They could establish a much greater level of empathetic understanding with respondents than the other researchers. However, by having some team members with no prior payment experience, we minimised the risk of the research being overly coloured by the experiences of the researchers who also met the criteria to be respondents.

As an extension to this practice, at each new research phase of the longitudinal cases, we ensured that at least one team member had no prior involvement in the project. This was done to ensure that the results of the current phase were not unduly influenced by exposure to findings from the previous stages of the research. The same practice should be applied when multiple research projects are conducted in the same topic area. It is impossible for a researcher not to bring assumptions about the likely results of new research to a project when they have experience with similar projects in the past. Clearly their experience is useful, but this should be balanced by input from other researchers who with no such previous topic orientation.

Having such prior experiences disclosed may also assist readers of the results to understand varying perspective coming into the research. For example, in the gambling case study, the focus groups were run according to the respondent’s stated attitude to gambling. A team of two researchers moderated the four focus groups. The first was a light gambler with a positive attitude and the second was a non-gambler with a negative attitude to the category. This was explained to the research sponsor. The non-gambler moderator spent a greater proportion of time on topics that related to the problems associated with gambling than did the moderator who gambled. The moderators tended to steer the discussion into areas where they could share feelings with the group participants. This was needed as the topic was one that was quite sensitive to many respondents and a high degree of trust and rapport was needed before respondents would be open about their own behaviour and perceptions. These moderator biases were explained in the client’s research report. It provided for a more open and diverse set of research findings on which to base the quantitative instrument design.
To conclude, using such a team approach in research is useful “to enhance access through triangulation across potentially diverse perspective’s” (Groves and Belk 1994) and “re-energise the team in the tiring process of fieldwork” (Wallendorf and Belk 1989).

The use of multiple researchers offers a number of options in terms of design, which we now outline.

**Research Design and Methodology**

Intersubjective certifiability can also be influenced by the research process and methodology selected. In terms of process, if the individual researchers can record their own interpretations before discussions with other researchers then "this makes triangulations across sets of notes more legitimate by not altering each other's recall or understanding of what was said or done (Wallendorf and Belk 1989). The various ways in which we suggest this can be done in evaluation research are as follows:

- At the start of in depth interviews, send researchers out in teams of two to identify interviewers’ preconceptions and thus improve intersubjective certifiability through researcher triangulation. One researcher conducts the interview while the second observes. A debrief follows where the researchers compare their interpretation of the interview. Any areas of non-consensus are identified and a strategy developed to resolve the issue. A new team of two is then put together (or the researchers swap the interviewer/observer role) and the process is repeated. This approach was adopted in both the family assistance qualitative phases. We found that respondents were comfortable having two researchers present as long as it was clearly explained that we were at the start of the interviewing process and that one would lead most of the discussion with the other coming in only if an issue was missed or required elaboration. In some instances, the leader was the researcher who most closely matched the respondent in age and family situation. In others, this was deliberately reversed. The process of having two researchers in an indepth interview also makes this detection of body language and observation of the respondent’s surroundings much easier, resulting in richer research findings.

- Similarly, in a focus group situation, one researcher moderates and the second observes (either remotely or in the same room). The researchers write up their conclusions independently and then compare results. Areas of non-consensus are clearly identified and resolved. The post-viewing by the moderator is also recommended as a means of uncovering other responses – perhaps body language indicating disagreement when verbalising agreement or physical expressions of discomfort when certain topics are raised. In the gambling research, this approach was especially useful. Two researchers viewed the tape of the focus groups before writing the quantitative instrument. They had conflicting perceptions about the degree to which respondents had stated that gambling perceptions were formed. Acknowledgment of this disagreement and a careful reviewing of the tapes resolved the issue. Without such a process, undue bias may have been introduced into the research findings.

The chosen research methodology can also influence intersubjective certifiability. It is possible that different findings can be obtained in a focus group situation compared to, say, indepth interviews with the same respondents. Differences may come about if issues are considered private or are prone to peer bias by respondents. In such instances, differing responses would emerge in a focus group situation than in an indepth interview and so the results may overstate/understate certain findings. Where possible, we suggest using a combination of qualitative methodologies to minimise the risk of such systemic bias. Our suggestions for evaluation research are:

- When little is known about a research issue, or respondents are likely to find the topic difficult to discuss (due
to lack of knowledge or involvement, it is useful to start the research process with a focus group or some other situation where respondents can interact. This allows for the snowballing of ideas and the identification of "public" opinions. It also clarifies what thoughts and feelings respondents are comfortable to express in a group setting, with people they do not know. This can then be replicated in a family setting (or other nuclear group the research question relates to). A focus group of family members may identify if there are issues that respondents are willing (or unwilling) to state in front of people they do not know but will (or will not) state in front of people they are familiar with. Hidden issues can be further identified by following (or preceeding) such group work with individual, confidential, indepth interviews.

We used this process in two different ways in our evaluation work. Firstly, in the gambling research, we chose a focus group methodology as we felt that one-on-one interviewing would be too confronting for the respondents, especially when they were talking about their own gambling behaviour. Previous work in the area had shown that respondents tended to understate their own participation in gambling when interviewed individually {Roy Morgan Research, 1999 #6483}. By structuring the focus groups according to stated frequency (even if it was under-reported) we were able to get groups of respondents together who were able to talk openly about gambling and were more relaxed as they were with 'peers'.

In the evaluation projects, we adopted a different approach. When we interviewed both partners together, we found that one tended to dominate the interview. This was usually the main caregiver, who was also the payment recipient and so had a higher level of knowledge about the issues being discussed. We found that the other partner was often loath to reveal how little they knew and would attempt to appear more knowledgeable and involved in the process than when they were interviewed separately from their partner. On reflection, the researchers felt this was due to their desire to appear as an active carer in the family, to hide feelings of guilt for not being involved in issues to do with the children. When interviewed separately, a greater proportion of such respondents were simply unable to answer the questions, as they could not refer to their spouse for 'help'. Added to this, the main caregivers were also more likely to state that their partner was 'hopeless' on such issues but did not tend to make such bold statements about how the household decision making occurred when their partner was present.

• If the evaluation work involves a family, interview both partners separately and look for areas of consensus and non-consensus. Then, bring the two parties together and explore further the areas of non-consensus. With such an approach, it is important to explain the process before starting so that trust is established and the respondents do not feel the researcher is either testing them, or attempting to make the family members disagree with each other.

We used this approach in the family assistance research cases. Some of the main areas of non-consensus were concerning workforce participation of the main carer and household budgeting practices. When these areas of non-consensus were explored with the partners together, it often appeared that these issues had not been fully discussed between them, rather than real non-consensus. If the respondents had been interviewed together first off, we would not have been able to as accurately identify who held the knowledge and strongest attitudes in the partnership about these issues.

It is also possible to run methodologies in parallel and look for areas of consensus and non-consensus in the research findings. For example, a series of face to face interviews in a family run at the same time as a focus group of main caregivers.

**Respondents**

Our final suggestion focuses on using different but related respondent groups to assess the intersubjective certifiable of findings (Hirschman 1986). This way it is possible to see if there are differences between what the
target respondents perceive, and how another party who knows them well perceives them. For example, their
partner or significant other, caregiver, or service provider. Determining if the findings from the different groups are
similar increases intersubjective certifiability.

Such an approach was used in the gambling research when we were seeking information about motivations for
gambling. Respondents were able to discuss at length the reasons why others gamble but were far more
constrained when it came to discussing their own behaviour. This was especially true in the case of problem and
heavy gamblers. Respondents appeared comfortable in discussing how boredom and loneliness were reasons to
gamble but none admitted this was their motivation for participation.

Another aspect of respondents in evaluation research that can raise issues is that of recruitment and retention.
Sometimes when the topic is particularly sensitive, recruitment for respondents is often not possible (or very
productive) by traditional phone method. For example, trying to find women with particular health issues. In
Australia we are fortunate to have a reasonably extensive structure of community health centres even in rural and
remote areas and have often been assisted in the recruitment process by creating networks with these centres. It is
often a far easier task to explain to a health care worker the value of examining the reasons why some female
demographics do not undertake regular health screening, than to a lay person recruited by random telephone
contact.

In terms of retention, we have developed and tested several methods to assist in panel retention. Outside of the
evaluative research setting, we have found the supply of research results a particularly strong incentive in the
business to business context; far more successful that the cash or lottery-type incentive that was thought to be
successful. However, often the issue is not incentive, but simply the dynamics of the population. We estimate that
in a twelve-month period, approximately 20% of the population move address. This places onus on the researcher
to develop a methodology that encompasses ways of recontacting other than by phone or mail direct with the
respondent. We now ask for e-mail addresses along with the name and address of one friend and one relative not
residing with the respondents. The use of electoral rolls and electronic telephone pages services can also assist in
tracing respondents. However, further we would suggest that regular contact is made with respondents, even
outside of the survey period. This can be either by mail or phone and while there is obviously a considerable trade-
off financially, there is much greater likelihood of retention when these activities form part of the process.

**Implications**

This paper has identified ways to improve intersubjective certifiability through preventing the risk of a single
researcher, process, or respondent perspective unduly influencing the research results. The adoption of our
recommendations should result in a richness of interpretation that may not be achieved by one researcher,
respondent group, or research methodology. They also minimise the risk of unduly biased results being reported.

However, thought needs to be given to establishing correct interpretations when following the use of the multiple
researcher framework, there is differing opinion among researchers. Here the use of video and audio tapes along
with interviewer and observer notes cannot be overlooked and can be used in trying to achieve a correct
interpretation. We have found the adjudication of a further party necessary at times. While this might appear
cumbersome it can assist in progressing further to exact an often-regarded unexact research method.
References


Herschell, Robert M., Dr. (1999) *Some Principles for "Quality Control" in Qualitative Research a Phenomenographic Treatise*, First Annual Conference of Association for Qualitative Research, Association for Qualitative Research, Melbourne, 8-10 July.


