The Canadian Journal of Program Evaluation Vol. 16 No. 2 Pages 87–99
ISSN 0834-1516

SOFTLY, SOFTLY CATCH THE MONKEY: INNOVATIVE APPROACHES TO MEASURE SOCIALLY SENSITIVE AND COMPLEX ISSUES IN EVALUATION RESEARCH

Anne Sharp
Catherine Eddy
University of South Australia
Adelaide, Australia

Abstract: Many government program evaluations require capture of information that is hard to measure, of a sensitive nature, and difficult for the respondent to articulate. This article suggests research designs and methodologies to assist in overcoming such problems in evaluation research. Our discussion is illustrated by three evaluation case studies. Suggestions for research design focus on increasing reliability through intersubjective certifiability and the use of triangulated respondent groups, as well as varying the composition of the research team at different stages of the research. Methodological suggestions are for multifaceted research processes, run in parallel and in sequence, to uncover topics on which findings vary and to find information "hidden" in other approaches. Methods for improving recruitment and retention of respondents are also discussed. We conclude by critically evaluating the outcomes of applying these new approaches and discussing the implications of gaining different or new information from adopting such innovative approaches.

Résumé: De nombreuses évaluations de programmes gouvernementaux exigent des informations difficiles à mesurer, délicates, et que le répondant à de la difficulté à articuler. Cet article suggère des concepts et méthodologies de recherche qui aideront à surmonter ces problèmes dans la recherche évaluative. La discussion se fonde sur trois études de cas. Les suggestions de concept de recherche se concentrent sur l'augmentation de la fiabilité par moyen d'intersubjectivité attestable et d'utilisation de groupes triangulés de répondants, ainsi que par la composition variée de l'équipe de recherche à divers stades de l'étude. Au plan méthodologique, on suggère des processus de recherche à plusieurs facettes, exécutés parallèlement et en séquence, afin de mettre au jour les sujets sur lesquels les résultats varient et de

trouver des informations « cachées » dans d'autres approches. L'article traite également de méthodes pour mieux recruter et retenir les répondants. La conclusion constitue une évaluation critique des résultats de l'application de ces nouvelles approches et expose les conséquences de l'adoption de ces approches innovatrices sur l'obtention d'informations différentes ou nouvelles.

Evaluation research often calls for the 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 program). 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 and application are critical, as some of the sensitive or hidden issues, such as a youth's degree of shared values with his/her parents, cannot be fully explored with some of the potential research approaches.

Because of these factors, we need to be extremely careful to avoid introducing uncontrolled-for bias into the research either through the researchers, the respondents we select, or the research design. Although the topic of bias introduced by respondents has received some attention in the literature (Interviewer, 1998), that introduced by researchers and research design has received far less attention (Herschell, 1999; Rust & Cooil, 1994; Towstopiat, 1984). Yet these issues all affect the quality and reliability of the final research results (Welch, 1985). In this workwe 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 other means, such as intuition or guesswork, is intersubjective certifiability (Hunt, 1976, 1991). By *intersubjectively certifiable* we mean that two or more researchers or proc-

esses will draw similar results and interpretations after exposure to the same material. This issue is often addressed in research literature under the concept of reliability.

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 objectivity have been called for by numerous authors (Gabriel, 1990; Herschell, 1999).

This concept of reliability in results 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 or approaches are compared (Wallendorf & Brucks, 1993). Where the same results are arrived at, the research is said to be intersubjectively certifiable. Where it is not, 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 program evaluation research there are sometimes areas where respondents' behaviours and attitudes are hidden. For example, feelings of equity and role sharing in a relationship may not be as fully disclosed when a couple in a partnership are interviewed together as when they are interviewed separately, with no feedback to the other partner regarding their responses. This variation in what the respondent is willing, prepared, or can remember to say affects 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 program evaluations. Some groups of interest may be suited to some methodologies but not others. For example, indigenous populations living in traditional settings may find a focus group methodology very intrusive; a face-to-face methodology may be more appropriate. However, in exploring another subgroup, for example job seekers, the researcher may find it helpful to have more than one respondent to raise issues of interest with and to assess corroboration amongst the different respondents. Having said this, sometimes a group appears to be similar when in fact there are differences that are not immediately obvious and may arise in the context of the focus group. Participants' willingness or preparedness to disclose information may be largely determined by the research setting.

THE EVALUATION CASE STUDIES

We draw upon three case studies to illustrate our discussion. The first case involved a longitudinal evaluation of a federal government family assistance program. This program provided financial payments to the main caregiver in two-parent households with dependent children, where the main caregiver was not engaged in full-time work. The program objective was, through the payment, to provide recognition for the role of caring and to increase equity within the family. It also sought to increase choice for families concerning 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, and perceived barriers to achieving the respondent's desired level of workforce participation. The evaluation took place over three years, with the same 1,300 respondents, where possible, being recontacted at 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 as well as 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 in-depth interview component before telephone interviews with families were conducted.

The third case study involved another piece of longitudinal research, again over three years. The research was conducted for a state government regulatory body. The objective was not to evaluate any program 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 the 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 respondents' stated attitude to gambling (positive or negative).

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

THE RESEARCHERS, RESEARCH DESIGN AND METHODOLOGIES, AND RESPONDENTS

Methods for both improving intersubjective certifiability and maximizing 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 (Herschell, 1999). 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 they bring to bear, and so it may be idiosyncratic (Colwell, 1990; Peter & Olson, 1983). The inter-rater reliability research stream does offer some insight into the reliability of human observers' judgements, but this literature is more about measuring such error 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 perspectives 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 were careful to use a research team that consisted of single people, partnered people with no children, and partnered people with children. This was because each different circumstance was likely to give a person quite different views and experiences of family assistance, and as a result that person would relate to the respondent groups in different ways.

It follows logically that intersubjective certifiability is increased with the use of several researchers. By each disclosing personal experience and information at the beginning of the research process, other researchers in the team can identify possible biases. 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 that they brought a wealth of experience to the research that the other research team members lacked. This experience was extremely useful in conducting the interviews and designing the research instruments, 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 minimized the risk of the research being overly coloured by the experiences of the researchers who also met the criteria for respondents.

As an extension to this practice, at each new research phase of the longitudinal evaluations, we had at least one team member with 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 researchers not to bring assumptions about the likely results of new research to a project when they have past experience with similar projects. Clearly their experience is useful, but this should be balanced by input from other researchers with no such previous topic orientation.

Having such prior experiences disclosed may also assist readers of the results to understand varying perspectives informing the research. For example, in the gambling case study the focus groups were run according to respondents' stated attitude to gambling. Two

researchers moderated the four focus groups: one was a light gambler with a positive attitude and the second was a non-gambler with a negative attitude to gambling. This was explained to the research sponsor. The non-gambling 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. Having different perspectives 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 perspectives" (Groves & Belk, 1994, p. 398) and to "re-energise the team in the tiring process of fieldwork" (Wallendorf & Belk, 1989, p. 73).

The use of multiple researchers offers a number of research design options, which we 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 discussion 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 & Belk, 1989). The 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. One researcher conducts the interview while the second observes. A debrief follows where the researchers compare their interpretations of the interview. Any areas of non-consensus are identified, and a strategy is 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 of the family assistance qualitative phases. We found that respondents were comfortable with 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 at an in-depth interview also makes the observance of body language and 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 verbalizing 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 differing 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 with a focus group methodology compared to, say, in-depth interviews with the same respondents. Differences may arise if issues are considered private or are prone to peer pressure bias on the part of respondents. In such instances, differing responses would emerge in a focus group situation than in an in-depth interview, and so the results may overstate/understate certain findings. Where possible, we suggest using a combination of qualitative methodologies to minimize 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 expressing 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 preceding) such group work with individual, confidential, in-depth interviews.

We used this process in two different ways in our evaluation work. First, in the gambling research we chose a focus group methodology, as we believed 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). 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 could talk more openly about gambling and were relaxed, as they were with "peers."

In the family assistance research, 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 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 further explore 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 them disagree with each other.

We used this approach in the family assistance research. Some of the main areas of non-consensus concerned the workforce participation of the main caregiver 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 existing. If the respondents had been interviewed together first off, we would not have been able to identify as accurately who in the partnership held the knowledge and strongest attitudes 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 with another family.

Respondents

Our final suggestions focus on using different but related respondent groups to assess the intersubjective certifiability of findings (Hirschman, 1986). This makes it possible to see if there are differences between the perceptions of the target respondents and those of another party who knows the respondents well, for example, their partner, caregiver, or service provider. Determining that 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 be-

haviour. 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 not possible (or very productive) by the 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 women do not undertake regular health screening, than to explain to a lay person recruited by random telephone contact.

We have also developed and tested several methods to assist in retention for longitudinal evaluations. Outside of the evaluative research setting, we have found that providing a summary of the research results is a particularly strong incentive when seeking cooperation from business respondents, far more successful than the cash or lottery-type incentives. However, often the issue is not incentive, but simply the dynamics of the population. We estimate that in a 12-month period, approximately 20% of the population move address. This places the onus on the researcher to develop a methodology that encompasses ways of recontacting the respondent other than by phone or mail. In our longitudinal evaluation work, high levels of retention were achieved through collecting e-mail addresses along with the name and address of one friend and one relative not residing with the respondent. The use of electoral rolls and electronic telephone pages services also assisted in tracing respondents. However, the key to retention appears to us to be regular contact with respondents, even outside of the survey period. This can be by either mail or phone, and although there is obviously a considerable trade-off financially, there is much greater likelihood of retention when these activities form part of the process. Respondents develop a sense of ownership of the research and as a result are more likely to notify of changes of address and be willing to participate in further phases of the research.

IMPLICATIONS AND FUTURE RESEARCH

This article has identified ways by which to improve intersubjective certifiability through preventing the risk of a single researcher, process, or respondent perspective unduly influencing evaluation 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. The suggestions we make also minimize 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, and 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 the "correct" interpretation. We have also found the adjudication of a further party necessary at times. Although this might appear cumbersome, it can assist in progressing further to *exact* an often-regarded *unexacting* research method.

The suggestions we have discussed have been successfully trialed in evaluation research, both longitudinally and cross-sectionally. We would encourage further documenting of such practices so that socially sensitive and complex issues can be better captured and measured through research efforts.

REFERENCES

Colwell, J. (1990). Qualitative market research: A conceptual analysis and review of practitioner criteria. *Journal of the Market Research Society, 32*(1), 13–36.

Gabriel, C. (1990). The validity of qualitative market research. *Journal of the Market Research Society, 32*(4), 507–519.

Groves, R., & Belk, R. (1994). The odyssey down under: Interpretations of aboriginal consumption. In B. Sharp (Ed.), *Proceedings of the Australian Marketing Conference* (pp. 398–413). Adelaide: Marketing Science Centre.

Herschell, R.M. (1999). Some principles for "quality control" in qualitative research: A phenomenographic treatise. *Proceedings of the First Annual Conference of Association for Qualitative Research*. Melbourne: Association for Qualitative Research.

Hirschman, E.C. (1986). Humanistic inquiry in marketing research: Philosophy, method, and criteria. *Journal of Marketing Research, 23*(August), 237–249.

Hunt, S. (1991). *Modern marketing theory: Critical issues in the philosophy of marketing science*. Cincinnati, OH: South-Western Publishing.

Hunt, S.D. (1976). *Marketing theory: Conceptual foundations of research in marketing*. Columbus, OH: Grid.

Interviewer, Quality Control Australia. (1998). *IQCA recruitment for qualitative research standards*. Sydney: Association of Market Research Organisations/Market Research Society of Australia Limited.

Peter, J.P., & Olson, J.C. (1983). Is science marketing? *Journal of Marketing, 47*(Fall), 111–125.

Roy Morgan Research. (1999). *Sixth survey of community gambling patterns and perceptions*. Melbourne: Author.

Rust, R.T., & Cooil, B. (1994). Reliability measures for qualitative data: Theory and implications. *Journal of Marketing Research, 31*(February), 1–14.

Towstopiat, O. (1984). A review of reliability procedures for measuring observer agreement. *Contemporary Educational Psychology, 9*, 333–352.

Twyman, W.A. (1996). Designing advertising research for marketing decisions. *Journal of the Market Research Society, 38*(4), 527.

Wallendorf, M., & Belk, R. (1989). Assessing trustworthiness in naturalistic consumer research. In E.C. Hirschman (Ed.), *Interpretive consumer research*. Provo, UT: Association for Consumer Research.

Wallendorf, M., & Brucks, M. (1993). Introspection in consumer research: Implementation and implications. *Journal of Consumer Research, 20*(December), 339–359.

Welch, J.L. (1985). Researching marketing problems and opportunities with focus groups. *Industrial Marketing Management, 14*(4), 245–253.