IS EVALUATION PRACTICE IN AUSTRALASIA UNETHICAL?

DR JENNY NEALE
Deputy Pro Vice Chancellor/Deputy Dean
Faculty of Humanities and Social Sciences
Victoria University of Wellington
PO Box 600
Wellington
NEW ZEALAND
Email jenny.neale@vuw.ac.nz

ASSOCIATE PROFESSOR JOHN OWEN
Director
Centre for Program Evaluation
University of Melbourne
Parkville
Victoria 3053
AUSTRALIA
Email j.owen@edfac.unimelb.edu.au

ABSTRACT

Professional associations of evaluators have recently devoted much time and energy to the development of codes of behaviour to guide the work of evaluators. However, there is a body of anecdotal evidence that many evaluations are carried out by practitioners without reference to any code of practice. This paper sets out a training strategy we have used to encourage evaluators to make sense of codes of behaviour, and to incorporate them into their work. By grouping standards into broad categories evaluators, and those commissioning evaluations, can concentrate on the relevant aspects for the particular stage of the evaluation.

1. INTRODUCTION

have reference points from which to judge whether evaluation practice is acceptable." (Owen and Rogers 1999:151)

Owens and Rogers define a code of behaviour as a "set of standards of practice. Each standard can be regarded as a principle mutually agreed to by people engaged in a given profession – a principle that, if met, enhances the fairness and quality of that practice" (1999: 152).

In recent years "there have been increasing discussions in the professions about how to make sure that proper ethical conduct is not only advocated as an ideal but also practiced" (Schmeiser 1995: 2).

Experience suggests that knowing about the standards or principles is not enough: there is a need for them to be acted on (Owens and Rogers 1999; 164-5).

While it is acknowledged that codes of behaviour are needed they do not solve every problem, and they should primarily been seen as a starting point for ongoing discussion and the locus or frame of reference around which a culture of ethics is developed and maintained (Kimmel 1988, Owens and Rogers 1999, Faisander 1998, Morris and Cohn 1993, Schmeiser 1995, and Homan 1991). Evaluators using codes of ethics or of behaviour and program standards must pay just as much attention to the situation in which the standards are being applied, as to the standards themselves. The existence of standards does not in any way absolve the evaluator from the responsibility of as acting ethical decision-maker" (Owens and Rogers 1999: 152-3),

The existence of standards or codes of practice may also have the unintended effect of foreclosing further discussion because they are interpreted as the final definitive statement on the issue. They are seen as prescriptive and followed to the letter rather than taking the spirit in which they are meant (Homan 1991: 179, Owens and Rogers 1999, Faisander 1998). Codes of practice should raise the awareness of researchers in regard to ethical considerations (Homan 1991). Thus codes of practice or ethical standards should generate debate around ethical issues (Homan 1991), and be understood as part of the overall evaluation process, to be discussed and debated (Owens and Rogers 1999, Schmeiser 1995). The true aspiration of ethical standards is (or at least should be) to "increase the awareness of hical uses of assessment in various contexts such as teaching, counseling, evaluation, and research" (Schmeiser 1995:1). Therefore, the problem remains of how to translate ethical standards from paper to practice.

Applying any code of behaviour or set of standards must be done in a way appropriate to the particular context, and agreed upon or negotiated between the evaluator/researcher and the client/organisation prior to undertaking any project (Owens and Rogers 1999). From the outset codes of practice/sets of standards should be an integral part of any evaluation project.

Morris and Cohn (1993) argue that while ethical considerations in program evaluation "have received a great deal of attention in the relevant scholarly literature, *empirical* studies of the ethical challenges encountered by evaluators are extremely rare" (Morris and Cohn 1993: 621). However, there is some research that has addressed the issues.

Newman and Brown (cited in Morris and Cohn 1993) asked evaluators to state both the seriousness and frequency of violations of the 30 Program Evaluation Standards developed by the Joint Committee on Standards for Educational Evaluation. Of the four groups of standards, evaluators reported the utility and feasibility standards were violated more often than the accuracy and propriety standards. However, the researchers found that there was no difference in the seriousness between the four categories. Other researchers indicate that the most common and most serious problems evaluators report facing are related to reporting and disseminating results (Owens and Rogers 1999, Kimmel 1988, Homan 1991, and Morris and Cohn 1993).

Research by Morris and Cohn (1993) asked 459 respondents to describe both the most frequent and the most serious ethical problems they had faced. Overall there was found to be "a great deal of commonality in the issues" (Morris and Cohn 1993; 628). Most of the problems reported were experienced at, or were related to, the reporting stage of an evaluation, such as "presentation of findings, disclosure agreements, and misinterpretations/misuse of the final report" (Morris and Cohn 1993: 639). However, ethical problems were experienced at the entry stage too, and involved identifying, negotiating and contracting with stakeholders (Morris and Cohn 1993). None of the main problems reported (either the most serious or the most frequent) occurred or related to the "design, data collection, or data analysis phases" (Morris and Cohn 1993: 639). At every stage of the evaluation process problems were reported, evidence of the fact that "ethical problems, can and do, arise in every stage of the evaluation process" (Morris and Cohn 1993: 639). Overall this research found that the pressures faced by evaluators result from a discrepancy between what the client expects the evaluator to find, and what the evaluator actually finds and reports, and then what happens to the findings once the report is presented. As Morris and Cohn (1993: 625) conclude "the stage of the evaluation in which the conflict occurred or was perceived to have occurred appeared to be an important consideration". Ethical issues and problems can arise at any time, any stage of the evaluation process.

How do you then make standards and principles a meaningful part of any evaluation process? One of the ways advocated to move codes of behaviour (or ethical standards) from paper to practice is through training and education (Owens and Rogers 1999, Schmeimser 1995, Homan 1991, and Morris and Cohn 1993). Homan (991:181) argues that "ethics as a component of professional training will…affect the norms of that culture".

Good training also involves teaching standards and/or principles in an experiential format (Owens and Rogers 1999: 165).

'The Program Evaluation Standards' were formulated because there was no "clear definition of what constitutes a reasonable evaluation" (Sanders 1994: 1). The Standards are designed

to help "evaluators identify and confront political reality" (Sanders1994: 4) by providing a framework to be used throughout the evaluation process from design to assessing the evaluation after completion.

AES has been an active participant in the discussion and debate around standards and has developed its own "Guidelines for the Ethical Conduct of Evaluations" which are "complemented by the Program Evaluation Standards) (Evaluation News and Comment June 1998). This discussion was picked up by Scott Bayley (Evaluation News and Comment June 2000) who had an interest in attempting to "explain how the standards can be applied to the major steps of conducting evaluations".

2. METHOD

Drawing on the debate generated through AES, an exercise was carried out as part of a postgraduate evaluation course in June/July 2000 and February 2001. Participants were given a list of 'The Program Evaluation Standards' and a table outlining the four stages of the evaluation process: the planning and negotiation stage, data management stage, reporting and disseminating stage, or the overall management of the evaluation process; and then asked to place each of the standards in one of the stages. Bayley (2000) defined the stages of the evaluation process slightly differently from those given on the table provided for the class exercise. Therefore, to enable a comparison to be made between these two schema, Bayley's stages of the evaluation process have been grouped into the same categories as those developed by Owen (2000) for the class exercise.

Table 1 Stages of the evaluation process

Bayley's stages of the evaluation process	Corresponding stages from the class exercise in which Bayley's stages were placed
Deciding whether to evaluate	Planning and negotiation stage
Defining the evaluation problem	Planning and negotiation stage
Designing the evaluation	Panning and negotiation stage
Collecting information	Data management stage
Analysing information	Data management stage
Reporting the evaluation	Reporting and disseminating stage
Budgeting the evaluation	Planning and negotiation stage
Contracting for the evaluation	Planning and negotiation stage
Managing the evaluation	Management of the evaluation
Staffing the evaluation	Planning and negotiation stage

Those completing the class exercise were asked where possible to place each of the standards in only one of the stages. In comparison Bayley placed each standard in more than one category. In the first class 21 forms were filled out representing the views of 23 students from a post-graduate masters' course where there was a wide range of evaluation experience. In the second class 13 forms were filled out representing the views of 20 people all of whom had some evaluation experience.

3. USE OF STANDARDS

3.1 Planning and negotiation stage

There was some general agreement around standards or principles that, according to Bayley and the majority (at least 50%) of those who responded to both the class exercises, require attention when planning and negotiating an evaluation. Firstly, there is identifying and describing the programme (A1) and the context in which it exists (U1), the stakeholders, their biases, values and power relationships (F2). Also important is the identification and description of the evaluator's own values and beliefs (U4), as it is through all these that the findings will be interpreted. Relevant to this stage of the evaluation process is the fact that the evaluation should help insure the needs of programme participants and other relevant stakeholders are meet (P1). Information should be collected that will help meet those needs, and which answers the key questions asked of the evaluation (U3). Finally, a formal agreement must be entered into that address all of the above including what is to be done by whom and when (P2).

The principle that evaluations should be respect and protect the rights of respondents (P3) is relevant to this stage of the evaluation process according to those who completed the class exercise, but not according to Bayley. He sees the importance of this standard, which states explicitly that evaluations should be '*designed* and conducted to respect and protect the rights and welfare of human subjects" (Sanders 1994: 93) (emphasis added), for all the other aspects instead.

Bayley also indicated that some standards require consideration during the planning process that (the majority of) those who completed the class exercise did not. These centred on issues related to the responsibility and ability of the evaluator, the methods used to gather (or generate) information and the treatment of that information.

3.2 Data management stage

Issues central to this stage of the research process concern the way information is gathered and ensuring the quality of that information. The methods used should "be practical and keep disruption to a minimum" (Sanders1994: 65), and should also lend themselves to generating valid (A5) and reliable information (A6) in terms of its intended use. Information should also (from the moment it is collected or generated through to the analysis and final report) be constantly reviewed and checked for errors, and any errors found corrected (A7). The way the data is analysed whether it be quantitative (A8) or qualitative (A9) should be systematic and appropriate to the questions the evaluation is asked to answer. The evaluation needs to be a fair assessment of the strengths and weakness of the programme (P5) which suggests that the data collected (or generated) covers both these aspects. Finally, in gathering (or generating) this data the evaluator must ensure that respondents are protected from harm (P4).

Bayley indicates that a much wider range of standards are relevant to the data management process. Firstly those that relate to dealing with people, and secondly, those that relate to

dealing with the data gathered from people or other sources and the programme being evaluated. These include issues around dealing with stakeholders and the need for the evaluator to identify the values, biases, and/or power relationships, and ensure formal agreements are adhered to. Other issues, such as transparency (of procedures, information sources, and conclusions), and assessment of all of the above (meta evaluation), are seen to be relevant.

3.3 Reporting and disseminating stage

The standards which require consideration during the reporting stage of the evaluation process, deal mainly with generating and collecting data, and with reporting that data. Firstly, enough detail about the data sources used must be given, so that their adequacy can be assessed (A4). The report itself should be presented on time (U6), include all relevant findings, be made available to all those affected by the evaluation (P6), and be presented in a way appropriate to stakeholders (U5). The report should also be presented in an impartial way, which takes account of the possible distortions that could be made of the findings as a result of the biases of any party (Sanders 1994: 181) (A11). Finally, all conclusions should be "explicitly justified, so that stakeholders can assess them" (Sanders 1994: 177) (A10).

Bayley, however, also considers a wider range of principles and issues relevant to the reporting stage, including those related to dealing with people, how to treat the data, and the programme itself.

3.4 Managing the evaluation

With regard to the overall management of the evaluation process, firstly, the expenditure of resources requires consideration, both in terms of evaluator accountability and responsibility (P8), and in terms of the information the evaluation produces, it must justify the resources expended to generate it (Sanders 1994: 77) (F3).

The credibility of the evaluator (U2) is crucial. The evaluator should be able to ensure those connected to the evaluation are not harmed or threatened (that the dignity of human interactions are respected) (Sanders 1994: 99) (P4). The evaluator should also be capable of dealing with conflicts of interests that may arise (or be revealed), without compromising the integrity of the evaluation (P7). Evaluations should be designed to ensure the welfare of stakeholders (P3). Finally, during the overall management of the evaluation there needs to be assessments carried out in using either these (The Program Evaluation Standards) or other standards (Sanders 1994: 185), so that stakeholders themselves can assess the strengths and weaknesses of the evaluation (A12).

The majority of those who completed the class exercises indicated that consideration of the impact of the evaluation (U7) should be a part of the overall management of the evaluation. Bayley on the other hand sees this aspect being of primary concern in the planning and reporting stages rather than overall management of an evaluation.

Identifying those affected by the evaluation, and their needs (U1), so that the evaluation can then assist those needs being met (P1), is seen by Bayley as something that requires ongoing

consideration throughout the evaluation process. A formal agreement is also part of the overall management of an evaluation, not just something negotiated at during the planning stage. It should either be adhered to or renegotiated if for any reason one or more of the parties can no longer adhere to it (P2). The other principles Bayley identifies as relevant to the management of the evaluation process concern the information gathered, and the presentation of it.

4 DISCUSSION

Given the context where in the class exercises people were asked to allocate a standard to only one stage of the evaluation process there are some interesting differences in approach. For example, Bayley looks at the involvement of people at each stage whereas those involved in the class exercises tend to suggest that once ethical decisions have been made around letting the contract in the first instance, the evaluators will then behave ethically/to a suitable standard at all subsequent stages.

There were also interesting differences between the 'student' group and the 'evaluators' group possibly reflecting the latter group's greater experience with evaluation practice. The student group indicated that the identification and description of the evaluator's own values and beliefs (U4) was important at the evaluation planning stage whereas the evaluators group were almost unanimous in seeing it as crucial at the reporting stage. The evaluators group were more likely to think that consideration should be paid to whether or not the evaluator is the appropriate person to carry out the evaluation (U2) at the planning stage while the student group saw this aspect as being more important in the overall management. Respecting and protecting the rights of individuals (P3) is considered important by over half of the student group at this planning stage and with regard to the overall management of the evaluation. However, while nearly two-thirds of the evaluators also see this aspect (P3) as important at the planning stage, under a quarter deem it important for the overall management of the evaluation.

While there was agreement that the data must lend itself to a fair assessment of the programme (P5), the student group placed the emphasis on this standard at the data management stage while the evaluators saw its importance when reporting an evaluation. Also, a significant minority of the evaluators group (and not Bayley) saw as important at the reporting stage that the way qualitative data is analysed should be systematic and appropriate to the questions the evaluation is asked to answer (A9).

That consideration of the impact of the evaluation (U7) should be a part of the overall management of the evaluation was not seen as a concern for Bayley but was for the majority of those completing the class exercises. However, while 90% of the student group indicated this, it dropped to 70% of the evaluators group possibly reflecting the reality of the working experience of all those involved.

5. CONCLUSIONS

Ethical standards and/or codes of practice inform the evaluators' decisions and/or judgments according to the context within which an evaluation takes place. Standards are not just relevant during the course of an actual evaluation but can be used to generate discussion and consideration of both ethical considerations and the contexts in which they are applied (Sanders 1994, and Owens and Rogers1999). As Fraser (2001) indicates we need both as they address somewhat different questions. Standards deal with the quality of the evaluation while ethics deal with the behaviours of the people involved. Thus, ethical standards and codes of behaviour should be understood and taught as a dynamic part of the evaluation process.

Both the AES discussion and class exercises (despite some differences in form and content) provide good fora for an exploration of the place of evaluation standards in the evaluation process. Numbers participating in the class exercises are too small to place a lot of importance on the differences between the two groups and the similarities and differences with the Bayley schema. However, there appear to be areas where more experience with the evaluation process and/or with different types of evaluation give more weight to some aspects in comparison to others. Areas where there was generally strong agreement could be emphasised. Further work could to be done to look in some detail at the reasons for placing standards at certain stages of the evaluation process and not in others ensuring on-going debate on the issues

ACKNOWLEDGEMENTS

Our thanks to Debbie Small, former M.A.(Applied) in Social Science Research student at Victoria University for her invaluable contributions to this paper

REFERENCES

Bayley, S (2000). Applying 'The Program Evaluation Standards'. **Evaluation News and Comment**, (9), 1, 22-23.

Faisander, S. (1998). "Evaluation and ethics: getting the job done and respecting **Evaluation News and Comment**, (7), 1, 12-13.

Fraser, D. (2001).. "Beyond ethics: why we need evaluation standards". **Evaluation Journal of Australasia**, (1), 1, 53-58.

Homan, Roger (1991). **The Ethics of Social Research.** New York: Longman

Morris, M and Cohn, R (1993). "Program evaluators and ethical challenges: a . **Evaluation Review,** Vol. 17, No.6.

Owen, J. M. (1998). Evaluation models for community crime prevention. Stage III Report. A report to the Institute for Social Programme Evaluation, Murdoch University, Perth, Australia.

Owen, J. M and Rogers, P, J. (1999) **Program evaluation: forms and approaches (2nd ed)**. Sydney: Allen and Unwin.

Sanders, J R. (1994). (Ed). **The program evaluation standards.** The Joint Committee on Standards for Educational Evaluation. Thousand Oaks, CA: Sage. 2nd Edition).

Schmeimser, C (1995). **Ethics in assessment.** North Carolina: Office of Educational Research and Improvement. URL; http://www.libr.ac.nz/cgi-bin...ICit+457&D=ericz&S=AIPPPPOPEPMCML

Worthen, E (1996). Mechanisms for making the Program Evaluation Standards meaningful to graduate education. Paper presented at the Annual Meeting of the American Evaluation Association, November, 1996, Atlanta, GA.

APPENDIX 1 Standards and stage in evaluation process

Standard	Stage in evaluation process											
	Planning			Data management			Reporting			Overall management		
	n	n	B	n	n	B	n	n	b	n	n	B
U1	21	13	x	-	1		-	-	x	-	-	x
U2	8	8	x	2	-	X	1	2		15	4	x
U3	11	8	x	13	5	x	-	-	x	-	-	
U4	13	3	x	7	-	x	9	11	x	-	-	
U5	1	-		-	-		21	11	x	1	-	
U6	-	-		1	-		17	10	x	7	3	x
U7	6	2	x	6	1		6	4	x	19	9	
F1	8	1	x	15	10	x	0	-		1	5	x
F2	19	10	x	2	1	x	1	1		7	4	
F3	6	7	x	1	-		1	1		17	8	x
P1	19	12	x	-	-		1	-	x	4	-	x
P2	20	13	x	-	-	x	-	-		4	-	x
P3	11	8		7	2	x	6	-	x	12	3	x
P4	5	1		12	9	x	4	2		14	4	x
P5	4	1	x	13	4	x	9	9	x	1	-	
P6	0	1		1	-		19	13	x	2	1	
P7	8	5	x	3	-		3	-		13	9	x
P8	4	1	x	1	-		2	-		21	13	x
A1	14	8	x	-	-	x	5	-	x	2	2	
A2	16	8	x	4	6	x	1	-	x	1	-	
A3	9	-	x	3	2	x	7	2	x	9	6	x
A4	2	-	x	10	5	x	11	7	x	1	-	
A5	9	6	x	14	6	x	1	-		2	1	
A6	10	5	x	13	10	x	1	-		1	-	
A7	1	-		18	8	x	1	3		1	2	x
A8	1	-	x	18	11	x	1	-		1	-	
A9	1	-	x	18	13	x	1	5		1	-	
A10	1	-	x	3			15	12	x	-	-	
A11	1	-	x	2			16	8	x	-	2	
A12	1	-	x	1		x	4	-	x	19	11	x
TOTAL*	21	13		21	13		21	13		21	13	

*NB In each case some grids were filled out in pairs and some individually hence 21 represents 23 people (the ‘student) group) and 13 represents 20 people (the ‘evaluators’