BIOGRAPHICAL SUMMARY

Ms Jenni Livingston (B.Sc. Dip.Ed.) Director, Quem Associates Pty. Ltd. (Engineering and Evaluation Consultants)

Jenni Livingston has a longstanding interest in evaluation, coming from a background in science education and professional development. Her first experiences in evaluation work were concerned with federally funded programs – the Science Education Professional Development Program and Using Cases in Professional Development with Lawrence Ingvarson at Monash University. More recently, she has worked with John Owen, Rosalind Hurworth and Gerald Elsworth from the Centre for Program Evaluation, the University of Melbourne on a variety of evaluations. These evaluations include the Implementation of the Curriculum Standards Framework in Victorian schools, a design evaluation of a School Nurse Program, an evaluation of an innovative health promotion program and evaluations designed to assist in the implementation of the computer enhanced curriculum in schools. She is currently completing a Masters degree in Education in the field of evaluation.

THE PLACE OF PROGRAM CLARIFICATION IN EVALUATION.
THE EXPERIENCE OF EVALUATING A SINGLE, EXEMPLARY SCHOOL NURSE PROGRAM.

JENNI LIVINGSTON
QUEM ASSOCIATES PTY. LTD.

Introduction

A School Nurse Program (hereafter SNP) was evaluated by a small team of evaluators from the Centre for Program Evaluation at the University of Melbourne in 1996. This program had operated successfully in an inner city secondary school for over twenty years. Funded initially under the Commonwealth Disadvantaged School's program, the SNP had endured federal/state/local demarcation issues, school amalgamation, departmental re-structuring, curriculum and governance transformation, at least four different sources of funding, and staffing changes, a set of factors that had lead to the demise of many similar school based programs.

The evaluation was commissioned by the state health and community services department as a program clarification exercise. This department had taken over the funding of the SNP for a two year trial period, including an evaluation as one of the conditions for funding. The evaluation was to provide policy advice to a small planning unit. This unit wanted to know how the program operated and, as well, they asked the evaluation team to extract those elements central to its transferability into other contexts, given that it was considered to be exemplary by academic nurses. These questions place the evaluation firmly in the area of program clarification.

What Is Program Clarification?

Program clarification is one of a number of purposes Owen (1993: 14) outlines for evaluations. The purpose of an evaluation might be enlightenment, accountability, program improvement, program clarification, program development or purely symbolic, or a combination of these purposes. As well, Owen (1993: 20) posits a framework of five metamodels to help conceptualize and guide the practice of evaluation. These forms are Impact Evaluation, Evaluation in Program Management, Process Evaluation, Design Evaluation and Evaluation for Development (Owen, 1993: 21). The form called design evaluation is the vehicle for program clarification exercises. The purpose of design evaluation is explained thus;

> There is a need for the "program logic" of an existing program to be made clear to those either directly involved in or having a stake in the program... Design evaluation is an emerging area of interest in its own right because of the contribution it can make to the effective delivery of program outcomes. (Owen 1993: 17)

A Consideration Of The Place Of Program Clarification As Evaluation

Within the design evaluation form described by Owen (1993) I would include work on evaluability assessment (Wholey, 1979, 1994, Smith 1989), program logic (Funnell 1997), program templates (Scheirer 1987), outcomes hierarchies (Patton 1997) and program theory (Bickman 1987, Chen, 1990). The essential aspect, according to Owen (1993) is that "program clarification" is a legitimate purpose for evaluation. However this is easily justified with Owen's heuristic description of evaluation, namely;

> "The process of providing information designed to assist decision-making about the object being evaluated. The term evaluand has been used as an alternative generic term to "object". In practice the evaluand could be a program, a policy, an event, or a training manual or textbook." (Owen, 1993:3)

A further definition by Owen introduces the option of judgment;

> *"Program evaluation is the process of delineating, obtaining and disseminating information of use in describing or understanding the program, or making judgments or decision related to the program". (Owen, 1993: 7)*

In his Evaluation Thesaurus (Scriven 1991), Michael Scriven introduces four senses of evaluation. Scriven's ideas are important because he presents the most thoroughly argued stance within the literature. Apart from Scriven, there are other ways to describe evaluation that more easily accommodate the program clarification purpose for an evaluation (see previously, Owen 1993). Yet another approach denotes evaluation as commissioned research, with the implication that the act of reporting findings to stakeholders then makes the investigation an evaluation.

Scriven outlines what he calls "the key sense" of evaluation as;

> *The process of determining the merit, worth or value of something, or the product of that process...The evaluation process normally involves some identification of relevant standards of merit, worth or value; some investigation of the performance of evaluands on these standards; and some integration or synthesis of the results to achieve an overall evaluation or set of associated evaluations. (Scriven 1991: 139)*

Scriven's three other senses include the notion of evaluation as an investigative discipline, the "unfortunately narrow" sense of evaluation as being the work of evaluators, and lastly, the mathematical sense which he manipulates to reinforce the distinction between measurement and evaluation.

I want to concentrate upon the key sense to examine the claim that the SNP study, the clarification of a school nurse program, is a legitimate form, deserving of the title "evaluation".

Shadish et al (1991) refute the possibility of value-free evaluation although it was strongly espoused in the early years of evaluation. They consider three approaches to valuing, namely prescriptive, descriptive and metatheoretical. Scriven's "key sense of evaluation" (see above) is seen by Shadish et al as "a metatheoretical approach (which) constructs value statements about any entity" and they argue that "a logic like this is widely used in product evaluation, and is implicit in much program evaluation" (Shadish et al 1991:455). They note that "metatheoretic approaches do not specify which criteria to use in the first step of the logic". A prescriptive ethical theory generates its own criteria from the moral act of evaluation of programs aimed at poverty or racism, for example. Descriptive valuing is simple description of stakeholder values.

The next aspect of Scriven's sense of evaluation second is concerned with the "selection of criteria of merit". Shadish et al (1991:457) tell us that "early practice was to use program goals as criteria". For a number of reasons, outlined below, this reliance on goals proved to be an error;

> *(a)   goals are often vague, contradictory or latent;*
> *(b)   program implementation is so heterogeneous and locally controlled that central goals do not overlap heavily with local goals;*
> *(c)   programs have unintended effects that can be just as important as planned goals; and*
> *(d)   managers often specify modest goals so as to succeed at something less important rather than fail at something more important; or else*
> *(e)   they find themselves stuck with unrealistically high goals that overpromised to secure initial funding. (Shadish et al, 1991: 457)*

In setting criteria of merit, there has been no agreement within the evaluation field as to a replacement for program goals. Different evaluators have relied upon a variety of sources, including;

(a)   claims by clients, service deliverers, and managers about program achievement and factors leading to success of failure (for example, Stake);

(b)   factors feeding into decisions that policymakers or managers must make (for example, Weiss);

(c)   systems models that explain how inputs relate to processes and outcomes (for example, Chen and Rossi);

(d)   the degree to which the material needs of clients are met (for example, Scriven);

(e)   information needs of stakeholders (for example, Bryk) or of managers and policymakers (for example, Wholey)

Interestingly, the SNP evaluation elicited criteria of merit from each of these categories.

The third step in Scriven's logic, and the key area where program clarification might be charged with fraud, as evaluation is the measurement of performance. This is not an issue with the SNP evaluation, where the designation as exemplary implied that performance had been judged. However, Shadish et al claim that inferences about causal connections are integral to this concept of measurement, although Scriven claims this is a "limited task" (Shadish, 1991: 460). Since this conundrum has been exposed, many evaluators have sought to find out why the evaluand works, with a view to providing useful knowledge for program improvement. In other words, Scriven's metalogic is seen to be inadequate since it fails to address the issues tackled in program theory construction.

This issue can be tackled from another point of view. Many authors claim an atheoretical or methodologically driven approach predominates and characterizes the evaluation field (Chen 1994). As well, Scriven (1991: 286) argues that program theory is largely irrelevant to the work of evaluators. But in the case of Scriven's (1991:180) Goal Free Evaluation, he argues that an unbiased evaluator who has not been told of the purpose of the program "does the evaluation with the purpose of finding out what the program is actually *doing* without being cued as to what it is *trying* to do". He argues that "merit is determined by relating program effects to the relevant needs of the impacted population, rather than to the goals of the program for the targeted or intended population". Interestingly, the SNP evaluation functioned in a very similar way to goal free evaluation.

Bickman (1987: 12) draws attention the role program theory description can play in uncovering unintended effects. Bickman (1987: 12) says:

According to Scriven's (1980) goal free philosophy, theory construction and development allows the evaluator to look for either positive or negative effects where they might not be expected by program designers and implementers. The underlying theory allows the evaluator to depart from political objectives and focus upon objectives that can be inferred from the operation of the theory.

The issue is not contradictory, however. As mentioned previously, Scriven (1991: 286) argues that normative program theory should not be the work of evaluators; rather it should be the "concern of mainline social scientists". He argues that:

"although payoffs in program theory territory will often occur and should be looked for and reported on, they are not the business of the evaluator and making them part of the business will deflect resources, effort and intellectual focus from evaluation, which is hard enough as it is.

In summary, Scriven's objections to program clarification as evaluation are not epistemological; rather they are logistical.

Program Theory And Theory Based Evaluation in General

Worthen (1996) has traced briefly the origins of theory based evaluation, noting the work of a number of authors to conclude that the roots of "theory based evaluation" originate with the work of Weiss (1972) and Fitz-Gibbon and Morris (1975). For Fitz-Gibbon and Morris (1975), a theory based evaluation;

> *Is one in which the selection of program features to evaluate is determined by an explicit conceptualization of the program in terms of a psychological, social psychological or philosophical theory. (Fitz Gibbon and Morris 1975/1996)*

According to Fitz Gibbon and Morris, the evaluator's role is to check the actual program against the intended theory, theory that has been derived from professional social inquiry (Lindblom and Cohen, 1979). Weiss (1975) gives the evaluator a different role, hypothesizing a sequence of events or implied causal connections within the program. Again, this process of conceptualizing the expected paths of change is to drive the evaluation design. However, Weiss does not source her theory explicitly in professional social inquiry. Her evaluator has a role as well.

Worthen (1996) then outlines the major written contributions over the past decade under the editorships of Bickman, and Chen and Rossi. Worthen also mentions the work of Quay and Wholey as earlier watersheds contributing to the thinking about theory based or theory driven evaluations.

The work of Wholey (1979, 1994) and Smith (1989) which led to the conceptualization of evaluability assessment has re-introduced debate about what constitutes evaluation. The initial view of evaluability assessment was, as the name implies, a precursor to an outcome evaluation, to be undertaken to describe program goals and processes and to determine whether further evaluation effort was justified. Shadish, Cook and Leviton (1991) place the origin of evaluability assessment with Horst et al (1974) who described it as a tool to "ameliorate the problems of lack of definition, clear logic and management" in programs. The value of this tool was appreciated by a number of evaluators (for example, Smith 1989) who had been concerned at the failure of a number of evaluations for just these reasons.

Chen (1990, 1994), Chen and Rossi (1987) and Bickman (1987,1990) see the need to establish a "program theory" initially to clarify connections between aspects of program operation. The evaluation then establishes, or not, the causal connections between these components. McClintock (1990) contributes to this field by emphasizing the importance of establishing program philosophy. Suggestions for data collection methodologies, data analysis techniques and models for depicting the findings usually accompany each contribution.

However, most examples in the evaluation literature using program theory have centred upon using the program theory in one of two ways. The program theory in these cases is derived by evaluators from stakeholders or from social science theory, or most recently, a combination of both sources. (See Chen 1990, 1994, Chen and Rossi (1987), Bickman (1987, 1990.) In these situations the program theory is largely derived in the early stages of an evaluation, to either posit causal linkages or to establish a normative theory, with the actual situation being tested against this theory in the methodologically driven, usually quantitative, outcomes-focused stages of an evaluation.

Chen and Rossi (1987) have redefined the theory in "theory-driven evaluation" in the following ways;

> *Theory is substantive knowledge that is action-oriented, preferably grounded in previous research, concerning a program (or policy) and the social problem that the program (or policy) attempts to alleviate…*

And;

Program theory is: A specification of what must be done to achieve the desired goals, what other impacts may also be anticipated and how these goals and impacts would be generated (Chen 1990)

Most program theory oriented evaluations proceed by defining the "ideal" program, most commonly by integrating stakeholder and social science theory or by relying upon documentation prior to implementation of the program. Evaluators write at length of the paucity of documented program descriptions let alone comprehensive statements. (Lipsey 1985). The actual program is then evaluated against the derived or documented program theory.

Bickman (1996) writes of the significant role of program theory in a quasi-experimental evaluation, the Fort Bragg Demonstration, where the theory tested was that "a comprehensive, integrated and coordinated continuum of care is more cost-effective than a fragmented service system with a limited variety of services" (p112). Using his own definition (Bickman 1987) of program theory as "a plausible and sensible model of how a program is supposed to work", the use of program theory to describe program operations, proximal and ultimate outcomes and linkages between these aspects of the program is using the program theory as preliminary action, to drive the evaluation questions for the study. This theory served as the original organizing principle of the evaluation and identified key assumptions underlying the program.

The Fate of the SNP Evaluation

As a member of the SNP evaluation team, my role involved intensive observation, key informant interviews and document review, data analysis and writing of the report for the commissioning unit (See Livingston et al, 1996). Other team members worked with the literature search, data analysis and writing. However, we were all involved in the intensive negotiations with the commissioning team. This fruitful interaction meant that findings were conveyed regularly, minor modifications made along the way and a draft report was well received. However, at the meeting where the final report was presented, the newly appointed regional nominee on the committee greeted the report with horror, indicating that if the report was ever released, "every school in the region would be on my back to have one". Needless to say, the program struggled on for another twelve months with funding from the suicide prevention program, finally coming to a halt at the end of 1998. The planning unit disappeared and the grand policy plans disappeared.

The vagaries of political policy making are not worth loosing sleep over, but I had deeper concerns about the extent to which we had really understood the program. The report was called "It's not just waiting for sick kids" and my challenge was to articulate what the nature of the program was if it is "not just waiting for sick kids". To achieve this understanding I have re-examined the data, the literature and the way the evaluation proceeded. This reflective process has taken place outside the evaluation, without the constraints of time and money that are fundamental to the evaluative process. Whether a nursing person would have described the program theory of this successful **interventionist** school nursing program more easily than a non-nursing person is worth debating as well. I suspect it would have mattered little, since I know of no comparable work in the literature. The point is that a successful program has been described in a way that it can be taken up and implemented at some time or place in the future, and the learnings are not lost.

Future roles for Program Clarification

I have argued that the SNP program clarification exercise counted as evaluation for two reasons. The first is based on the designation of the program as exemplary, with the implication that it successfully achieves worthwhile goals. As the evaluation progressed, I was concerned at the source of this information but it was not within the ambit of the evaluation to do so. However, the presence of successful outcomes, and the idea that the program was worthwhile were givens for the evaluation.

The second is derived from the process of program clarification. Program clarification necessarily involves a search for what makes a program valuable, with the only variation being the source of the values and the program logic knowledge. These can be from the program alone, a combination of social science theory and the program, or from social science theory alone. The parameters for evaluating a program are more likely to be true to the program when an intensive study of the program is made. For example, SNP evaluations in the literature relied upon self reporting in terms of an externally derived checklist with a single key item per 15 minute period. My observations showed that collecting data using these parameters was absurd. Some consultations were incredibly brief encounters to check on progress; others started with the presenting complaint but were moved by the nurse to cover other far more important issues. Although much data was tabulated to indicate the extent of the nurse's work, the stories of consultations between students and the nurse gave a far stronger picture of the nature of this SNP.

A more important question concerning program clarification, however, is whether any evaluation can proceed without some form of program clarification taking place. Program clarification demands a deep understanding of the program content and process by the evaluator. Whether this can be achieved given the normal constraints of an evaluation is worthy of debate. My current position, based upon my work on establishing a program theory for an interventionist school nurse program and other program clarification evaluations is that some form of program clarification should be an essential component of every evaluation.

Bibliography

Bickman, L. Ed (1987) *Using Program Theory in Evaluation*. New Directions in Program Evaluation 33. San Francisco: Jossey Bass.

Bickman, L. ed. (1990) *Advances in Program Theory*. New Directions in Program Evaluation 47. San Francisco: Jossey Bass.

Chen, H-T. (1990) *Theory Driven Evaluation: A Comprehensive Perspective*. Newbury Park, CA: Sage

Chen, H-T. (1994) Current Trends and Future Directions in Program Evaluation. *Evaluation Practice* 15:3, 229 – 237.

Chen , H-T., and Rossi, P. H. (1987) The Theory Driven Approach to Validity. *Evaluation and Program Planning* 10, 95 – 103.

Conrad, K. J. and Miller, T. Q. (1987) Measuring and Testing Program Philosophy. *New Directions in Program Evaluation* 33, 19-42.

Fitz Gibbon, C. T. and Morris, L. T. (1975) reprinted (1996) Theory Based Evaluation *Evaluation Practice* 17:2, 177 – 184.

Funnell, S. (1997) Program Logic: An Adaptable Tool for Designing and Evaluating Programs. *Evaluation News and Comment*. 6:1, 5 – 17.

Leviton, L. C., Collins, C. B., Laird, B. L. and Kratt, P. P. (1998) Teaching Evaluation using Evaluability Assessment. *Evaluation* 4: 4, 389 – 409.

Lindblom, C. E. and Cohen, D. K. (1979) *Usable Knowledge*. New Haven: Yale University Press.

Lipsey, M.W., Crosse, S., Dunkle, J., Pollard, J., and Stobart, G. Evaluation: The State of the Art and the Sorry State of Science. *New Directions in Program Evaluation*. 27, 7-28.

Livingston, J. A., Hurworth, R. E. & Peddey, L. (1996) *It's not just waiting for sick kids…* Centre for Program Evaluation, University of Melbourne.

McClintock, C. (1987) Conceptual and Action Heuristics: Tools for the Evaluator. *New Directions in Program Evaluation* 33, 43 – 57.

McClintock, C. (1990) Evaluators as Applied Theorists. *Evaluation Practice* 11:1, 1 - 12

Owen, J. M. (1993) *Program Evaluation: Forms and Approaches*. New South Wales, Australia: Allen and Unwin.

Owen, J. M., and Rogers, P. (1999) *Program Evaluation: Forms and Approaches (2nd edition)* New South Wales, Australia: Allen and Unwin

Patton, M. Q. (1986) *Utilization Focused Evaluation*. (2nd Ed.) Beverley Hills, CA: Sage

Patton, M. Q. (1997) *Utilization Focused Evaluation*. (3rd Ed.) Thousand Oaks: CA: Sage

Scheirer, M. A. (1987) Program Theory and Implementation Theory: Implications for Evaluators. *New Directions for Program Evaluation*. 33, 59 – 76

Scriven, M. (1991) *Evaluation Thesaurus*. Newbury Park CA: Sage.

Shadish, W. R., Cook, T. D., and Leviton, L. C. (1991) *Foundations of Program Evaluation: Theories of Practice*. Newbury Park, CA: Sage.

Weiss, C. (1972) reprinted (1996) Excerpts from Evaluation Research: Methods of Assessing Program Effectiveness. *Evaluation Practice* 17:2, 173 – 175.

Worthen, B. (1996) Editor's Note: The Origins of Theory-Based Evaluation. *Evaluation Practice* 17: 2 169 – 171

Yin, R. K. (1988) *Case Study Research: Design and Method*. Newbury Park, CA: Sage.