

# 9

The usefulness of program theory *for* monitoring and evaluation can be enhanced by incorporating information about *the* context in *which* the *program* operates, by defining success *criteria* and *comparisons for* judging and interpreting *performance* information, *and* by *identifying* sources of *performance* information.

## Developing and Using a Program Theory Matrix for Program Evaluation and Performance Monitoring

*Sue C. Funnell*

Inadequacies of performance information systems, including those used for program evaluations, have been well documented (Perrin, 1998; Winston, 1999). Sometimes these inadequacies arise from incomplete program theories or simplistic applications of program theory that overlook fundamental evaluation principles. One deficiency concerns incomplete or inappropriately focused information systems. This can occur for several reasons. First, a performance information system may concentrate on inputs, processes, and activities that apply to the lowest levels of outcomes and may overlook those that are used by the program to achieve higher levels of outcomes. Second, a performance information system may focus too much on inputs, processes, and outputs and too little on outcomes. Third, there is often a failure to link performance information to explicit and defensible evaluative criteria and to some basis for interpreting and judging performance. Finally, program theory that only looks at the impact of the program and ignores other causal factors can encourage implicit and uncritical attribution of outcomes to the program.

The *program* theory matrix approach **has** been developed in order **to** provide a means of systematically addressing these concerns.<sup>1</sup> It is an adjunct to other approaches to the use of program theory for performance measurement and evaluation. The approach has been refined through application to many different types of programs over more than fifteen years. It is particularly useful in helping people who might be less familiar with key evaluation principles, such as those identified previously, or who are uncertain about how to apply those principles in conjunction with program theory. This chapter

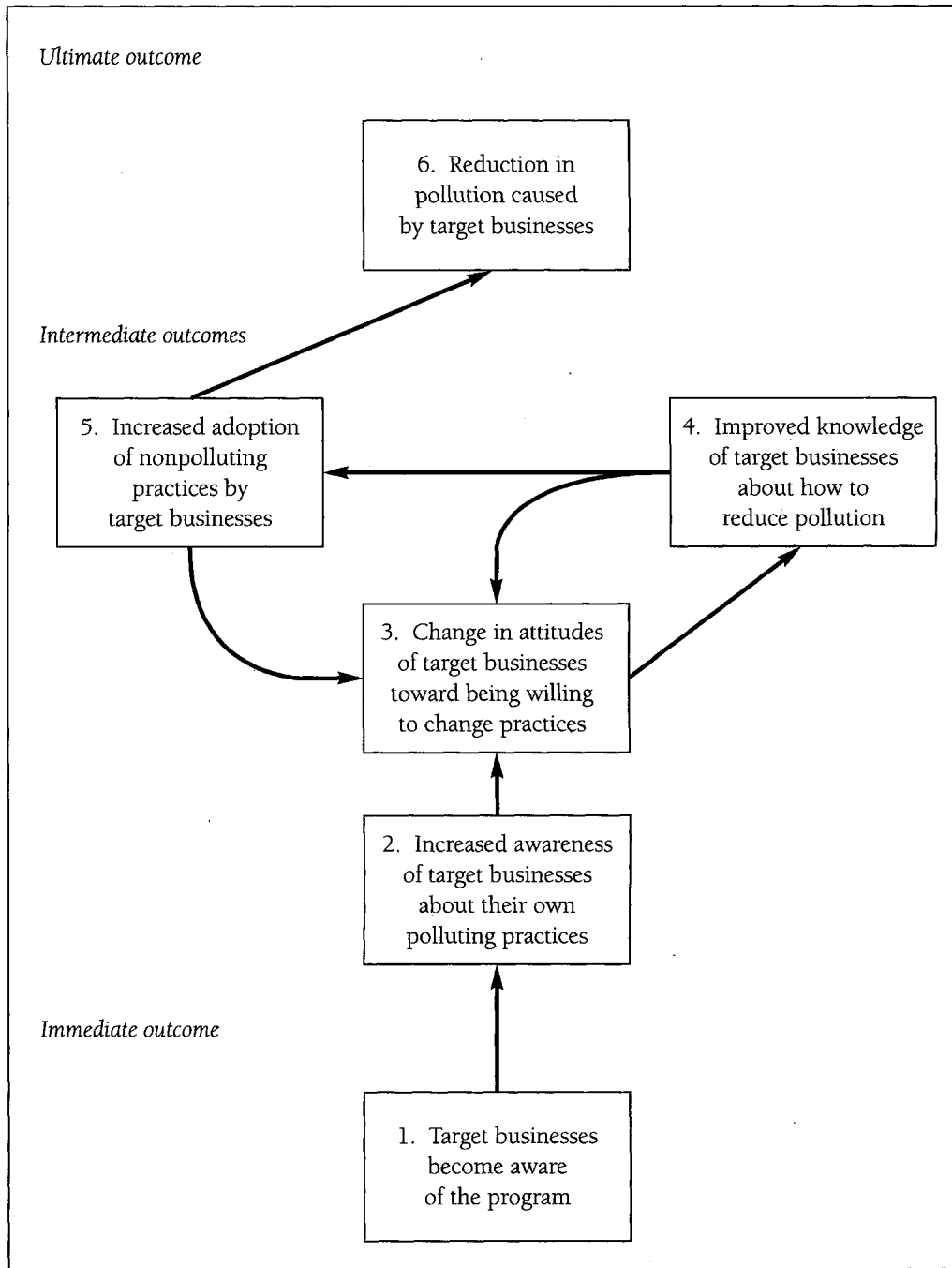
outlines the essential features of the approach and discusses how it has been used to address the concerns just identified.

The principles encapsulated in the program theory matrix were originally developed in 1985 as an initiative of one state government, New South Wales (Lenne and Cleland, 1987). The Australian government later recommended the application of these principles as an important part of planning evaluations of federal programs (Australian Department of Finance, 1994). The approach has been used by municipal, state, and federal government agencies and by not-for-profit agencies across a wide range of programs. For example, at the 1990 Conference of the Australasian Evaluation Society, twelve groups of authors presented papers showing adaptations and applications of the approach to the evaluation and performance monitoring of a range of different programs—management information systems, health promotion, a public information service, a trade waste program of a water utility, a tourism media campaign, a drug and alcohol program, a sports drug agency program, a state library program, a government information service marketing program, business and strategic planning in a utility, a range of HIV-AIDS programs, and a residential care program. An overview paper described the methodology and identified lessons learned concerning the methodology, drawing on the twelve papers (Funnell, 1990).

### **Essential Features of the Program Theory Matrix Approach**

A *program theory* as described in this chapter consists of seven components that are typically portrayed in matrix form. The program theory matrix approach begins with the articulation of a sequenced hierarchy of intended outcomes. The hierarchy commences with immediate outcomes (for example, the target group is successfully reached by the program). These are followed by a chain of intermediate outcomes (for example, changes in knowledge and practices of target group) that in turn are followed by wider and often long-term impacts (for example, alleviation or satisfaction of need that gave rise to the program). Then for each identified outcome, a series of questions is posed, the answers to which are recorded in the matrix: What would success look like (for example, the nature of the desired changes in knowledge and with whom)? What are the factors that influence the achievement of each outcome? Which of these can be influenced by the program (for example, quality of service delivery)? Which factors are outside the direct influence of the program (for example, economic climate, past experiences of clients, competing programs)? What is the program currently doing to address these factors in order to bring about this outcome (for example, staff training, risk management)? What performance information should we collect (quantitative and qualitative indicators and comparisons)? How can we gather this information (for example, interviews, observations, administrative records)? These questions, although clear and easily understood, address complex evaluative issues, such as deciding on evaluative criteria,

**Figure 9.1. Hierarchy of Intended Outcomes for Small Businesses Pollution Program**



identifying potential data sources, and recognizing the impact of factors outside the program.

The hierarchy of outcomes used for the evaluation of an advisory program to change the polluting practices of small businesses is shown in Figure 9.1, which should be read from the bottom up. Table 9.1 shows the

**Table 9.1. Examples of Application of Program Theory Matrix to One Level of the Outcomes Hierarchy for the Small Businesses Pollution Program**

1. Intended Outcome <sup>a</sup>	2. Success Criteria <sup>b</sup>	3. Program Factors Affecting Success	4. Nonprogram Factors Affecting Success	5. Activities and Resources of Program	6. Performance Information <sup>c</sup> — Examples for Columns 2 to 5	7. Sources of Data
Change in attitudes of target businesses toward being willing to change practices	Agreement by businesses to meetings with program advisers with a view to identifying possible solutions; few refusals	Availability of confidential advisory assistance Extent to which program can convince businesses of benefits of change	Extent of target businesses' "illegal" practices and whether businesses are prepared to take the risk of exposing their problems to the program in order to receive assistance	Promotes advisors and makes commitments about confidentiality Identifies and promotes proven benefits to similar businesses—uses case examples Offers and undertakes follow-up advisory visits Develops partnerships with industry associations Tailors paperwork and expectations to the type of business involved	Percentage of businesses that request and receive advisory assistance (columns 2 and 5), compared with targets set for each industry Percentage of businesses receiving advice that rate the advice as credible and useful (columns 3 and 5), compared across target industries, and whether there is any relationship between perceived usefulness and apparent impact on willingness	Administrative records of requests for assistance and receipt of assistance Review of educational materials and strategies developed for the program and for each industry Post-program anonymous survey of businesses Advisors' client records Interviews with advisors Anonymous survey of businesses Site visits and case studies by evaluator
Preparation of action plans that include defined key elements Business-specific examples of increased willingness	Extent to which the program processes are burdensome for businesses Success with which the program engenders industry association support	Businesses' beliefs and past experiences concerning cost and benefits of change Whether the views and actions of competing businesses favor reducing pollution Size and nature of businesses and capacity to com-	Businesses' beliefs and past experiences concerning cost and benefits of change Whether the views and actions of competing businesses favor reducing pollution Size and nature of businesses and capacity to com-	Percentage of businesses that request and receive advisory assistance (columns 2 and 5), compared with targets set for each industry Percentage of businesses receiving advice that rate the advice as credible and useful (columns 3 and 5), compared across target industries, and whether there is any relationship between perceived usefulness and apparent impact on willingness	Administrative records of requests for assistance and receipt of assistance Review of educational materials and strategies developed for the program and for each industry Post-program anonymous survey of businesses Advisors' client records Interviews with advisors Anonymous survey of businesses Site visits and case studies by evaluator	Administrative records of requests for assistance and receipt of assistance Review of educational materials and strategies developed for the program and for each industry Post-program anonymous survey of businesses Advisors' client records Interviews with advisors Anonymous survey of businesses Site visits and case studies by evaluator

mit resources to preparation of plans  
Whether the offer of assistance comes at the "right time"

with target and planning guidelines (column 2)

Percentage of businesses that cite the advisory program as having contributed to willingness (column 2 and 3)  
Examples of impact on willingness (column 2)

Reasons for nonparticipation or nonpreparation of plan and extent to which reasons relate to program factors and activities (columns 3 and 5) or nonprogram factors (column 4)  
Extent to which successful partnerships are developed with the industry associations (columns 2 and 5)

and so forth, kept by advisors  
Anonymous survey of businesses  
Interviews with advisors  
Interviews with advisors and with key informants in industry associations  
Review of actions taken by industry associations

---

<sup>a</sup>Level 3 in Figure 9.1.

<sup>b</sup>Targets differ for different industries.

<sup>c</sup>Including comparisons.

development of the matrix for one level of outcome in this hierarchy. Many outcomes hierarchies have feedback loops and branches, and the matrix approach is as applicable to such hierarchies as it is to a simple linear one. Figure 9.1 shows some simple feedback loops.

This chapter explains the use of the matrix. However, understanding the principles underpinning the program theory matrix is more important than filling in the boxes. Moreover the columns of the matrix can be adapted to suit purpose, audience, and program context. For example, if it is useful to do so, one can split the activities and resources column into separate columns.

In addition, the underlying principles are applicable to many projects undertaken by evaluators. For example, the principles of the program theory matrix were used to develop criteria for a commissioned review of the quality of performance information in the annual performance reports of all federal government portfolios to Parliament (Funnell, 1993).

### **Common Difficulties Associated with Performance Indicators and Performance Monitoring**

Four common difficulties with performance indicators have been identified in the earlier discussion. In the following section, I discuss how the program theory matrix and the principles underpinning the matrix can help overcome those difficulties.

**Insufficient Attention to the Measurement of Inputs, Processes, and Outputs Needed to Achieve the Higher Levels of the Outcomes Hierarchy.** Typical inputs-processes-outputs-outcomes approaches to program theory (for example, Bennett, 1979; Suchman, 1967; Wholey, 1983), although incorporating the type of outcomes hierarchy shown in column 1 of Table 9.1, can as a consequence of their linear nature lead to dissociation between particular activities and the particular outcomes they are intended to achieve. By contrast, the program theory approach that uses a matrix (as portrayed in Table 9.1) rather than a line emphasizes that inputs (resources) and processes (activities) operate at all or most levels of the outcomes hierarchy. This approach encourages reflection on and measurement of the relationships among inputs, processes, outputs, and outcomes at all levels of the hierarchy. For example, in the small businesses pollution program, one set of inputs, processes, and outputs was applied to the lower-level intended outcome concerning change in target group awareness, and a different set of inputs, processes, and outputs was applied to a higher-level intended outcome concerning change in target group practice. Performance measures were developed for program processes at all levels of the hierarchy as a means of measuring program implementation.

This approach also reinforces the fact that lower-level outcomes do not in themselves lead to higher-level outcomes. Rather, the cause-effect process within the chain of outcomes is typically mediated by additional inputs, processes, and outputs at each level of the chain.

**Measuring What Is Easy Rather Than What Is Important: Insufficient Attention to Outcomes.** A common problem associated with the selection of performance indicators is that the indicators may relate only to what is easy to measure (typically inputs, processes, and outputs), leaving out other important aspects of performance, especially outcomes. The matrix approach makes the hierarchy of intended outcomes the backbone of the program theory to which other parts of the program theory (for example, inputs, processes, and outputs) are attached. This extra prominence of the hierarchy of outcomes can help ensure that important intermediate outcomes are not overlooked. Evaluators using the matrix are encouraged, for each outcome in the hierarchy, to make a judicious selection of measures relating to each of columns 2 to 5 in the matrix and to the relationships among them.

The outcomes hierarchy can also provide a structure for reporting the findings of an evaluation. Each chapter in the report can address one level of the outcomes hierarchy and can also include relevant information about program implementation and nonprogram factors that affect achievement of the outcome. This is an alternative to an evaluation report whose chapters are structured around methods of data collection. Examples of the use of the program theory matrix to structure an evaluation report include the evaluation of the small businesses pollution program illustrated in Figure 9.1 and Table 9.1 (Funnell and Ford, 1998) and an evaluation of a United Nations AIDS program (Funnell, 1999).

Sometimes insufficient attention to the measurement of outcomes occurs because at the time of developing the program theory, little thought is given to how outcomes might be measured. Ideally, such thought is applied early in the development of a program so that measures of outcomes can, where possible, be incorporated in routine data collections. The program theory matrix encourages the identification of sources and methods of data collection for each item of required performance information (see column 7 in Table 9.1). The identification of sources of information, although not an essential part of program theory, is the bridge between the program theory and the performance-monitoring system or evaluation design.

**Failure to Link Performance Indicators to Explicit and Defensible Evaluative Criteria and Standards.** Failure to explicate, for each intended outcome, the success criteria by which performance can be evaluated can foster measurement of the easily measurable. Explicating criteria is a step toward safeguarding against the slide to the easily measurable and the possibility of goal displacement. Being explicit about criteria draws attention to the fact that where it is not feasible to collect information about some criteria (typically qualitative criteria) through routine monitoring, then a more in-depth program evaluation may be required from time to time.

Those who apply more traditional approaches to program theory will at some point need to identify the success criteria by which the achievement of

the various components of the program theory will be assessed. The program theory matrix simply provides a systematic process for undertaking and recording this step. Many traditional approaches to identifying success criteria, such as stakeholder analysis and consultation and review of relevant literature, can be used in conjunction with the matrix. For example, the development of criteria for evaluating the utilization and impacts of several program evaluations conducted by a state government agency drew on the literature on utilization of program evaluation (Funnell and Harrison, 1993). Generic program theories, such as those developed by the author and colleagues (Funnell and Lenne, 1990), are another means of gaining a rapid point of entry for identifying evaluation criteria for programs that are manifestly different but appear to have similar underlying structures.

Column 2 in Table 9.1 shows how the matrix approach to program theory encourages explication of criteria. In this example, the success criteria show how the program management and staff would like to see the broad intermediate outcome—willingness to take action—manifested in the behavior of businesses with which the program is working. In another example, the program theory developed for an evaluation of an employment program for mature-age people specified that one of its outcomes should be that participants retain jobs obtained through the program. To make this outcome measurable in a way that reflected the program intent, it was necessary to define the terms of this outcome statement. Specifications included that people placed by the program should, ideally, be in the same or a different job for twelve months or more, that the employment should be continuous, that it should be full-time permanent rather than casual, and that the results should be equitably distributed across various specified subgroups. These criteria were derived from literature on employment programs and long-term chances of success, program financial break-even information, and policy statements concerning priorities for different subgroups. The specification of these criteria was the bridge to the selection of performance indicators both for performance-monitoring purposes and for a discrete program evaluation (Funnell and Mograby, 1995). The performance information in relation to this outcome also included levels of performance that were less than the ideal—for example, employed for six months or three months, employed on a noncontinuous or casual basis. In this way, it is possible to use the step of identifying success criteria (column 2 in the matrix) to set up a goal attainment scale from “least desirable” to “most desirable” performance against each outcome.

Many data are collected that never become information because nobody ever thought very seriously about how the data might be used to draw conclusions. The program theory approach shown in Table 9.1 (see column 6) encourages program managers and evaluators to make explicit the comparisons that they will use to make judgments about the adequacy of performance or to draw cause-effect conclusions about performance. Typically, comparisons are with standards, targets, norms, past performance, and sometimes with



other programs. It is important that there is some reasonable basis for selecting standards and targets (Funnell, 1993). Examples of comparisons used for the employment program for older people were comparisons of outcomes for different subgroups to determine whether results were equitable and comparisons with the employment rates of similar groups in the wider population. In the case of the small businesses pollution program that involved twelve very different types of industries (for example, automotive repairs, marinas, shopping centers, chemical industries, market gardeners), the specific criteria and targets differed depending on the size, nature, complexity, and current polluting practices of each target industry.

**Uncritical Attribution of Outcomes to the Program.** The program theory matrix approach tackles the issue of causal attribution by identifying the possible effects both of program factors (column 3) and of external factors (column 4) on intended outcomes of the program and by encouraging measurement of both types of factors. By so doing, this approach explicitly incorporates what Lipsey (1993) has referred to as the *exogenous factors* that should be included in a program theory. It is also important to identify and measure the way in which program activities (column 5) are being implemented to influence or manage those factors. A performance information system or evaluation that wishes to draw causal conclusions will need therefore to have information from each of columns 2 to 5 in Table 9.1 and about the relationships among them. In addition, prior to performance measurement and empirical evaluation, an assessment of the internal logic of the program can be undertaken, which focuses on the degree of correspondence among the various columns and the completeness of each. One such assessment for the small businesses pollution program would be whether the program incorporates actions to allay concerns of businesses about having their polluting practices exposed. This question can be asked independently of the further question of whether those actions are effective.

Performance information relating to causality can be drawn from a range of evaluation designs—experimental, quasi-experimental, or even nonexperimental. However, all designs typically depend on there being some identification of the program factors (in classic terms, the independent variables) and nonprogram factors (confounding variables) that are likely to affect the outcomes (dependent variables). The program logic matrix provides a framework for drawing out the most important factors in a systematic way. For example, the evaluation of the previously mentioned employment program used some quasi-experimental comparisons that were incorporated in column 6 of the matrix for that program. It compared outcomes for program participants with those of the relevant population. It made comparisons with past studies, taking account of the economic climate at the time (an external nonprogram factor). And it made comparisons that suggested it was unlikely that the program was achieving its success through *creaming*—selecting only those potential participants who were most likely to succeed.

Identification of factors likely to affect results and the likely relative impact of program versus nonprogram factors positions the evaluator to make an educated judgment about how far up the outcomes hierarchy it would be sensible to attribute impacts to the program. When the nonprogram factors substantially outweigh the program factors that affect an outcome, then caution must be exercised when attributing a particular outcome to a program.

### **Other Uses of the Program Theory Matrix**

This chapter has focused on the use of the program theory matrix for purposes of performance measurement and evaluation and has touched on its use for assessing the internal logic of a program and for communicating findings. However, there are many other uses of program theory matrices, including using them to negotiate accountabilities and using them for team building and staff morale. As the manager of a drug and alcohol program for prisoners who was using program logic to provide a framework for developing and testing her intervention model said, "Participation in the process reassured staff that they are part of a program which has a commonsense, coherent and believable rationale. This reassurance provides a buffer against external criticism. . . . However, our evaluation process is not simply about giving a warm inner glow to staff. By exposing the assumptions underlying the program, and making them testable, we have raised expectations that they will be tested and will continue to be tested from time to time. We certainly have a commitment to doing so" (Matthews and Funnell, 1987, p. 7).

### **Conclusion**

The program logic matrix offers a systematic process for developing and applying program theory in a way that guards against some of the problems that commonly bedevil performance measurement systems and provides a constructive approach to designing program evaluations. It does this, first, by encouraging the development of a comprehensive approach to measurement and evaluation that gives balanced attention to inputs, processes, and outcomes. Second, it draws attention away from what is easy to measure and toward what is important to measure and encourages early consideration of the means by which all important types of information will be obtained. Third, it explicitly incorporates evaluative criteria and comparisons. Finally, it provides a systematic basis for exploring causal attribution by making provision for the identification and measurement of effects of both program factors and external factors on intended outcomes of the program.

### **Note**

1. The approach has usually been referred to as *program logic matrix*.

## References

- Australian Department of Finance. *Doing Evaluations: A Practical Guide*. Canberra: Australian Government Publishing Service, 1994.
- Bennett, C. F. *Analyzing Impacts of Extension Programs*. Washington D.C.: U.S. Department of Agriculture, 1979.
- Funnell, S. "Developments in the New South Wales Approach to Analysing Program Logic." *Proceedings of the Annual Conference of the Australasian Evaluation Society*, 1990, 2, 247–256.
- Funnell, S. "Reporting the Performance of Public Sector Programs." *Evaluation Journal of Australasia*, 1993, 5(2), 16–37.
- Funnell, S. "An Evaluation of the United Nations Best Practices Collection: Its Strengths and Weaknesses, Accessibility, Use and Impact." Unpublished report commissioned by United Nations AIDS, Geneva, 1999.
- Funnell, S., and Ford, C. "An Evaluation of Solutions to Pollution: The Small Business and Local Government Environmental Management Program of the New South Wales Environment Protection Authority." *Proceedings of the Annual Conference of the Australasian Evaluation Society*, 1998, 2, 391–412.
- Funnell, S., and Harrison, C. "Utility Is in the Eye of the User." *Proceedings of the Annual Conference of the Australasian Evaluation Society*, 1993, pp. 265–292.
- Funnell, S., and Lenne, B. "Clarifying Program Objectives for Program Evaluation." *Program Evaluation Bulletin* 1990, no. 1. Sydney: Program Evaluation Unit of New South Wales Premier's Department, 1990.
- Funnell, S., and Mograby, A. "Evaluating Employment Programs Delivered by Community Organisations." *Proceedings of the Annual Conference of the Australasian Evaluation Society*, 1995, 2, 531–552.
- Lenne, B., and Cleland, H. "Describing Program Logic." *Program Evaluation Bulletin* 1987, no. 2. Sydney: Public Service Board of New South Wales, 1987.
- Lipsey, M. "Theory as Method: Small Theories of Treatments." In L. Sechrest and A. Scott (eds.), *Understanding Causes and Generalizing About Them*. *New Directions for Program Evaluation*, no. 57. San Francisco: Jossey-Bass, 1993.
- Matthews, A., and Funnell, S. "Evaluability Assessment of Drug and Alcohol Programs for Prisoners in New South Wales." Paper presented at the annual conference of the Australasian Evaluation Society, Canberra, July 1987.
- Perrin, S. "Effective Use and Misuse of Performance Measurement." *American Journal of Evaluation*, 1998, 19(2), 367–379.
- Suchman, E. A. *Evaluative Research: Principles and Practice in Public Service and Social Action Programs*. New York: Russell Sage Foundation, 1967.
- Wholey, J. S. *Evaluation and Effective Public Management*. New York: Little, Brown, 1983.
- Winston, J. A. "Performance Indicators—Promises Unmet: A Response to Perrin." *American Journal of Evaluation*, 1999, 20(1), 95–99.

SUE C. FUNNELL is director of Performance Improvement, an Australian consultancy company specializing in program planning and evaluation and is immediate past president of the Australasian Evaluation Society.