5

This chapter highlights the continued salience of Weiss's questions about theory-based evaluation, especially given the often simplistic uses of program theory in evaluation.

Theory-Based Evaluation: Reflections Ten Years On

Patricia J. Rogers

Theory-based evaluation has developed significantly since Carol Weiss's chapter was published ten years ago. In 1997 Weiss pointed to theory-based evaluation being mostly used in the areas of health promotion and risk prevention. The use of program theory is now commonplace and Weiss's chapter has been cited in a wide range of program areas, including evaluations of energy conservation (New York State Energy Research and Development Authority, 2006), comprehensive community-based initiatives (Judge and Bauld, 2001), supported housing (Rog and Randolph, 2002), gaming and simulation (Kriz and Hense, 2006), and anticorruption activities (Marra, 2000).

Along with this proliferation of activity has come a proliferation of terminology. Weiss referred to the use of the different terms *program theory* and *logic models* to refer to essentially similar concepts. Now there is an ever-longer list of labels that have been used, not with consistently distinct definitions, including *theory-based, theory-driven, theory-oriented, theory-anchored, theory-of-change, intervention theory, outcomes hierarchies, program theory,* and *program logic.* Despite the use of the term *program,* the method has been used for planning and evaluating interventions ranging from small projects to multisite projects, multiyear strategies and even whole-of-government processes (for example, Public Service Commission, South Africa, 2003).

One of the biggest changes in the use of program theory since 1997 has been its increasing incorporation in program management processes. This phenomenon had occurred earlier in some places and program areas—for

example, state and federal governments in Australia had mainstreamed the use of program theory in the 1980s and 1990s (Funnell, 1990), and in the area of international development, many aid agencies had required the use of log frames, a particular type of program theory (*Logical Framework,* 1971). Now many organizations and funders require proposals for projects, programs, and policies to include a logic model or program theory from the beginning in recognition of its value for planning and management, as well as for evaluation. This development has been a primary factor in increasing the number of evaluations that use program theory or logic models of some type and the availability of hard copy and online resources to support the use of program theory.

Despite these developments, Weiss's chapter remains highly relevant because of her warnings about the traps and challenges in using program theory. Many so-called program theory evaluations continue to demonstrate one or more of these limitations, and evaluators would do well either to read or to reread Weiss's discussion of three particular issues, and to examine recent examples that have addressed them.

First, there is the issue of the type of program theory that is used. In her review, Weiss found many evaluations were based on an implementation theory that specifies the activities and some intermediate outcomes, rather than a programmatic theory that specifies the mechanisms of change. Many organizations that claim to have adopted program theory still focus only on implementation theory. They have institutionalized a version involving five or so boxes arranged linearly: inputs, activities, outputs, outcomes, and impacts (again, the terminology is not consistent, and sometimes additional boxes are used for need, context, assumptions, and external factors). These versions, which are often referred to as *program logics* or *logic models* rather than program theories, can be a very good start, particularly for organizations unaccustomed to focusing on outcomes rather than activities, or for programs and policies that are being developed as well as evaluated. However, these simple diagrams fall short of the conceptual summary involved in a *programmatic theory* because they do not examine the causal mechanisms involved in programs and policies—simply using unlabeled arrows to show the links between the components—and do not show the different complementary or alternate causal strands involved in achieving the outcomes. They therefore provide little useful information for replication or improvement. There is nothing wrong with starting with the simple intervention theory; in fact, it is probably difficult to articulate a programmatic theory without at least an implicit implementation theory, but there is a great deal wrong with thinking this is all there is to program theory. Evaluators looking for an example of a program theory with clearly labeled mechanisms could look to Pawson and Tilley's classic realist example of how closed-circuit television in parking lots might work to reduce auto theft through the mechanisms of detection, capture, and removal of thieves; through deterrence; through passive surveillance; and so on (Pawson and Tilley, 1997; Tilley, 2000).

The second issue raised in Weiss's chapter is the quality of the program theory. Weiss found that many of the program theories were based only on practitioners' assumptions and logical reasoning and hence were "simplistic, partial, or even downright wrong." Although it can be useful to articulate practitioners' assumptions about how a program is intended to work, this is often not an adequate program theory for planning and evaluating the program. Unfortunately, many examples of program theory evaluation are still based on poor theories—for example, health promotion programs based solely on the discredited theory that improved knowledge will change attitudes and hence behavior. But how realistic is it to expect an evaluation to include the development of a better program theory or a full-fledged research theory? This is the issue behind Stufflebeam's (2001) trenchant criticism: "There really is not much to recommend theory-based evaluation, since doing it right is usually not feasible and since failed or misrepresented attempts can be counterproductive" (p. 80).

Examples of good practice in program theory evaluation demonstrate that it can be feasible and useful to improve the quality of the theory by better logical analysis of alternative causal explanations, better use of existing research theories, and better use of alternative perspectives on how programs work, including understanding how program clients or intended beneficiaries understand it, and through a process of competitive elaboration and testing against the data. For example, the Centre for Communication Programs at Johns Hopkins University has moved on from using a "Knowledge-Attitudes-Practice" model of behavior change to underpin its evaluations, and now uses the frameworks of "Ideation and Communication for Participatory Development." Murray-Johnson and others (2000–01) compared the utility of four different theories: health belief model, theory of reasoned action, extended parallel process model, and social cognitive theory.

The third issue is how program theory is used in evaluations. In 1997 Weiss observed that many evaluators developed program theory but then did not use it at all to guide the evaluation. This seems less of a problem now, in part because of the widespread practice of developing performance measurement or operationalizing variables on the basis of the program theory. However, the ways program theory are used to guide evaluation are often simplistic. In many cases the evaluation consists only of gathering evidence about each of the components in the logic model, and answering the question "Did this happen?" about each one. Although this can be useful in reporting on the program or policy in terms of a coherent performance story, it does not use the full potential of program theory evaluation, including its ability to address the issue of causal attribution.

There are three possible responses to the challenge of causal attribution. One is to give up the attempt to use program theory evaluation for this purpose, deciding to use it only "to improve, not to prove." Another option is to combine program theory with other methods for causal

attribution—for example, Cook (2000) discussed "the false dichotomy" between experimental designs and program theory, and how program theory could be used to design better experiments. Alternatively, a Popperian approach can be taken, and program theory can be used to develop "testable hypotheses," which are then investigated using nonexperimental methods (Pawson and Tilley, 1997; Tilley, 2000). Using the variations among different levels of implementation and different contexts for implementation not as "noise" to be screened out but rather as opportunities to test hypotheses, one can build a stronger case that the program not only contributes to the observed outcomes but also to explaining how. In an international climate of increasing focus on rigorous methods for impact evaluation, this may be the aspect of program theory evaluation most deserving of further development.

References

Cook, T. D. "The False Choice Between Theory-Based Evaluation and Experimentation." In P. J. Rogers, T. A. Hasci, A. Petrosino, and T. A. Huebner (eds.), *Challenges and Opportunities in Program Theory Evaluation.* New Directions in Evaluation, no. 87. San Francisco: Jossey-Bass, 2000.

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.

Judge, K., and Bauld, L. "Strong Theory, Flexible Methods: Evaluating Comprehensive Community-Based Initiatives." *Critical Public Health,* 2001, *11*(1), 19–38.

Kriz, W. C., and Hense, J. U. "Theory-Oriented Evaluation for the Design of and Research in Gaming and Simulation." *Simulation & Gaming,* 2006, *37*(2), 268–283.

"Logical Framework." *An unpublished approach and training materials developed for U.S. Agency for International Development,* 1971.

Marra, M. "How Much Does Evaluation Matter? Some Examples of the Utilization of the Evaluation of the World Bank's Anti-Corruption Activities." *Evaluation,* 2000, *6*(1), 22–36.

Murray-Johnson, L., and others. "Using Health Education Theories to Explain Behaviour Change: A Cross-Country Analysis." *International Quarterly of Community Health Education,* 2000–01, *20*(4), 323–345.

New York State Energy Research and Development Authority. *Quarterly Report to the Department of Public Service. Quarter ending Sept. 30, 2006.* Retrieved Feb. 5, 2007, from http://www.nyserda.org/Energy_Information/SBC/sbcsept2006.pdf.

Pawson, R. D., and Tilley, N. *Realistic Evaluation.* London: Sage, 1997.

Public Service Commission, South Africa. *The PSC's Public Administration Monitoring and Evaluation System: First Consolidated Report.* 2003. Retrieved Feb. 5, 2007, from http://www.pmg.org.za/docs/2003/appendices/031112psc.ppt.

Rog, D. J., and Randolph, F. L. "A Multisite Evaluation of Supported Housing: Lessons Learned from Cross-Site Collaboration." In J. M. Herrell and R. B. Straw (eds.), *Conducting Multiple Site Evaluations in Real-World Settings.* New Directions in Evaluation, no. 94. San Francisco: Jossey-Bass, 2002.

Stufflebeam, D. L. *Evaluation Models.* New Directions for Evaluation, no. 89. San Francisco: Jossey-Bass, 2001.

Tilley, N. *Realistic Evaluation: An Overview.* 2000. Retrieved Feb. 5, 2007, from http://www.danskevalueringsselskab.dk/pdf/Nick%20Tilley.pdf.

Weiss, C. "Theory-Based Evaluation: Past, Present and Future." In D. J. Rog and D. Fournier (eds.), *Progress and Future Directions in Evaluation: Perspectives on Theory, Practice and Methods.* New Directions for Evaluation, no. 76, San Francisco: Jossey-Bass, 1997.

PATRICIA J. ROGERS is associate professor in public sector evaluation and director of CIRCLE (Collaborative Institute for Research, Consulting and Learning) at the Royal Melbourne Institute of Technology, Australia.

Theory-based evaluation examines conditions of program implementation and mechanisms that mediate between processes and outcomes as a means to understand when and how programs work.

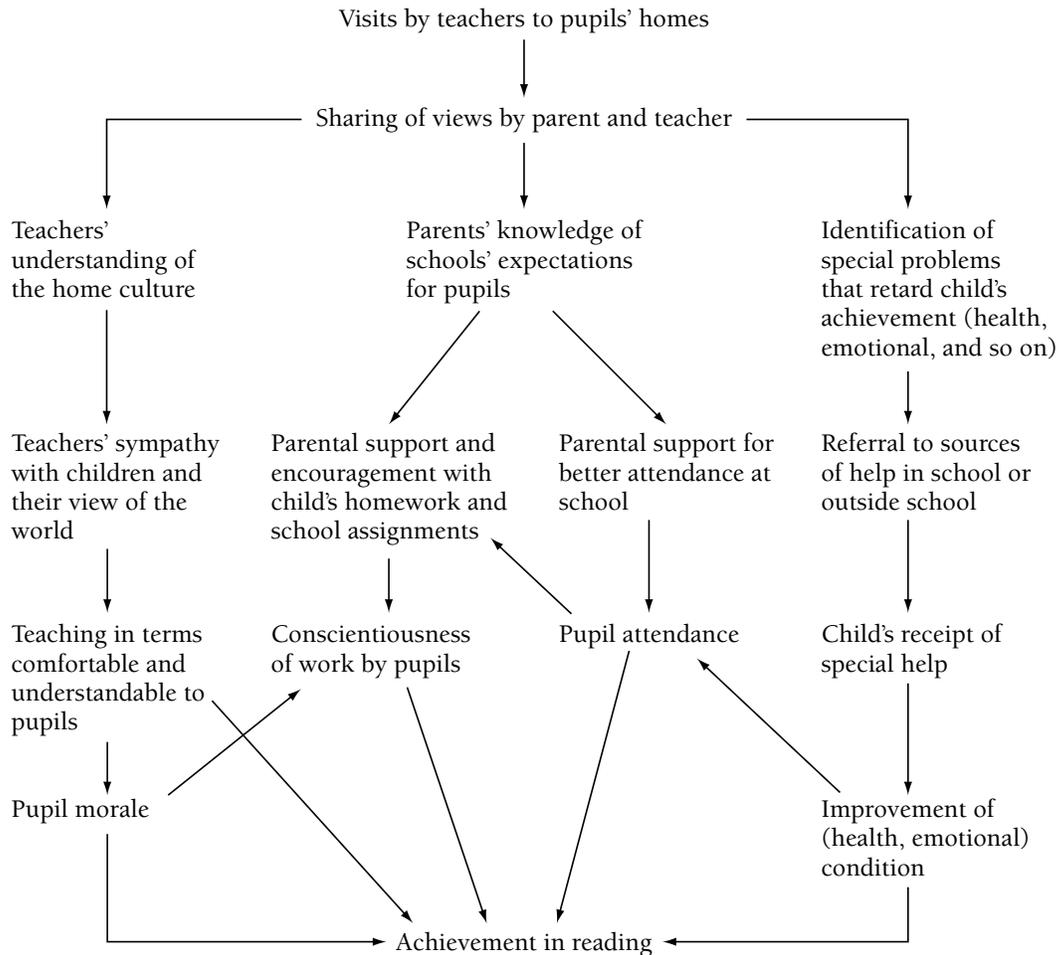# Theory-Based Evaluation: Past, Present, and Future

Carol H. Weiss

Theory-based evaluation has surged to attention in recent years. Evaluators are writing about it, and evaluations structured around theory are beginning to appear in numbers in the literature.

The Past

The concept of theory-based evaluation has been around for over twenty-five years. In the spring-summer 1996 issue *Evaluation Practice* published two early papers—an excerpt from my 1972 book *Evaluation Research* and Fitz-Gibbon and Morris (1975)—along with a historical introduction by Blaine Worthen (1996). I have been trying to go back further. In his 1967 book *Evaluative Research*, Edward Suchman referred several times to the notion of programs' theories. Suchman discussed two kinds of reasons for an unsuccessful program: failure of the program to put the intended activities into operation (implementation failure) and failure of the activities to bring about the desired effects (theory failure). My 1972 book offered the first discussion Worthen and I have found of the central idea of basing evaluation on the program's theory. I included a diagram of several alternative theories on which a program of teachers' home visiting might be based. See Figure 5.1. I called the subject a "process model," and I urged that the evaluator collect data on the posited links.

In succeeding years there were a few papers on the subject. Joe Wholey's work on evaluability assessment stressed the need to find out whether the implicit theory underlying a program made sense (Wholey, 1979, 1983). Wholey's idea was that prior to the start of a formal study, the evaluator should analyze the logical reasoning that connected program inputs to

Figure 5.1. Theory of a Program of Teacher Home Visits

Visits by teachers to pupils' homes

Sharing of views by parent and teacher

| Teachers' understanding of the home culture | Parents' knowledge of schools' expectations for pupils | Identification of special problems that retard child's achievement (health, emotional, and so on) |

| Teachers' sympathy with children and their view of the world | Parental support and encouragement with child's homework and school assignments | Parental support for better attendance at school | Referral to sources of help in school or outside school |

| Teaching in terms comfortable and understandable to pupils | Conscientiousness of work by pupils | Pupil attendance | Child's receipt of special help |

Pupil morale

Improvement of (health, emotional) condition

Achievement in reading

Source: Weiss, 1972, p. 50.

desired outcomes to see whether there was a reasonable likelihood that goals could be achieved.

Huey-Tsyh Chen and Peter Rossi discussed the idea in a series of publications (Chen and Rossi, 1980, 1983, 1987; Chen, 1990, 1994). Their addition to the discussion included the idea that the theory should be a social science theory, not just a series of ad hoc logical premises. Chen (1990) also distinguished between normative theory and causal theory. Normative theory "provides guidance on what goals and outcomes should be pursued or examined" (p. 43), whereas causal theory was the set of assumptions about how the program works. Causal theory is what most of the previous authors and most of the subsequent ones have talked about.

By the late 1980s, program-based evaluation was becoming a popular idea. Although not many examples of theory-based evaluation were yet published, the ideas were becoming increasingly visible. Leonard Bickman edited two issues of New Directions for Program Evaluation (1987 and 1990)

that elaborated and advocated the strategy, and Lee Sechrest and A. G. Scott edited one in 1993. Lipsey wrote several articles, one explicating four different versions of program theory. Several articles addressed the subject of how to analyze data that followed the underlying assumptions of a program through time (Judd and Kenny, 1981; Smith, 1990; Marquart, 1990; Trochim, 1985). Dozens of papers appeared.

At the same time, other writers were writing about logic models. Logic models seem to be similar to program theories; at least they are if the word *theory* does not overwhelm us. If we take the word *theory* to mean the professional logic that underlies a program, then the two concepts appear to be much the same.

I wrote a paper on theory-based evaluation in 1995, published in what I thought would be an obscure book, that has received considerable attention (Weiss, 1995). The idea of basing evaluation on programs' theories of change in community-based programs received a warm welcome among evaluators and sponsors of these kinds of programs. One reason seems to be that it promised (or at least hinted at a promise) that theory-based evaluation could strengthen the validity of evaluations when random assignment is impossible, as it is in place-based programming. If the evaluation can show the series of micro-steps that lead from inputs to outcomes, then causal attribution *for all practical purposes* seems to be within reach. Although such an evaluation cannot rule out all the threats to validity we have come to know and love, it has the advantage (if things go well) of showing what processes lead to the outcomes observed; if some of the posited steps are not borne out by the data, then the study can show where the expected sequence of steps breaks down.

The Present

If the past was a period when people developed and elaborated the idea of theory-based evaluation, the present is a time when evaluators are putting the ideas into practice. A graduate student, Jo Birckmayer, and I have done a search for theory-based evaluations in the periodical literature. Mark Lipsey sent me abstracts of studies that his staff coded as "integrated theory" in the evaluation data base he collected some years ago. We are still collecting cases, and I will welcome your papers and articles. We have now inspected about thirty studies that have at least a modicum of theory orientation. This preliminary inspection has given rise to some tentative ideas.

First, quite a number of articles claim that the *programs* are theory-based, and depending on how forgiving our definition of *theory* is, many of them do seem to have a set of coherent ideas providing the basis for intervention. But many of the evaluations do not follow through on that theory; they do not collect data on crucial theoretical constructs. For example, Gottfredson (1987) reports the evaluation of a program to reduce school disorder that is reportedly based on organization development principles.

However, the evaluation as published does not look at the organizational development aspects. Similarly, Campbell and Ramey (1995) evaluate an early childhood intervention program that is based on theories about enhancement of early cognitive development and the ensuing development of academic confidence, motivation, and success. The evaluation does not track the steps of the program theory.

But a number of evaluations trace the emergence of the stages posited in theory. A few of the theories are simple, what Lipsey and Pollard (1989) would probably call "one-step" theories. Sheard, Marini, Bridges, and Wagner (1976) report on lithium given to aggressive prisoners in a medium-security institution. The evaluation examined the number of violent infractions that they committed. The evaluation thus tested the theory that lithium would reduce violence. The mechanism in this case is the physiological effects that lithium produces in the body. (Actually, the theory is not simple. But the physiological element, the action of lithium in the body, has been investigated by bio-medical researchers and need not form part of the evaluation.)

But some are more complex theories, and the evaluations make valiant efforts to follow them along their course. Cohen and Rice (1995) trace the effects of involving parents in prevention of adolescent substance abuse prevention. They find that parents were difficult to engage, and even when parents attended the program, they did not believe that their children's friends used drugs and so did not monitor their friendships.

Much of the work in theory-based evaluation is going on in the fields of health promotion and risk prevention. Evaluators are using theory-based approaches in programs to reduce smoking, stress, risky sexual behavior, drug abuse, adolescent pregnancy, and similar ills (for example, see Goodman and others, 1996). We have also located a few theory-based studies in mental health and health care (for example, see Bickman, 1983).

Another interesting thing is that evaluators abroad are embracing the approach. When I posted a request for examples of theory-based evaluation on the listserver EVALTALK, I received papers from Rush and Ogborne (1991), in Canada; Torvatn (1995), in Norway; Kelly and Maloney (1992), in Scotland; and Milne (1995), in Australia.

Conditions Conducive to Theory-Based Evaluation. When evaluators adopt a theory-based approach, it is often for one of two reasons. The first is that the evaluator is also the program developer. A program designer, usually an academic, is engaged in a cycle of program development to deal with a particular problem. He or she develops theory, operationalizes the theory in a set of program activities, tests the program and therefore the underlying theory through evaluation, and revises the intervention. Such a cycle has a long and honorable history in several fields of what I would call applied social psychology. Evaluation is part of the ongoing series of activities by which the intervention takes shape. When the work is not that of a single individual but part of the work of an academic center devoted to the

design of interventions, evaluation becomes a key feature of both theory development and program modification.

A good example is the work of Sandler and others (1992) at the Program for Prevention Research at Arizona State University who iteratively developed a family bereavement program for youngsters who had lost a parent. Based on prior research and pilot testing, the program aimed to influence four "mediators," hypothesized as implicated in child symptomatology: parental demoralization, parental warmth, discussion of grief, and stable positive events. The evaluation then collected data on the extent to which the program was associated with changes in the four mediators, and went on to study the child's psychosocial symptomatology.

Another condition that promotes a theory-based approach to evaluation is conscientious theory-based development of the program. Health promotion and risk prevention are fields where program planning is well developed: designers tend to spell out their theoretical assumptions in thoughtful detail and build programs on that foundation. Therefore it becomes easy for the evaluator to follow the tracks of theory in the evaluation. Some of the theories are relatively traditional and well established, such as social learning theory. A health promotion program provides knowledge (for example, about methods to break the smoking habit), which leads to a change in motivation and intention (willingness to try to reduce smoking), which leads to a change in practice (cessation of smoking). The change in practice is assumed to lead to the ultimate outcome, which may be reduction in cardiovascular disease. In addition, social-reinforcement theory may call for provision of social supports to encourage and sustain smoking cessation.

Social-cognitive theories of several kinds are prevalent in risk prevention. The operative mechanisms are expected to be change in knowledge, change in attitude, increase in feelings of self-efficacy, higher motivation, mastery of skills, and heightened sense of responsibility, which lead to intentions to change behavior and so on to the desired outcomes. Evaluations follow the anticipated sequence of changes over time.

Program Theory and Implementation Theory. While the subject of theory-based evaluation has been gaining adherents and attention, it has also gained confusion. As is frequently the case with an emerging idea, people have attached their own understandings to the same words. Therefore we have different varieties of recommended evaluation strategies all sailing under the flag of "theory-based evaluation." Figures 5.2–5.5 are diagrams that evaluators have used to explain how their studies have been guided by theory. The diagrams differ in level of specificity, complexity, and type of pictorial display. They also incorporate two different elements of theory, what I will call implementation theory and programmatic theory.

Implementation theory focuses on how the program is carried out. The theoretical assumption it tests is that if the program is conducted as planned, with sufficient quality, intensity, and fidelity to plan, the desired results will be forthcoming. The emphasis is on what Suchman would have

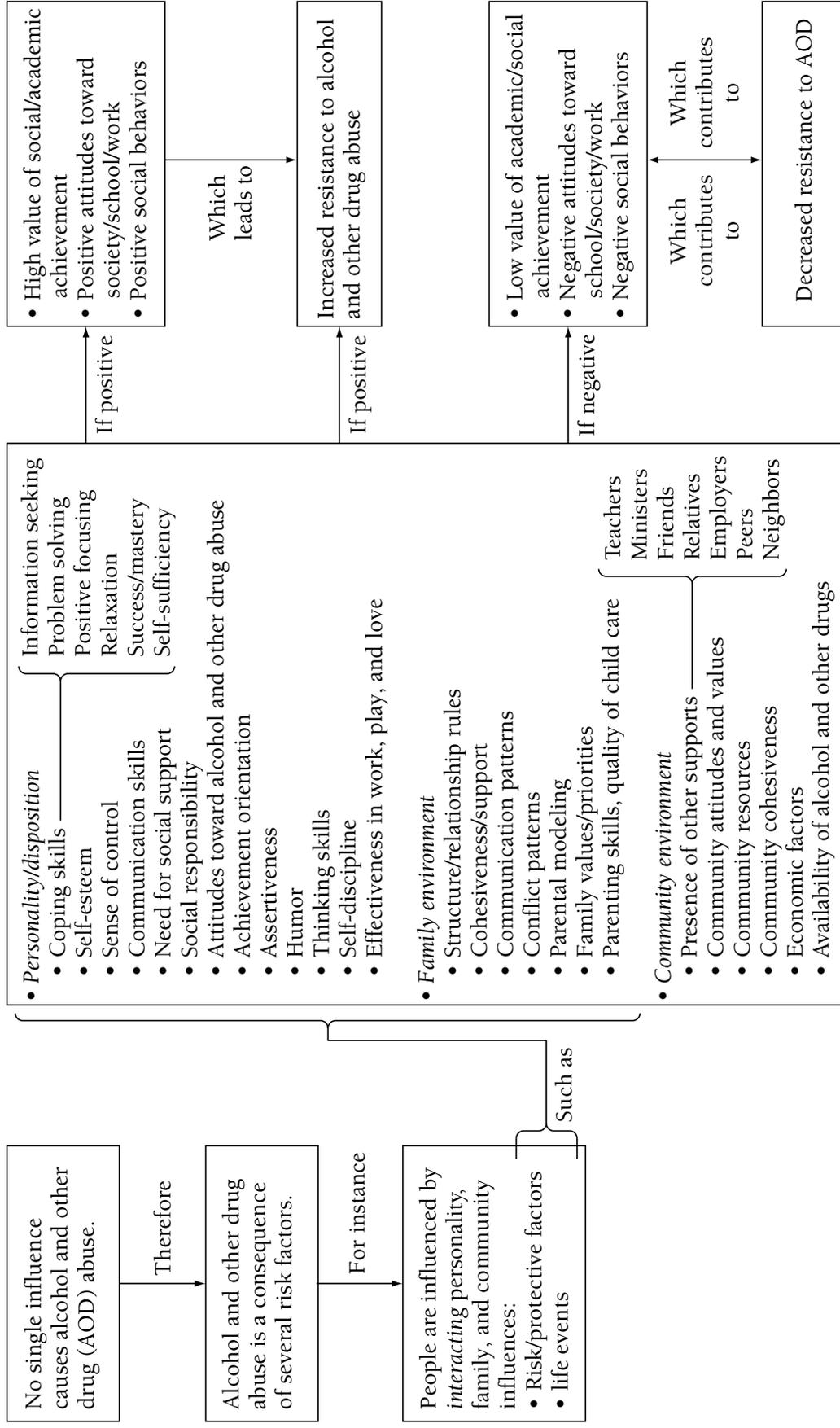**Figure 5.2. Theory of an Antismoking Program**
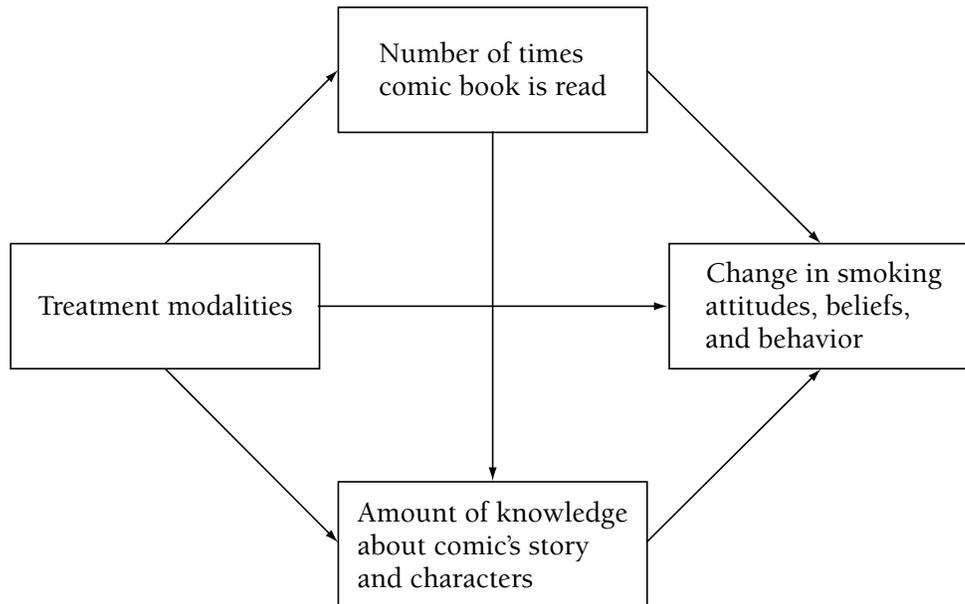
Source: Chen, Quane, Garland, and Marcin, 1988.

called implementation failure/success. Programmatic theory, on the other hand, deals with the *mechanisms* that intervene between the delivery of program service and the occurrence of outcomes of interest. It focuses on participants' responses to program service. The mechanism of change is not the program activities per se but the response that the activities generate. For example, in a contraceptive counseling program, if counseling is associated with reduction in pregnancy, the cause of change might seem to be the counseling. But the mechanism is not the counseling; that is the program activity, the program process. The mechanism might be the knowledge that participants gain from the counseling. Or it might be that the existence of the counseling program helps to overcome cultural taboos against family planning; it might give women confidence and bolster their assertiveness in sexual relationships; it might trigger a shift in the power relations between men and women. These or any of several other cognitive, affective, social responses could be the mechanisms leading to desired outcomes.

Similarly, programs that aim to teach students understanding of other cultures may assume that any good results observed are due to the teaching. But teaching is not the mechanism. The mechanism is what students get from the teaching—knowledge or heightened interest, motivation, even anxiety. An evaluation that attempts to track the theoretical underpinnings of the program has to devise ways to define and measure the psychosocial, physiological, economic, sociological, organizational, or other processes that intervene between exposure to the program and participant outcomes.

Much evaluation that is purportedly theory-based actually examines outcomes in terms of implementation variables. For example, McGraw and

Figure 5.3. Community Prevention Model for Alcohol and Other Drug Abuse

No single influence causes alcohol and other drug (AOD) abuse.

Therefore

Alcohol and other drug abuse is a consequence of several risk factors.

For instance

People are influenced by *interacting* personality, family, and community influences:
• Risk/protective factors
• life events

Such as

• *Personality/disposition*
  • Coping skills — Information seeking / Problem solving / Positive focusing / Relaxation / Success/mastery / Self-sufficiency
  • Self-esteem
  • Sense of control
  • Communication skills
  • Need for social support
  • Social responsibility
  • Attitudes toward alcohol and other drug abuse
  • Achievement orientation
  • Assertiveness
  • Humor
  • Thinking skills
  • Self-discipline
  • Effectiveness in work, play, and love

• *Family environment*
  • Structure/relationship rules
  • Cohesiveness/support
  • Communication patterns
  • Conflict patterns
  • Parental modeling
  • Family values/priorities
  • Parenting skills, quality of child care

• *Community environment*
  • Presence of other supports — Teachers / Ministers / Friends / Relatives / Employers / Peers / Neighbors
  • Community attitudes and values
  • Community resources
  • Community cohesiveness
  • Economic factors
  • Availability of alcohol and other drugs

If positive

• High value of social/academic achievement
• Positive attitudes toward society/school/work
• Positive social behaviors

Which leads to

Increased resistance to alcohol and other drug abuse

If positive

If negative

• Low value of academic/social achievement
• Negative attitudes toward school/society/work
• Negative social behaviors

Which contributes to

Which contributes to

Decreased resistance to AOD

Figure 5.4. Case Management Theory

Case management activities ⟶ Intermediate outcomes ⟶ Distal outcomes

Treatment planning:
  Comprehensive initial assessment

  Client and family participation

  Multidisciplinary team participation

Linkage:
  Regular communication among client, family, providers, and case manager

  Assistance with arrangements for treatment

  Single point of contact for family and providers

Monitoring:
  Review of treatment progress

  Review of restrictiveness of care

Advocacy:
  Assistance to family in negotiating with providers

  Respect for child and family rights

More appropriate treatment plan

Plan that is acceptable to child and family

Timely review of treatment process

Accessible services

Efficient use of services

Improved mental health outcomes

Quicker recovery

Increased client and family satisfaction

Reduced costs

Source: Bryant and Bickman, 1996, p. 123.

others (1996) evaluated a program to change students' dietary knowledge and food choices so as to reduce the risk of cardiovascular disease. The program, called CATCH (Child and Adolescent Trial for Cardiovascular Health), had several components, but the part recently reported related to the classroom component. The outcomes studied were the children's dietary knowledge, self-confidence that they could select better foods, and intentions to eat more wisely. See Figure 5.6. These were analyzed against input variables (mainly student and teacher characteristics) and program processes, such as the extent to which teachers completed the full course of CATCH classroom activities ("dose") and the degree to which teachers modified the activities ("fidelity" to plan). The analysis led to conclusions about

**Figure 5.5. Program Theory of an Employment Training Program**
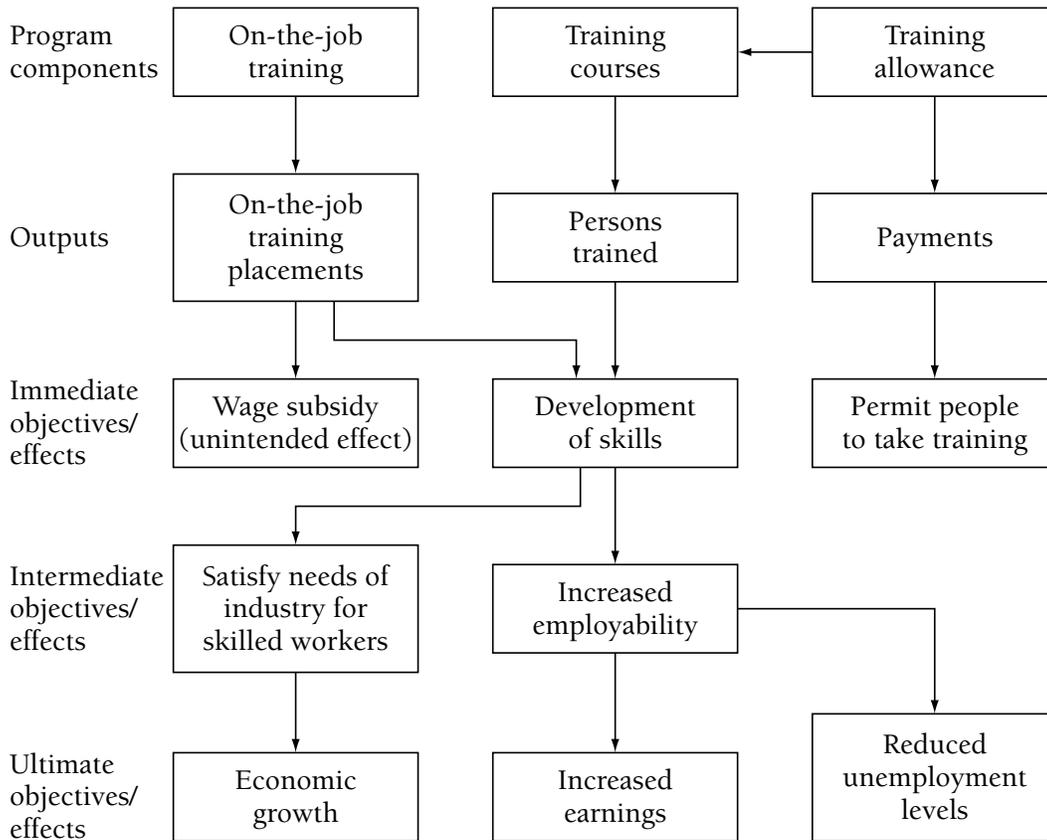
Source: Auditor General of Canada, 1981, p. 14.

the extent to which program activities as defined by CATCH planners led to desired health knowledge and intended behaviors.

This study makes excellent use of process measures in analyzing outcomes. But it does not provide a test of the programmatic theory of the program, at least in this paper. The theory underlying the program is described as "modification of psychosocial factors … lead[ing] to changes in risk-factor behaviors" (McGraw and others, 1996, p. 292). The evaluation reported does not address the modification of psychosocial factors—that is, it does not inquire into the mechanisms by which change is brought about. Similarly, Pentz and others (1990) do an excellent analysis of the relation between the implementation of a drug abuse prevention program and outcomes for adolescents. The emphasis is strictly on the extent to which implementation variables (such as exposure, adherence, fidelity, and amount of implementation) were associated with outcomes.

The difference between program theory and implementation theory is analogous to the distinction between mediator and moderator variables (Baron and Kenny, 1986). Mediator and moderator variables are both third variables that affect the relation between an independent and a dependent

**Figure 5.6. Program Theory Model of the Child and Adolescent Trial
for Cardiovascular Health**
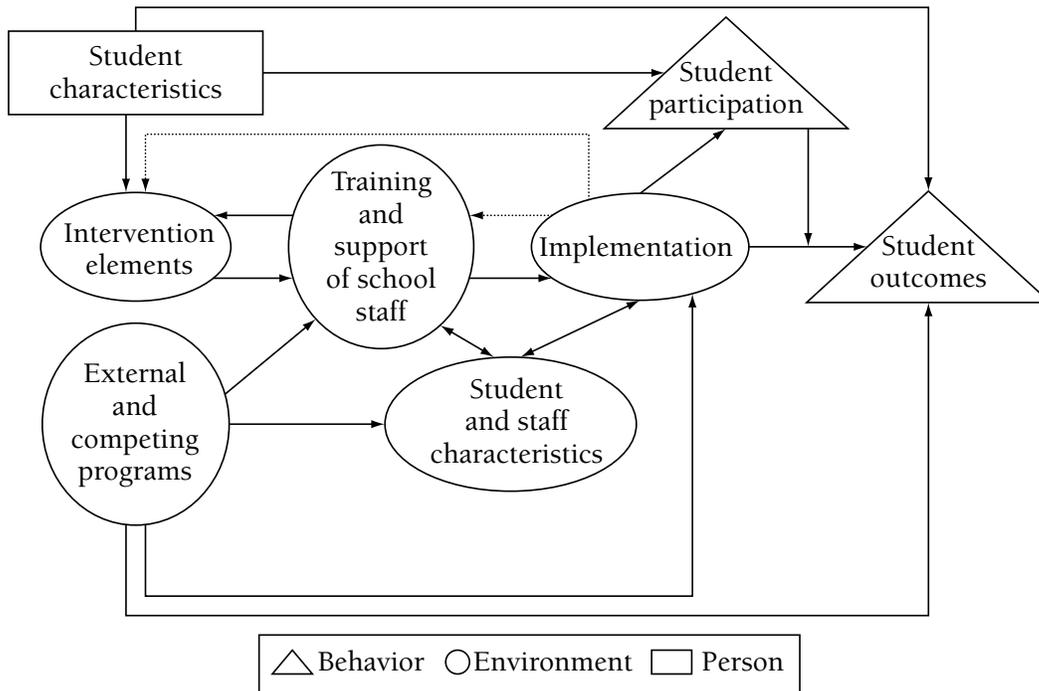
Source: McGraw and others, 1996, p. 294.

variable. The moderator variable is a characteristic, such as gender or frequency of exposure, the subcategories of which have different associations with the outcome variable. Girls do better than boys, or those who attend the program regularly do better than those who attend infrequently. On the other hand, a mediator variable "represents the generative mechanism through which the focal independent variable is able to influence the dependent variable of interest" (Baron and Kenny, 1986, p. 1173). That is, the moderator helps to explain which features of persons or situations have the strongest relationship to the outcome; mediators help to explain how the process works. The concepts are similar to the concepts of implementation theory and programmatic theory.

In most programs both kinds of theories will be implicated. Elsewhere I have used the term *theories of change evaluation* for evaluations that explore both elements.

The Future

Theory-based evaluation is demonstrating its capacity to help readers understand how and why a program works or fails to work. Knowing only outcomes, even if we know them with irreproachable validity, does not tell us enough to inform program improvement or policy revision. Evaluation needs to get inside the black box and to do so systematically.

One of the side benefits of this kind of evaluation is its contribution to wiser program planning even before the evaluation gets under way. When evaluators are involved in the planning phase, they have the opportunity to elicit program designers' own theories about how the program is expected to work. They can help designers to disaggregate the assumptions into the mini-steps that are implied and to confront the leaps of faith and questionable reasoning that are often involved. Evaluators can also offer theories and promising hypotheses based in the social sciences and evidence from prior evaluations that show which kinds of theories hold up in practice. In all these ways, evaluators can become profitably engaged in helping to plan programs that are rooted in better conceived premises.

Challenges lie ahead. One of the immediate needs is for better measures. Through repeated tests, evaluators have made great strides in developing valid measures of outcomes. But measurement of mediating variables is a relatively recent activity in most academic subfields, and evaluators will have to learn how to do it better. Given the large number of variables that are implied in many theories of change, measurement error makes it difficult to identify significant associations among variables—even when they are present. The field needs advances in measurement of mediating mechanisms.

Probably the central need is for better program theories. Evaluators are currently making do with the assumptions that they are able to elicit from program planners and practitioners or with the logical reasoning that they bring to the table. Many of these theories are elementary, simplistic, partial, or even outright wrong. Evaluators need to look to the social sciences, including social psychology, economics, and organization studies, for clues to more valid formulations, and they have to become better versed in theory development themselves. Better theories are important to evaluators as the backbone for their studies. Better theories are even more essential for program designers, so that social interventions have a greater likelihood of achieving the kind of society we hope for in the twenty-first century.

Theory-based evaluation can pursue two different strands in the coming years. One path is to build more detailed program theories, so that evaluations can trace micro-steps of process all along the pathways that lead to program effects. This is like the theory of the teacher home visiting program in Figure 5.1. Each program activity and each participant response is followed along the hypothesized chain of events. There is real promise in this direction, if we iteratively test the linkages between steps and substantially improve our knowledge of how processes work. This kind of evaluation will have much to tell program funders, managers, and practitioners about what works and what does not work under a range of different conditions.

The other path that theory-based evaluation can pursue is to limit the theory to the one or two central assumptions embedded in each program. They should be premises that are significant for program success, common across a range of programs, and particularly problematic. For example, many interventions are now based on the assumption that empowering

residents of low-income neighborhoods to help plan social, economic, and educational programs for their community will improve the nature of services. A central premise of these kinds of programs is that residents of the community will plan and allocate resources in ways that are more responsive to need than the professional systems of the past.

Or consider programs to make major alterations in teachers' behavior through staff development programs. The assumption here is that short-term training will be able to modify teaching patterns developed over years of education and professional practice. Another example would be efforts to change such behaviors as low school grades, delinquency, and domestic violence through programs that seek to raise self-esteem and self-confidence. Theory-based evaluation could be directed at investigating the viability of such central theoretical premises.

Evaluations that test such macro-theoretical assumptions will require multiple cases and will be difficult to do. The quest will probably be more appropriate for meta-analysis than for single studies (Cook and others, 1992). Furthermore, they will not hold much interest for funders or practitioners who are wedded to the premise being scrutinized. Professionals who run short-term staff development programs for teachers are not going to be receptive to studies that question whether staff development is a sensible approach to changing teacher behavior.

Still there are audiences who want to know. Program sponsors and funders, whether foundations or government agencies, should have an intense interest in whether the strategy in which they are investing is feasible across a range of conditions. A meta-analysis of evaluations that have measured and examined the same central assumption should have important news to report. It should be able to give insight into the conditions under which these hypotheses hold and the conditions under which they result in shortfalls of varying dimensions. My long-range hope is that evaluation will not only be based on theory but also *contribute* to the cumulation of theoretical knowledge.

Conclusion

Looking at conditions of program implementation that are associated with better outcomes is a real contribution to the improvement of programs. These process/outcome evaluations show which program processes yield positive benefits. Evaluations will provide even more valuable information when they address the mechanisms that mediate between processes and outcomes. Theory-based evaluations attend not only to what programs do but also to how participants respond. Such evaluations are not easy to do, but there are circumstances in which evaluators should proceed in that direction. As a starting point, we need plausible theories. We need to make the maximum use of logical reasoning, practitioner wisdom, prior evaluations, and social science research to generate program theories and then use our collective evaluation work to test them under realistic operating conditions.

References

Auditor General of Canada. *Audit Guide*: Auditing of Procedures for Effectiveness. Ottawa: Office of Auditor General, 1981.

Baron, R. M., and Kenny, D. A. "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology,* 1986, *51*(6), 1173–1182.

Bickman, L. "The Evaluation of Prevention Programs." *Journal of Social Issues,* 1983, *39*(l), 181–194.

Bickman, L. (ed.). *Using Program Theory in Evaluation*. New Directions for Program Evaluation, no. 33. San Francisco: Jossey-Bass, 1987.

Bickman, L. (ed.). *Advances in Program Theory*. New Directions for Program Evaluation, no. 47. San Francisco: Jossey-Bass, 1990.

Bryant, D. M., and Bickman, L. "Methodology for Evaluating Mental Health Case Management." *Evaluation and Program Planning,* 1996, *19*, 5–16.

Campbell, F. A., and Ramey, C. T. "Cognitive and School Outcomes for High-Risk African-American Students at Middle Adolescence: Positive Effects of Early Intervention." *American Educational Research*, 1995, *32*(4), 743–772.

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

Chen, H.-T. "Theory-Driven Evaluations: Needs, Difficulties, and Options." *Evaluation Practice,* 1994, *15,* 79–82.

Chen, H.-T., Quane, J., Garland, T. N., and Marcin, P. "Evaluating an Antismoking Program: Diagnostics of Underlying Causal Mechanisms." *Evaluation and the Health Professions*, 1988, *11*(4), 441–464.

Chen, H.-T., and Rossi, P. H. "The Multi-Goal, Theory-Driven Approach to Evaluation: A Model Linking Basic and Applied Social Science." *Social Forces*, 1980, *59*, 106–122.

Chen, H.-T., and Rossi, P. H. "Evaluating with Sense: The Theory-Driven Approach." *Evaluation Review*, 1983, 7, 283–302.

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

Cohen, D. A., and Rice, J. C. "A Parent-Targeted Intervention for Adolescent Substance Use Prevention: Lessons Learned." *Evaluation Review,* 1995, *19*(2), 159–180.

Cook, T. D., Cooper, H., Cordray, D. S., Hartmann, H., Hedges, L. V., Light, R. J., Louis, T. A., and Mosteller, F. *Meta-Analysis for Explanation*: *A Casebook*. New York: Russell Sage Foundation, 1992.

Fitz-Gibbon, C. T., and Morris, L. L. "Theory-Based Evaluation." *Evaluation Comment,* 1975, *5*(l), 1–4.

Goodman, R. M., and Wandersman, A. "FORECAST: A Formative Approach to Evaluating Community Coalitions and Community-Based Initiatives. *American Journal of Community Psychology*, 1994, special issue, 6–25.

Goodman, R. M., Wandersman, A., Chinman, M., Imm, P., and Morrissey, E. "An Ecological Assessment of Community-Based Interventions for Prevention and Health Promotion: Approaches to Measuring Community Coalitions." *American Journal of Community Psychology*, 1996.

Gottfredson, D. C. "An Evaluation of an Organization Development Approach to Reducing School Disorder." *Evaluation Review*, 1997, *11*(6), 739–763.

Judd, C. M., and Kenny, D. A. "Process Analysis: Estimating Mediation in Treatment Evaluations." *Evaluation Review,* 1981, *5*(5), 602–619.

Kelly, M. P., and Maloney, W. A. "A Behavioural Modelling Approach to Curriculum Development and Evaluation of Health Promotion for Nurses." *Journal of Advanced Nursing,* 1992, *17,* 544–547.

Lipsey, M. W., and Pollard, J. A. "Driving Toward Theory in Program Evaluation: More Models to Choose From." *Evaluation and Program Planning,* 1989, *12*, 317–328.

Marquart, J. M., "A Pattern-Matching Approach to Link Program Theory and Evaluation Data." In L. Bickman (ed.), *Advances in Program Theory*. New Directions for Program Evaluation, no. 47. San Francisco: Jossey-Bass, 1990.

McGraw, S. A., Sellers, D. E., Stone, E. J., Bebchuk, J., Edmundson, E. W., Johnson, C. C., Bachman, K. J., and Luepker, R. V. "Using Process Data to Explain Outcomes: An Illustration from the Child and Adolescent Trial for Cardiovascular Health (CATCH)." *Evaluation Review,* 1996, *20*(3), 291–312.

Milne, C. "Using Program Logic as a Practical Evaluation Tool: Case Studies from an Australian Evaluator." Paper given at First International Evaluation Conference, Vancouver, 1995.

Pentz, M. A., Trebow. E. A., Hansen, W. B., MacKinnon, D. P., Dwyer, J. H., Johnson, C. A., Flay, B. R., Daniels, S., and Cormack, C. "Effects of Program Implementation on Adolescent Drug Use: The Midwestern Prevention Project (MPP)." *Evaluation Review,* 1990, *14*(3), 264–289.

Quane, J. "Back from the Future: Can Evaluation Survive Dissension in the Ranks." Paper prepared for American Evaluation Association meeting, Boston, Nov. 1994.

Rush, B., and Ogborne, A. "Program Logic Models: Expanding Their Role and Structure for Program Planning and Evaluation." *Canadian Journal of Program Evaluation*, 1991, *6*(2), 95–106.

Sandler, I. N., West, S. G., Baca, L., Pillow, D. R., Gersten, J. C., Rogosch, F., Virdin, L., Beak, J., Reynolds, K. D., Kallgren, C., Tein, J-Y., Kriege, G., Cole, E., and Ramirez, R. "Linking Empirically Based Theory and Evaluation: The Family Bereavement Program." *American Journal of Community Psychology,* 1992, *20*(4), 491–521.

Sechrest, L. B., and Scott, A. G. (eds.). "Understanding Causes and Generalizing About Them." New Directions for Program Evaluation, no. 57. San Francisco: Jossey-Bass, 1993.

Sheard, M. H., Marini, J. L., Bridges, C. L., and Wagner, E. "The Effect of Lithium on Impulsive Aggressive Behavior in Man." *American Journal of Psychiatry,* 1976, *133* (12), 1409–1413.

Smith, N. L. "Using Path Analysis to Develop and Evaluate Program Theory." In L. Bickman (ed.), *Advances in Program Theory*. New Directions for Program Evaluation, no. 47. San Francisco: Jossey-Bass, 1990.

Suchman, E. *Evaluative Research*. New York Russell Sage Foundation, 1967.

Torvatn, H. "Chains of Reasoning: An Evaluation Tool." Trondheim, Norway: SINTEF-IFTM paper, 1995.

Trochim, W.M.K. "Pattern Matching, Validity, and Conceptualization in Program Evaluation." *Evaluation Review*, 1985, *9*, 575–604.

Weiss, C. H. *Evaluation Research: Methods for Assessing Program Effectiveness*. Englewood Cliffs N. J.: Prentice Hall, 1972.

Weiss, C. H. "Nothing as Practical as Good Theory." In J. Connell, A. Kubisch, L. B. Schorr, and C. H. Weiss (eds.), *New Approaches to Evaluating Community Initiatives*. New York: Aspen Institute, 1995.

Wholey, J. S. *Evaluation: Promise and Performance*. Washington, D.C.: Urban Institute, 1979.

Wholey, J. S. *Evaluation and Effective Public Management*. Boston: Little, Brown, 1983.

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

At the time of publication CAROL H. WEISS *was a professor of education at the Harvard Graduate School of Education.*

To refer to each group-based self, the theory chose the term identity, asserting that persons have as many identities as distinct networks of relationships in which they occupy positions and play roles. In Identity Theory usage, social roles are expectations attached to positions occupied in networks of relationships; identities are internalized role expectations. The theory asserts that role-choices are a function of identities so conceptualized, and that identities within self are organized in a salience hierarchy reflecting the importance of hierarchy as an organizational principle in socie Evaluators who wish to integrate theory-based approaches into their practices are encouraged to pursue additional readings (including those referenced in this document) and, as appropriate, to seek additional support in undertaking a theory-based evaluation. — — — The Centre of Excellence for Evaluation would like to thank those departments and agencies that reviewed and provided comments on draft versions of this document. Theory-based evaluations: Reflections ten years on. New Directions for Evaluation, 114, 63–67. Weiss, C. H. (1997). Theory-based evaluation: Past, present, and future. New Directions for Evaluation, 76, 41–55. Weiss, C. H. (2000). Cognitive evaluation theory (CET) is a theory in psychology that is designed to explain the effects of external consequences on internal motivation. Specifically, CET is a sub-theory of self-determination theory that focuses on competence and autonomy while examining how intrinsic motivation is affected by external forces in a process known as motivational "crowding out.". CET uses three propositions to explain how consequences affect internal motivation

Cognitive evaluation theory (CET) is a theory in psychology that is designed to explain the effects of external consequences on internal motivation. Specifically, CET is a sub-theory of self-determination theory that focuses on competence and autonomy while examining how intrinsic motivation is affected by external forces in a process known as motivational "crowding out.". CET uses three propositions to explain how consequences affect internal motivation: Based on the profile of people who have attended, and the nature of the discussions that they were interested in pursuing, it has been increasingly clear that what was often a fringe or frontier topic for discussion in many countries (the use of new technologies in education) has become increasingly mainstream, and in many instances, more and more central to educational planning. 8. Little evidence --> more rigorous impact evaluations related to learning Where Those were some personal, and perhaps idiosyncratic, qualitative reflections on how the nature of discussions at the annual Global Symposium on ICT Use in Education have changed over the past ten years. Theory‐based evaluation: Reflections ten years on: Theory‐based evaluation: Past, present, and futur June 2007 · New Directions for Evaluation. Patricia Rogers. Carol H. Weiss. This chapter highlights the continued salience of Weiss's questions about theory-based evaluation, especially given the often simplistic uses of program theory in evaluation. Theory-based evaluation examines conditions of program implementation and mechanisms that mediate between processes and outcomes as a means to understand when and how programs work. Read more. Article.