4

If there **is** little consensus about the assumptions under-
lying a program, theory-based *evaluators* can collect data
relevant to more than one *theory, selecting for* study the
specific links in those theories that answer key questions.

Which Links in Which Theories
Shall We Evaluate?

Carol *Hirschon* Weiss

Theory-based evaluation (TBE) offers many advantages to the evaluator
who conducts the study and the program individuals who receive the
results. It helps to specify not only the *what* of program outcomes but also
the *how* and the why. Theory-based evaluation tests the links between
what programs assume their activities are accomplishing and what actu-
ally happens at each small step along the way. It also has clear limitations
(Weiss, 1997).

   Other chapters in this issue explore the opportunities and challenges
that enter into the decision to use this approach to evaluating programs. I
want to enter the scene after all the actors have decided to take a theory-
oriented approach and now have to put the approach into practice. What
theory do they use? Do they settle on one theory, or do they consider several
theories? In how much detail do they spin out the theories? If they have an
elaborated theory (or theories), which links in the theory do they study?
What criteria do they use in deciding which links are worth studying?

Studying the Mechanisms of Social Change

TBE is an effort to examine the mechanisms by which programs influence
successive stages of participants' behavior. Table 4.1 shows a possible **the-
ory** of a job-training program. **A** theory-based evaluation can examine
whether trainees learn the skills taught, whether learning the skills leads
to the search for a job, whether the search for a job leads to interviews
with prospective employers, whether interviews lead to getting hired, and
so **on.**

Table 4.1. Theory of a Job-Training Program

Program publicizes a job-training program.
> Youth hear about the program.
> Youth are interested and motivated to apply.

Program enrolls eligible youth.
> Youth sign up.

Program provides occupational training in an accessible location.
> Youth attend regularly.

Training matches labor market needs.

Training is carried out well.
> Youth learn skills.

Training teaches good work habits.
> Youth internalize values of regular employment and appropriate behavior on the job.

Program refers youth to suitable jobs.
> Youth apply for jobs.
> Youth behave well in job interviews.
> Employers offer jobs.
> Youth accept jobs.
> Youth show up for work regularly.

Program assists youth in making transition to work and helps with problems.
> Youth accept authority on the job.
> Youth do their work well.
> Youth behave well with coworkers.
> Youth stay on the job.

Source: Adapted from Weiss, 1998, p. 59.

TBE is an attempt to see how far the program succeeds in accomplishing all the intervening phases between enrollment in the program and long-term job holding. If trainees do well all along the route from participation in the training program to staying on a job, there is at least plausible reason to believe that the program was responsible for the trainees' work success. (See Chapter Two by Jane Davidson for further discussion of establishing causality.)

But let us take a step back. Table 4.1 shows the expected steps in the implementation of the program. It is what might be called the *implementation theory* of the program. But why are the trainees going to follow through and wind up in long-term jobs? The table does not delve into underlying psychosocial mechanisms. What is going to keep the trainees engaged in what must seem at first an uncongenial period of training? Perhaps the youth are rational enough to want to acquire skills that will help them get ahead in the job market, or perhaps being with a group of peers provides the social support that keeps them engaged, or perhaps program staff instill a sense of group esprit and a sense of excitement about the benefits of work that support the youth in staying with the program.

These kinds of mechanisms are the things that will largely determine whether the implementation theory succeeds in moving through the steps

described in Table 4.1 from the top to the bottom. They are what I would call the real *program theory*. Together, the implementation theory and the program theory can be called the *theory of change* that the program posits as its route to success.

Which Theory to Select

Some programs are designed on an explicit theoretical basis, and a TBE can investigate whether the assumptions of the theory hold in practice. But many programs are the product of experience, intuition, and professional rules of thumb. A theory-based evaluator has to dig to uncover the implicit assumptions underlying the program. Often there are multiple views on what will make the program successful. Take a program that offers counseling to teenagers at risk of dropping out of school, for example. The counselors are young black and Latino men and women who grew up in the same inner-city neighborhood as the teenagers and studied counseling in a nearby community college. They are expected to steer the teens to a better awareness of the advantages of education and to encourage them to stay in school. Some people involved with the program also expect them to help the youth deal with difficult life circumstances, such as an abusive parent or involvement in gang activities. Some program people also expect them to intercede for troubled youth with social workers, police, or probation officers or to help the teens secure services from health clinics or other service agencies.

Several theories of action might be operating. Some people, maybe the program administrators, think that the counselors are role models for the teens. Because of common ethnic backgrounds and life circumstances, the teens can identify with them, will take their words of advice seriously, and will follow a more positive social path. Another theory might be that the counselors understand the perils and pressures that the teens face and will give advice that is better suited to the real world of the inner city than would a middle-class teacher or counselor. They will know how to advise on family problems because of the commonality of their family backgrounds. Another theory might be that the counselors, understanding the local culture, can use threats and penalties effectively, something that white middle-class counselors would be loath to do. Yet another theory is that the counselors will be well acquainted with all the available services in the community and therefore can refer the youth to an appropriate source of help. All of these assumptions grow from the match of counselors to the ethnic and socioeconomic status of the teenagers.

A different set of assumptions would refer to the specific steps and actions that the counselors use in their relations with the teens, perhaps growing from the particular training that they received in the community college. They may have received training in the use of rewards for small steps that a youth takes in a positive direction, such as offering a movie pass for attending school five days in a row. Or they may have been trained to help with the development of peer support groups, where a group of

youngsters help one another maintain good school attendance and proper completion of schoolwork. One might also imagine that a counselor could be effective by tutoring young people in the subjects that give them the most trouble in school and help them overcome cognitive deficits. There are a plethora of theoretical bases on which one might expect the program to be successful in encouraging young people to remain in school and do good work.

If the evaluator is embarking on a theory-based evaluation, which theory does she[1] hook the study to? Does she follow the counselor's encouragement of school attendance? His intervention into family disputes? His referrals to service agencies? His establishment of support groups? His coaching in math? Or what? One study can rarely collect data on all possible activities and their cascading consequences. It would be burdensome to follow each chain of possible events, and the evaluation would become complex and ponderous. Choices have to be made. The evaluator has to decide which of the several theories to track through the series of subsequent steps.

Overall, there are two major sources of theory—the social science literature and the beliefs of program stakeholders. The advantage of social science theories is that they are likely to be based on a body of evidence that has been systematically collected and analyzed. The main disadvantage is that available social science theory may not match the program under review, and even when it does, it may be at such a high degree of abstraction that it is difficult to operationalize in the immediate context. Nevertheless, when social science provides theory and concepts that ground and support local formulations, it can be of great evaluative value (Chen and Rossi, 1987). The evaluator should bring her knowledge of the social science literature to bear on the evaluation at hand.

A way to begin the task of choosing a theory to follow is to ask the program designers, administrators, and practitioners how they believe the program will work. They may have clear-cut ideas about the chain of actions and reactions that they believe will lead to better school achievement of the youth. But it is not unusual to find that different people in the program hold different assumptions about the steps by which inputs will translate into desired outcomes. What can the evaluator do?

First, she can convene a meeting of the stakeholders in the program, perhaps including the youth who are the program's clients, and ask them to discuss their assumptions about how the program will reach the desired results. They should discuss the ministeps of counselor action and youth response that will lead to success. Through such discussion, their originally hazy ideas may become clear, and they may reach consensus about what the program truly aims to do and how it aims to do it.

Program staff will often find a discussion of this type revealing and eminently practical. They will learn what their colleagues assume should be done (and what they are doing). Staff may all be performing the same functions but doing them with different assumptions about why they will be

successful. Or they may actually be doing different things. In discussion, they can find out whether they are working at cross-purposes or are on the same wavelength. If they are working in different directions, the program is apt to be fragmented and ineffective. Staff will often find the effort to reach consensus a stimulating and useful exercise. It may help the program attain coherence and direction.

Including Several Theories

In some instances, some program staffs cannot reach consensus. They have markedly different theories about where they should put their time and what kind of actions they should take in order to engage problem youth in school. In such cases, it may be necessary to include several different theories in the evaluation design. The evaluation can follow the chains of assumption of several theories to see which of them is best supported by the data.

When a number of different assumptions are jostling for priority, a TBE is wise to include multiple theories. If only one theory is tracked, and that theory is wrong or incomplete, the evaluator may miss important chains of action. The final result may show that positive outcomes were achieved but not through the series of steps posited by the theory. The evaluator will be unable to explain how success was attained (see Brug, Steenhuis, Van Assema, and De Vries 1996; Puska, Nissinen, and Tuomilehto, 1985). Or if the program has disappointing results, and only one theory was tracked, the evaluator may face readers who say, "But that's not how we thought good results would come about anyway." When programs rest on fuzzy assumptions, it is often useful for TBE to represent a range of theoretical expectations.

But the more theories that are tracked, the more complex and expensive the evaluation. It is worthwhile to try to winnow down the number of possible theories to a manageable number. Three or four would seem to be the maximum that an evaluator could explore in a single study. How can the evaluator decide which of the several theories is worth including in the evaluation?

Criteria for Selecting Theories

The first criterion is the beliefs of the people associated with the program, primarily the designers and developers who planned the program, the administrators who manage it, and the practitioners who carry it out on a daily basis. Also important may be the beliefs of the sponsors whose money funds the program and the clients who receive the services of the program. What do these groups assume are the pathways to good outcomes? What are the ministeps that have to be taken if the clients are to reap the benefits that the program promises? What the people who are deeply involved in the program believe is critical because their behavior largely determines how

the program runs. When they hold divergent assumptions about the route to success, the several theories that they proffer become candidates for inclusion.

A second criterion is plausibility. Can the program actually do the things that a theory assumes, and will the clients be likely to respond in the expected fashion? The evaluator needs to see what is really going on. One way is to follow the money. Where is the budget being spent? Where is the program really putting its chips? Which resources are they providing for what kinds of assistance? If the program makes available to each counselor a list of accessible service agencies, their eligibility criteria, and hours of operation, then it is a reasonable bet that they think the referral route is important. If nobody gives the counselors any information about available resources, then this theory is probably not an active candidate for study. If program designers and administrators talk a good deal about ethnic match between counselor and client but end up hiring primarily white middle-class counselors, ethnic match is not an operative theory in this program. Similarly, if the counselors do not know enough about plane geometry or nineteenth-century American history to tutor youth, then assumptions about success through tutoring are not apt to be the route to follow (unless the counselors find other people to do the tutoring). The evaluator needs to take a hard look at the program in action, not just in its planning documents, in order to see which theories are at least plausible in this location.

A third criterion is lack of knowledge in the program field. For example, many programs seem to assume that providing information to program participants will lead to a change in their knowledge, and increased knowledge will lead to a positive change in behavior. This theory is the basis for a wide range of programs, including those that aim to reduce the use of drugs, prevent unwanted pregnancy, improve patients' adherence to medical regimens, and so forth. Program people assume that if you tell participants about the evil effects of illegal drugs, the difficult long-term consequences of unwed pregnancies, and the benefits of complying with physician orders, they will become more conscious of consequences, think more carefully before embarking on dangerous courses of action, and eventually behave in more socially acceptable ways.

The theory seems commonsensical. But social scientists—and many program people—know that it is too simplistic. Much research and evaluation has cast doubt on its universal applicability. Although some programs that convey knowledge in an effort to change behavior have had good results, many have been notoriously unsuccessful. In an effort to add to the stock of knowledge in the program arena, an evaluator may find it worthwhile to pursue this theory in the context of the particular program with which she is working. She may want to carefully track the conditions of the program in order to gather more information about when and where such a theory is supported or disconfirmed by the evidence (and what elements of context, internal organization, and reinforcement make a difference).

So much effort is expended in providing information in an attempt to change behavior (through public service campaigns, material posted to Web sites, distribution of printed materials, lectures and speeches, courses and discussion groups, promotional messages disseminated through multiple media) that careful investigation of this theory is warranted. Furthermore so much uncertainty exists about the efficacy of providing information of different kinds to different audiences that program developers need a better sense of the prospects. The evaluator who pursues this theory in a TBE may look to social science theory for a sophisticated understanding of when and where information is likely to have effects and under what circumstances. She can build this knowledge into the evaluation. When the results of the evaluation are ready, she can offer program developers and staff a greater understanding of the extent to which information creates change within the immediate program context. Many studies have shown that information can lead to change in knowledge and attitudes but not often to change in behavior. The current evaluation can examine whether and where the sequence of steps in the theory breaks down and what forces undermine—or reinforce—the power of information.

A final criterion for choosing which theories to examine in a theory-based evaluation is the centrality of the theory to the program. Some theories are so essential to the operation of a program that no matter what else happens, the program's success hinges on the viability of this particular theory. Let us take the example of a comprehensive community program. The program involves the provision of funds (by government or a foundation) to a group of community residents, who then decide which enhancements the neighborhood needs in order to improve the lot of its inhabitants. The residents can choose to use the funds to add more services (mental health, education, and so on), clean up the streets and parks, rehabilitate buildings, hire private police, attract new business to the neighborhood in order to create jobs for local people, begin a car service for elderly residents, or whatever other services they decide are most likely to improve the local quality of life.

An evaluation can study the services chosen and find out the consequences of adding police or rehabilitating buildings or whatever other new services have been added. But a fundamental premise of this community-based approach is that local residents are knowledgeable, committed, hard working, and altruistic enough to find out what is most needed and to go about getting those services into the community. Further, they are assumed to represent the needs and wants of a wide swath of the community. So an underlying theory has to do with the role of citizen groups in developing and directing a comprehensive community initiative. The effectiveness of a group of residents in representing the interests of their neighborhood and securing priority services is key to the success of the program. This assumption becomes a prime candidate for the evaluation.

Which Links in a Theory to Study

Many theories, if drawn out in detail, consist of a long series of interlinked assumptions about how a program will achieve its effects. Let us go back to the job-training program in Table 4.1. If the evaluation does not have the resources or the time to study all of the steps laid out in the theory, which of them should the evaluation explore? Much of the answer to this question will depend on the practicalities of the situation. At what point is the evaluator brought to the scene? Is it after the first several steps have already been taken? How much money does the evaluation have to collect data? How difficult is it to get some kinds of data? For example, what kind of data will the evaluator need in order to know whether the training is carried out well? How will she find out whether the trainees adopt and internalize the values of regular employment? If some kinds of data are difficult or expensive to collect, that will set practical limits.

Second, program staff may have particular concerns about some segments of the implementation theory. They may want to know, for example, whether trainers are giving proper emphasis to good work habits and other "soft skills" or whether the youth in fact learn the occupational skills that the trainers seek to convey. They may want to know whether staff refer them to relevant jobs and whether the youth comport themselves appropriately in job interviews, so that it is clear why they do or do not get jobs.

It may be even more important to examine some links in the program theory about the psychosocial processes that underlie the program. Here is where much of the uncertainty in social programming lies. What impels developing countries to seek to attract more girls into the school system? What gets faculty members in urban universities to teach in interdisciplinary courses in order to retain students in school? In our example, what are the reasons that trainees persist in the training course and learn both job skills and work readiness skills? Is it the capacity of the trainers to develop supportive communities among the youth? Is it the strength of external rewards and punishments?

An evaluation can concentrate on understanding these kinds of mechanisms and the extent to which they operate within the program milieu. The evaluator can collect data on whether peer groups develop during the course of training and the messages and supports that these groups provide to their members. Do youth affiliate in subgroups? Do members of the various groups support the aims of the training program? (Or do they denigrate the effort to learn skills that will yield "chump change"?) Do the trainers actively encourage the formation of subgroups and provide leadership? What messages circulate in the different subgroups about the value of work and the willingness to accept authority on the job? Regarding the theory about external threats, how important to participants in the training program is the reduction in safety net supports?

Because evaluations to date have told their readers relatively little about the *why* of program success and failure, such inquiries may have great resonance. Studies that explore the psychosocial processes of program theory

will have much to tell program designers, lessons (however tentative) that may be suggestive for a whole range of programs.

Criteria for Selecting the Links to Study

The criteria for choosing which links to study are similar to the criteria for choosing which theories to study. Two are probably most important. The first criterion is the link or links that are most critical to the success of the program. It seems wise to invest resources in studying the particular assumption on which the program most basically rests. If the program is predicated on the assumption that what keeps youth enrolled in the full training program is the support of their peers, then that assumption warrants investigation.

The second criterion is the degree of uncertainty about the linkage. If nobody knows whether the assumption is likely to be supported empirically, or if prior studies have produced conflicting findings on the subject, that link may be worthy of systematic study. Some linkages are unsettled in the social science and the evaluation literatures. Some linkages seem to be supported in the social science literature (or in common sense), but evaluations of earlier programs show that they do not work in practice. An example would be the premise of case management within a multiservice program. A large number of multiservice programs have employed case managers who analyze the services that a family needs, locate and coordinate a range of services, and help the family members obtain appropriate services from relevant agencies. The idea of a family coordinator, an advocate and consultant to the family, sounds so utterly sensible that it is unsettling to find that evaluations have usually not found such programs successful (for example, Bickman and others, 1995; St. Pierre, Layzer, and Goodson, 1997). What are the assumptions that underlie case management? What is the case manager assumed to do, with what immediate consequences, leading to what next steps, with what later consequences? Including some of these kinds of links in the evaluation would yield important information.

Conclusion

In selecting the theory or theories to use as scaffolding for a TBE, the evaluator should consider these criteria.

- The assumptions of the people associated with the program. What are their constructions of the interlinked steps by which program inputs are transmuted into program outcomes?
- The plausibility of the assumptions, given the manner in which the program is allocating its time and resources.
- Uncertainty about the applicability of current assumptions. Given the often inchoate or contested nature of available evidence, do these assumptions hold? Under what conditions do they hold?

- The centrality of the assumptions to the program. If the program is based directly on a particular theory, it would be sensible to make this theory the centerpiece of the TBE.

Once the evaluator decides which theory or theories to use for structuring the evaluation, she ought to spell out all the links in the theory chain—what the program will do, how participants will respond, what the program does next, and so on. Many evaluations will not realistically be able to follow all the links in each chain, and the evaluator needs to choose the links on which to focus. Considerations for making that choice include the practicalities of access, resources, and methodological capability for studying given links and the particular knowledge needs of program staff, who want to know which elements of the program they need to modify or shore up.

In making both choices—which theories to select and which links to study—the evaluator needs to consider the underlying *mechanisms* on which the program rests, what I have called the *program theory* in contradistinction to the *implementation theory*. In some cases, the strongest contribution that TBE can make will be to analyze the psychosocial and political assumptions that undergird the program. TBE can then answer the question *why* as well as *how*.

I doubt that TBE should be a routine part of every evaluation. In many programs and for many purposes, an investigation of theoretical assumptions is too elaborate, too demanding, and probably irrelevant. What many program sponsors and managers want to know can be discovered by simpler and less probing strategies. But I also believe that TBE need not be the exhaustive and exhausting exercise that its image sometimes evokes. It can be domesticated and housebroken to fit even quite routine needs, as long as a key interest is *how* and *why* observed results come out the way they do.

It would be nice to think that over time repeated evaluations of a particular kind of program will yield consistent evidence about the validity of the theories on which the program is based, whether pro or con. In my most optimistic moments, I succumb to the notion that evaluations may be able to pin down which links in which theories are generally supported by evidence and that program designers can make use of such understanding in modifying current programs and planning new ones. I would like to believe that replicated evaluations can explain why some apparently commonsensical program strategies fail to work time after time and thus give clues for program improvement. Such hopes are no doubt too sunny. Given the astronomical variety of implementations of even one basic program model, the variety of staffs, clients, organizational contexts, social and political environments, and funding levels, any hope for deriving generalizable findings is romantic. Nevertheless, theory-based evaluation can add to knowledge. Even relatively small increments of knowledge about how and why programs work or fail to work cannot help but improve program effectiveness. And that is what program evaluation is all about.

Note

1. In this chapter, I use the feminine pronoun to refer to the evaluator and the masculine pronoun for all other actors to avoid the awkward "he or she" construction.

References

Bickman, L., Noser, K., and Summerfelt, W. T. "Long-Term Effects of a System of Care on Children and Adolescents." *Journal of Behavioral Health Services and Research,* 1999, 26(2), 185–202.

Bickman, L., and others. Evaluating Managed Mental Health Services: The Fort Bragg Experiment. New York: Plenum, 1995.

Brug, J., Steenhuis, I., Van Assema, P., **and** De Vries, H. "The Impact of a Computer-Tailored Nutrition Intervention." Preventive Medicine, 1996, 25, 236–242.

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

Puska, P., Nissinen, A., and Tuomilehto, J. "The Community-Based Strategy to Prevent Coronary Heart Disease: Conclusions for the Ten Years of the North Karelia Projects." Annual Review *of Public* Health, 1985, 6, 147–193.

St. Pierre, R. G., Layzer, J. I., Goodson, B. D., and Bernstein, L. S. National Impact *Evaluation of* the Comprehensive Child Development Program: Final Report. Cambridge, Mass.: Abt Associates, 1997.

Weiss, C. H. "How Can Theory-Based Evaluation Make Greater Headway?" Evaluation *Review,* 1997, 21, 501–524.

Weiss, C. H. Evaluation: Methods for Studying Programs and Policies. (2d ed.) Englewood Cliffs, N.J.: Prentice Hall, 1998.

CAROL HIRSCHON WEISS is *professor of* education *at* the *Harvard Graduate School of* Education.