Nothing as Practical as Good Theory: Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families

Carol Hirschon Weiss

The topic on the table is the evaluation of comprehensive cross-sector community-based interventions designed to improve the lot of children, youth, and families.' These types of initiatives draw on a history of experience, from the Ford Foundation's Gray Areas Program in the early 1960s, continuing through the federal programs of the President's Committee on Juvenile Delinquency, the large Community Action Program of the War on Poverty, the Model Cities Program, community development corporations, services integration programs, and others. Most of the government programs incorporated requirements for systematic evaluation; for foundation-supported programs, evaluation was more sporadic and informal. None of the programs was satisfied that it had achieved either maximal *program* benefit from its efforts or maximal *evaluation* knowledge about program consequences from the evaluations it undertook.

In recent years a new generation of comprehensive community initiatives (CCIs) has been funded. Supported in large part by private foundations, the initiatives aim to reform human service and collateral systems in geographically bounded communities. They work across functional areas-such as social services, health care, the schools, and economic and physical redevelopment-in an effort to launch a comprehensive

attack on the social and economic constraints that lock poor children and families in poverty. They bring local residents into positions of authority in the local program, along with leaders of the larger community, public officials, and service providers. Examples of foundation-sponsored initiatives include Annie E. Casey Foundation's New Futures Initiative, Pew Charitable Trusts' Children's Initiative, and the Ford Foundation's Neighborhood and Family Initiative. Recent federal programs, such as the Empowerment Zone and Enterprise Community Initiative, include some parallel features.

A number of evaluations have been undertaken to discover the effects of the recent initiatives. Much effort has gone into developing appropriate outcome measures that can indicate the degree of success-or at least progress-in attaining desirable results. The evaluation strategies being used and proposed have tended to follow standard evaluation practice, emphasizing quantitative measurement on available indicators of outcome, sometimes supplemented by case studies. Influential members of the foundation community have wondered whether these evaluation strategies fit the complexity of the new community initiatives and the knowledge needs of their practitioners and sponsors.*

It is in this context that I suggest an alternative mode of evaluation, theory-based evaluation. In lieu of standard evaluation methods, I advance the idea of basing evaluation on the "theories of change" that underlie the initiatives. I begin by describing this evaluative approach and discussing its advantages. I then make a preliminary attempt to elucidate the theories, or assumptions, on which current initiatives are based. Although this is a speculative enterprise, its aim is to suggest the kinds of questions that evaluation might address in the current case. The paper concludes with some issues concerning the feasibility of theory-based evaluation and a discussion of steps that might test its utility for the evaluation of CCIs. The paper is meant as a contribution to the discussion of how evaluation can derive the most important and useful lessons from current experience.

THEORY-BASED EVALUATION

The concept of grounding evaluation in theories of change takes for granted that social programs are based on explicit or implicit theories about how and why the program will work (Weiss 1972, 50–53; Shadish 1987;

Chen 1990; Lipsey 1993). The evaluation should surface those theories and lay them out in as fine detail as possible, identifying all the assumptions and sub-assumptions built into the program. The evaluators then construct methods for data collection and analysis to track the unfolding of the assumptions. The aim is to examine the extent to which program theories hold. The evaluation should show which of the assumptions underlying the program break down, where they break down, and which of the several theories underlying the program are best supported by the evidence.

Let me give a simple example. There is a job-training program for disadvantaged youth. Its goal is to get the disadvantaged youth into the work force (thus forestalling crime, welfare dependency, drug use, and so forth). The program's activities are to teach "job-readiness skills"-such as dressing appropriately, arriving on the job promptly, getting along with supervisors and co-workers, and so on-and to teach job skills. What are the assumptions-what is the theory-underlying the program?

The theory obviously assumes that youths do not get jobs primarily because they lack the proper attitudes and habits for the world of work and they lack skills in a craft. The program's sponsors may or may not have considered alternative theories-for instance, that high youth unemployment rates are caused by forces in the larger economy and by the scarcity of entry-level jobs with reasonable long-term prospects; or that youth unemployment is a consequence of youths' lack of motivation, their families' failure to inculcate values of work and orderliness, health problems, lack of child care, lack of transportation, a lack of faith in the reality of future job prospects, or ready access to illegal activities that produce higher financial rewards for less work.

Those responsible for the program may have rejected (implicitly or explicitly) those alternative theories, or they may believe that alternative theories are not powerful enough to overwhelm their own theory, or they may believe that other interventions are concurrently addressing the factors that their work neglects.

At the program level, the program theory is based on a series of "micro-steps" that make important assumptions-for example:

- Training for attractive occupations is (or can be) provided in accessible locations.

- Information about its availability will reach the target audience.

- When young people hear of the program's availability, they will sign up for it.

- They will attend regularly.

- Where necessary, stipends (and perhaps child care) will be available to youth while they are in training.

- Trainers will offer quality training and they will help youth learn marketable skills.

- Trainers will attend regularly and provide helpful and supportive counsel.

- Youth will learn the lessons being taught about work habits and work skills.

- Youth will internalize the values and absorb the knowledge.

- Having attained the knowledge and skills, the youth will seek jobs.

- Jobs with adequate pay will be available in the areas in which training was provided.

- Employers will hire the youth to fill the jobs.

- The youth will perform well.

- Employers will be supportive.

- Youth will remain on the job and they will become regular workers with good earnings.

When we examine the theory, we can see how many of the linkages are problematic. At the program level, we know that the quality of instruction may be below par. It can be difficult to recruit young people to job-training programs. Many attendees drop out of the programs; others attend erratically. In some job-training programs, the promised jobs fail to materialize; either the skills taught do not match the job market or employers do not hire the trainees. Many young people get jobs but leave

them in a short time, and so on. There are a host of reasons why the benefits originally expected from job-training programs are usually so small-in the best cases resulting in perhaps a 5 to 10 percent higher employment rate among program participants than among people who do not participate. The San Diego welfare-to-work program, the Saturation Work Initiative Model, was heralded in policy circles as a great success after two years, on the basis of evidence that about 6 percent more of the program participants than of the control group were employed after two years (Hamilton and Friedlander 1989). A five-year follow-up indicated that some of the difference between trainees and controls faded out over time (Friedlander and Hamilton 1993).

In fact, one reason for the current emphasis on community-based cross-systems reform is the need to deal with multiple factors at the same time—education, training, child care, health care, housing, job creation, community building, and so on-to increase the chances of achieving desired effects. The initiatives aim to work on the whole array of needs and constraints, including those that create opportunities, connect young people to opportunities, and prepare them to take advantage of opportunities.

The Case for Theory-Based Evaluation

Why should we undertake evaluation based on analysis of program theory? Basing evaluations on theories of the program appears to serve four major purposes:

1. It concentrates evaluation attention and resources on key aspects of the program.

2. It facilitates aggregation of evaluation results into a broader base of theoretical and program knowledge.

3. It asks program practitioners to make their assumptions explicit and to reach consensus with their colleagues about what they are trying to do and why.

4. Evaluations that address the theoretical assumptions embedded in programs may have more influence on both policy and popular opinion.

Focusing on Key Aspects of the Program. No evaluation, however well funded, can address every question that might be of interest to someone. With the current constraints on evaluation funding, the opportunity to look at a wide range of program processes and outcomes is further limited. In any evaluation of a program as complex as the current initiatives for children, youth, and families, careful choices need to be made about where to put one's evaluation energies. Central hypotheses about the program appear to represent potential issues that evaluation should address.

If good knowledge is already available on a particular point, then we can change its label from "hypothesis" or "assumption" to something closer to "fact," and move along. However, where a central tenet of the program is still in doubt, or in contention, then it might represent a question for which evaluation is well suited.

Generating Knowledge about Key Theories of Change. A whole generation of anti-poverty programs has proceeded on the basis of kindred assumptions, and we still lack sound evidence on the extent to which the theories hold up in practice. Many "effective services" programs, which began from somewhat different premises, have come to believe that "you can't service people out ofpoverty" (Schorr 1994), and have moved toward the same kinds of theories. Some assumptions have persisted since the Ford Foundation's Gray Areas Program. Although a great many evaluations were conducted on the community-based anti-poverty programs (including those in education, health, mental health, housing, community organization, and social services of many kinds), there has not been much analysis of the underlying assumptions on which they were based.

Effort was put into looking at outcomes-for example, school attendance, unemployment rates, and feelings of self-esteem. In later years increased attention was directed at studying how the programs were carried out-for example, styles of service and length of contact. Considerable knowledge accumulated about processes and outcomes. A small number of analysts have sought to synthesize the knowledge, but many of them have subordinated the synthesis to their own interpretations of the causes and cures of chronic poverty (for example, Bane and Ellwood 1994; Jencks 1992; Wilson 1987; Schorr 1988, 1991; Haveman 1977; Haveman and Wolfe 1994).

Creating a useful synthesis of the findings of evaluation studies on community-based programs has been difficult to do. The original evalua-

tion studies used a large assortment of indicators, periods of follow-up, sources of data, methods of study, definitions, and perspectives. Their research quality also varied widely. To add them up presents the familiar apples-and-oranges problem. Meta-evaluation, the quantitative technique that aggregates the results of different studies into an overarching conclusion, is suitable for studies of a single type of program, where the quantitative measures of program effects can be converted into a common metric of effect size. To synthesize the results of the hodgepodge of evaluation studies available on community-based cross-sector interventions at this point would require substantive knowledge and analytic skills of rare discernment.

Nevertheless, important questions about the implicit hypotheses of community-based programs endure. It would be very useful to direct new evaluations toward studying these theoretical hypotheses, so that knowledge accrues more directly on these key matters.

Making Explicit Assumptions, Defining Methods, and Clarifying Goals. A third benefit of theory-based evaluation is that it asks program practitioners to make their assumptions explicit and to reach consensus with their colleagues about what they are trying to do and why. Without such a conversation, it is likely that different participants have different tacit theories and are directing their attention to divergent-even conflicting-means and ends. Imagine, for example, a preschool teacher who believes in unconditional affection and nurturance for the children in her care, working under a supervisor who requires that the children demonstrate cognitive achievement (numbers, colors) before they can receive approval. At the extreme, the assumptions and practices of the teacher and the supervisor may be so divergent that their efforts tend to cancel each other out.

When they are asked to explicate the theories on which the program is based, the discussion among practitioners-and between them and program designers, managers, sponsors, community leaders, and residents-is likely to be difficult at first. Usually they haven't thought through the assumptions on which the program is based but proceed intuitively on the basis ofprofessional training, experience, common sense, observation, and informal feedback from others. Although reaching a consensus will be no mean feat, it is expected that discussion will yield agreement among program stakeholders and that the theories will represent a common understanding.

When the evaluator seeks to elicit formulations of program theory from those engaged in the initiatives, they may begin to see some of the leaps of faith that are embedded in it. Program developers with whom I have worked sometimes find this exercise as valuable a contribution to their thinking as the results of the actual evaluation. They find that it helps them re-think their practices and over time leads to greater focus and concentration of program energies.

Influencing Policy. Evaluations that address the theoretical assumptions embedded in programs may have more influence on opinions, both elite opinion and popular opinion.

Theories represent the stories that people tell about how problems arise and how they can be solved. Laypeople as well as professionals have stories about the origins and remedies of social problems (poor people want to work but the jobs have disappeared; services make people permanently dependent). These stories, whether they arise from stereotypes, myths, journalism, or research knowledge, whether they are true or false, are potent forces in policy discussion. Policies that seem to violate the assumptions of prevailing stories will receive little support. Therefore, to the extent that evaluation can directly demonstrate the hardiness of some stories (theories) and the frailty of others, it will address the underlying influences that powerfully shape policy discourse.

In a sense, all policy is theory. A policy says: If we do A, then B (the desired outcomes) will occur. As evaluative evidence piles up confirming or disconfirming such theories, it can influence the way people think about issues, what they see as problematic, and where they choose to place their bets. The climate of opinion can veer and wiser policies and programs become possible.

◊   ◊   ◊   ◊   ◊

In sum, the theory-driven approach to evaluation avoids many of the pitfalls that threaten evaluation. It helps to ensure that the developments being studied are good reflections of the things that matter in the program and that the results identified in the evaluation are firmly connected to the program's activities (Chen and Rossi 1987). Tracking the micro-stages of effects as they evolve makes it more plausible that the results are *due to* program activities and not to outside events or artifacts of the evaluation,

and that the results generalize to other programs of the same type. These are strong claims, and inasmuch as only a few large-scale theory-based evaluations have been done to date, it is probably premature to make grandiose promises. But certainly tracing developments in mini-steps, from one phase to the next, helps to ensure that the evaluation is focusing on real effects of the real program and that the often-unspoken assumptions hidden within the program are surfaced and tested.

THEORIES OF CHANGE UNDERLYING CCIs: A FIRST TAKE

The comprehensive initiatives with which we are engaged are extraordinarily complex. What services they will undertake, how they will manage them, how they will conduct them, who will be involved-all these facets are to be determined on the ground in each community, with the full participation of the unique constellation of individuals in positions of business, political, and community leadership, and professional service. Unlike the job-training example that I have given, it is almost impossible to develop a plausible set of nested theoretical assumptions about how the programs are expected to work. In one community the assumptions might have to do with a series of steps to coordinate existing services available from the public and private spheres in order to rationalize current assistance, and then fill in the gaps with new services. Another community might have theories related to the empowerment that accrues to local residents who gain a strong voice in the organization and implementation of social programs for the community, and the consequent psychological and political mobilization of residents' energies. One initiative may focus on enhancing the quality of life of *individuals* with the expectation that individuals in more satisfactory circumstances will create a better community. Another initiative may put its emphasis on building the community and its social networks and institutions, in the hope that a better community will make life more satisfying for its residents.

It is challenging, if not impossible, to spell out theories of change that apply across the board to all the existing foundation-sponsored initiatives and to such federal programs as Empowerment Zones and Enterprise Communities. They differ among themselves in emphasis, managerial structure, and priorities. They allow for complex interactions among participating entities; they give great autonomy to local community efforts;

they foresee a process of long-term change; they do not even try to foresee the ultimate configuration of action. But if we cannot spell out fine-grained theories of change that would apply generally, we can attempt to identify certain implicit basic assumptions and hypotheses that underlie the larger endeavor. That is what the rest of this paper is about.

An Examination of Assumptions

I read a collection of program documents about community-based comprehensive cross-sector initiatives for children, youth, and families (Chaskin 1992; Enterprise Foundation 1993; Pew Charitable Trusts, n.d.; Rostow 1993; Stephens et al. 1994; Walker and Vilella-Velez 1992), and here I outline the theoretical assumptions that I discerned. These assumptions relate to the service-provision aspects that appear to underlie confidence that the initiatives will improve the lot of poor people. (I limit attention to service provision here, even though additional assumptions, including those about structure and institutional relationships, are also important.) Some of the assumptions on which the initiatives appear to be based are well supported in experience; others run counter to the findings of much previous research. For most of them, evidence is inconclusive to date.

Assumption 1: *You can* make an impact with limited funds. A relatively modest amount of money (on the order of half a million dollars per community per year) will make a significant difference. Even though the War on Poverty in its heyday was spending massively more money than that, the assumption here appears to be that this money can stimulate activity on a broad array of fronts that will coalesce into significant change. Perhaps it is assumed that much has been learned from prior experience that can now be exploited.

1. *The money will leverage money already available in the community for public and private services*. One possible assumption may be that the carrot of additional money will stimulate greater willingness among public and private agencies to coordinate existing services. Each of the agencies serving the community may be willing to give up some of its autonomy and control in return for some additional money from the initiative, and engage in more collaborative action. Agencies are customarily short of uncommitted funds, and even minor infusions of money can be assumed to divert them to new ends or new means. A further assumption is that the resulting coordination will yield large benefits (see the third point in Assumption 3, below).

2. Another possible way in which the additional money can be expected to leverage *current expenditures is* by *finding the creation and ongoing operation of a high-powered board of community leaders (elected officials, business people, volunta y association leaders),* service professionals, *and local residents.* This board will have the clout to persuade service providers to be more respon-sive to community needs, to coordinate more effectively, and to plug up the gaps in service provision. Because of its stature, and the stature of the national foundation that stands behind it, the board can convene meetings and conferences among important segments of the community and exercise its influence to ensure that coordination is succeeded by true collaboration across sectors. The funds, in this formulation, provide for the staff work for a steering body of influential elites (including "elites" from among the residents), and the theory would posit that it is the influence of the elites that succeeds in attaining coordination and funding new services.

3. *A third way in which modest resources from the initiative might stimulate action would be* by *finding a central entrepreneurial staff.* It might be assumed that this staff would have the savvy to locate needs and opportu-nities in the community. For example, a shooting episode in a local school might spark public concern about serious violence, and the entrepreneurial staff might seize upon this opportunity to press for further services from the schools and the police, and for greater cooperation between them, as well seek additional funding for enhanced services. Or the occasion of a search for a new school superintendent might provide the initiative staff an oppor-tunity to set forth their agenda for what a superintendent should do, and therefore for the kind of person to be hired. If the "new" money supported an activist staff who could locate windows of opportunity and fashion ap-propriate agendas for action, it might be assumed to have multiple payoffs.

4. *Money could also be used to fund research, analysis, and evaluation.* The intent here would be to marshal1 the experience of earlier change efforts, to monitor the programs and projects supported by the initiative, to analyze opportunities, costs, and benefits, and to evaluate the consequences of action. The assumption would be that people respond rationally to the presentation of formal, systematic evidence, and that they use it to improve the work they are doing. It implies that research evidence helps to overcome preferences based on other grounds. For example, the assumption is that service staff will heed analysis showing that a particular program has been unsuccessful with a particular kind of family, and change their approach to service, despite such factors as their familiarity with traditional ways of

work, the structure of the service organization that supports accustomed practice, expectations from collateral agencies, professional convictions and allegiances, political pressures, and so on.[3]

Assumption 2: An effective program requires the involvement of local citizens. This assumption can rest on any of several grounds, or on a combination of them.

1. *Local residents bring Local knowledge, representativeness, and Legitimacy.* Local residents on a board may be expected to have a better understanding of local needs, and therefore be able to direct the program toward things that matter to the people on the site. Local residents on a board may be expected to have greater legitimacy to local residents, who will then be more trusting of actions that emerge from the local initiative and be more likely to give those actions their support.

Local residents on the board may also be expected to be "representative" of the community. Even if they are not elected, they may be seen as democratically empowered to speak in the name of the whole community. All communities have existing divisions (by ethnicity, age, gender, recency ofmigration, economic status, education, aspiration, law-abidingness, and so on), and poor communities have at least their share. Still, in some way the local residents invited to serve on the board may be viewed by the business and professional members ofthe board as manifesting the "general will" of the poor community.

Another scenario is that local residents on the board may be expected to be effective spokespersons to outside funders and other influentials. An articulate person who has spent three years on welfare and worked her way off can be expected to speak with conviction and be heard with respect, and thus may be effective in public relations and fund raising.

2. *It is expected that resident members of the board will be eager and effective participants.* They will want to participate on the board on a voluntary basis. They will attend meetings regularly. They will have the skills to deal with the matters that come before the board. They will have the time to give to participation. They will be conscientious in learning about matters up for discussion. If they are expected to represent the community, they will make at least some kind of bona fide attempt to canvass opinion in the community. They will be able and willing to articulate their preferences in a group that includes better-educated and higher-socioeconomic-status members.

3. Local residents on the board will not bring serious limitations to the task. They will not try to work the initiative for their own personal benefit (beyond an acceptable range). For example, they will not give their relatives priority in hiring regardless of qualifications or appropriate initiative property for their personal use.

4. A further hypothesis might be that the more participants the better.4 Extensive representation of residents is valuable because it brings to the table a wider range of ideas and experiences, and increases the diversity of opinions considered in planning and operations. Even though increased diversity is likely to generate conflict and slow the pace of action, nevertheless it enriches plans and ideas.

Assumption 3: Urban neighborhoods are appropriate units on which to focus program attention. Another assumption is that an urban neighborhood is a unit that makes sense for improving services and opportunities. Even though it is not a political subdivision, an urban neighborhood has natural boundaries that residents pretty much agree upon and that distinguishes it from nearby areas. It has social coherence so that residents feel at least some sense of common destiny. There is a real "community" and people who can speak for the community.

1. *Physical space.* Although assumptions on this topic are only hinted at in the documents I read, there may be theories about the improvement of physical space in the neighborhood. For example, improvement in outdoor physical space, such as improved street lighting, might be expected to lead to a reduction in crime and a reduction in fear of crime. As another example, improvement in the esthetics of the street, such as fill-in structures for snaggle-tooth blocks, will improve community morale. Or, expansion and improvement of recreational areas will provide play space and outlets for the energies of youth, with the expectation that this will reduce their engagement in illicit activity. Or turning rubbish-strewn empty lots into gardens will provide constructive activities for young people, give them a sense of pride in the neighborhood and even perhaps some potentially marketable skills, and give pleasure to residents.

Improving the housing stock can be expected to have a host of positive effects, so long as residents can afford the units that become available. Upgrading existing housing units and building additional units might be expected to improve the health of family members: such improvements will provide space and privacy so that tensions are reduced and family

relationships improve; children will have space to do homework and therefore will be more conscientious about it and thus do better in school; better cooking facilities will be available, which can be expected to improve nutrition; and so forth. If very-low-priced units (or rooms) are created, the numbers of homeless people on the streets can be expected to be reduced, with improvement in their lives and enhancement of the esthetics of the neighborhood.

2. *Economic development.* A series of assumptions are embedded in expectations for economic development of the neighborhood. Investment and loans for businesses and housing might be assumed to result in increased income for residents (if it is assumed that they are the ones employed in the businesses) and better housing conditions (assuming they get priority in the new or rehabbed housing) and increased income might be expected to lead to new enterprises (since residents are now more affluent consumers), which in turn are expected to create jobs and lead to prospering local retail and perhaps small craft and manufacturing businesses. Local businesses will employ local workers, and thereby give hope to potential trainees in job-training programs and students in educational programs.

3. *Socialdevelopment.* With the neighborhood as the unit for planning, services, economic development, and physical rehabilitation, further development of the positive aspects of the neighborhood can be expected in the form of local clubs and associations, religious congregations, schools, and informal interactions. Why should this happen? Perhaps because of symbiosis. An upward spiral of development might be expected because many of the separate activities will be successful and thus contribute to rising hope, satisfaction, optimism about the future, and a sense of common destiny. The bedrock hypothesis is that the visible success of early efforts will set off a chain of optimism and rising expectations.

Perhaps another theoretical strand would be that social and physical development can lead to a safer environment. Fewer people would commit crime; police would be more zealous about catching criminals; and crime rates would go down. People would feel safe to walk on the streets; instead of hiding behind their double-locked doors, they would engage in the kinds of social activities that bring liveliness and culture to the neighborhood.

4. *Social services.* A serious theoretical premise is that services can be effectively coordinated on a neighborhood level. Even though each separate service reports to a "downtown" bureaucracy, neighborhood caregivers from health, welfare, employment, policing, probation, sanitation,

health inspection, and education will be motivated to coordinate their services. They will not be constrained by the standard operating procedures of their agency, its longstanding regulations, traditions, and culture. They will embrace coordination, not sabotage the program's operation. In fact, staffshould press for changes in bureaucratic rules that will accommodate residents' wishes for integrated services, family-centered care, and cuts in red tape. They can even be expected to press for co-location of services if and where this is one of the residents' priorities.

Downtown bureaucracies are expected to accede to such pressures for greater decentralization of services and increased coordination at the neighborhood level, even when it reduces the authority of the central bureaucracy. This unusual organizational behavior may have its origins in the fact that a high-ranking representative from each social service department serves on the board of the initiative, and these representatives will promote the objectives of the initiative neighborhood within their own organizations. Perhaps there is also pressure from the city's electedofficials to accommodate the initiative (why?) or to respond to residents' demands because of their enhanced political organization and electoral mobilization. (If the latter is part of the theory, we need to adumbrate the set of assumptions about how political organization and electoral mobilization develop from the initiative's activities.)

Assumption 4: Neighborhood action will achieve the initiative's **goals.** A collateral hypothesis is that neighborhood involvement is sufficient to achieve the goals of the initiative, by using the influence of the neighborhood to leverage other resources. Additional action would be desirable at federal, state, and city levels or by corporations, banks, and supra-neighborhood private voluntary associations other than those involved. But while added resources and interventions would be beneficial, an important assumption is that the initiative board and staffoperating at the neighborhood level are sufficient to mobilize resources necessary to make the program successful.

Assumption 5: Comprehensive services will lead to success. Comprehensiveness of services is indispensable. The assumption is that many prior failures in programming were due to single-strand narrow-band programs. Each program addressed one need of a poor child, youth, or parent, but failed to recognize the extent to which families were trapped in a web of

constraints that single programs did not reach. No one program is sufficient to alleviate the multiple problems of a family suffering from low income, debt, poor health, lack of preschool day care, school failure of another child, and overcrowded, dilapidated housing. Only services across the whole range of need will help such a family escape from poverty.

1. *The nested assumption is that comprehensive service is possible to establish and maintain.* Agencies and direct-service workers can take the whole family as the unit of service and provide direct assistance themselves, direct assistance from another worker in the same or a nearby location, or easy, convenient referral to needed service elsewhere. Workers will be able to do at least a quick appraisal of the kinds of service required and know the appropriate care-givers who can provide that service. They will know the rules and regulations, eligibility standards, and operating procedures of hospitals, foster care agencies, probation services, welfare agencies, employment agencies, and the like, and can not only give referrals but can also follow up to see that family members receive appropriate help. They will have had sufficient training to prepare them for this changed role.

2. Perhaps *another* assumption *is that professional care-givers will intervene on behalf of their clients if proper assistance is not forthcoming.* Although such intervention is likely to bring care-givers into conflict with other social service providers (physicians, teachers, social workers, and so forth), they will run the gauntlet for the sake of their clients and press the other agency to alter its practice. Presumably they will usually be successful (or else the clients will lose confidence and hope, and the care-givers themselves will lose heart).

3. *Workers in the community initiative will seek policy* changes in service agencies to which clients are referred, and in other agencies, such as transportation and sanitation, so that they can collaborate in ensuring comprehensiveness of services.

4. *Implicit, too, is the expectation that these other agencies will alter their rules, regulations, and operating procedures to* adapt to the need for comprehensive provision of service to the community. (See item 4, under Assumption 3, above.)

Assumption 6: Social service interventions will succeed irrespective of employment conditions. Interventions in the social service sphere will make headway without regard to the employment structure. Business and industry, which control the availability of most jobs in the nation, are not

apt to be affected by the community initiatives (except perhaps in some distant future if the community has turned around and become a thriving market and source of able workers). Without changes in the availability of jobs, the assumption evidently is that families served by the initiative will move to the head of the job queue. They may thus displace applicants less capable of satisfying the needs of the job market.

Assumption 7: Services for adults confer benefits on **children.**A final set ofassumptions deals with the intra-familial allocation ofbenefits. There is an assumption that when an adult in a family receives services, benefits accrue to younger members of the family. A mother whose asthma is relieved has more energy to devote to her children; a father who receives training and gets a job becomes a positive role model for his children and is better able to support their needs. However, it is possible to imagine feedback loops that are less benign. A mother newly enabled to get a job may leave her children with a neglectful relative; a father who gains kudos through taking a leadership role in the community may lose interest in the relatively pale rewards of family life. Actions that assist adults may not automatically redound to the benefit of their children,

◊   ◊   ◊   ◊   ◊

In seeking to tease out the underlying hypotheses of the programs, I may have omitted a number of strategic points and perhaps included some that are tangential. I hold no brief for this particular list. My aim has been to give an example of what it would mean to begin an evaluation with an explication of the theories implicit in the program. The evaluation can then be directed toward testing those theories. I do not mean "test" in the sense of experimentation or even necessarily of quantitative assessment. I simply mean asking questions that bear on the viability of the hypotheses in these particular cases, through whatever methods of inquiry are chosen.

THE PROVISIONALITY
OF THE UNDERLYING HYPOTHESES

Some of the hypotheses in the list are well supported by evidence and experience. Some are contradicted by previous research and evaluation. For

example, Wilner's (1962) study of the effects of public housing on its residents failed to find any of the positive effects, compared with a matched comparison group, that had been posited. But that study was done a long time ago. Today public housing is different; neighborhoods are different; families are different. While the new high-rise public-housing projects of the 1950s represented great hopes for improvement not only in housing but also in family functioning, they proved to be disastrous in many locations. Public housing has now developed theories of the advantages of small low-rise units on scattered sites with tenant participation in management.

Another example: all the studies that I'm familiar with about coordination/integration of public social services have documented the extraordinary difficulties of changing the behavior of workers and agencies (see, for example, Arizona Department of Economic Security 1989 and State Reorganization Commission 1989). But perhaps there are success stories that give clues about necessary incentives and sanctions.

An important step will be to discuss the theories that practitioners and residents engaged in community-building initiatives actually have in mind as they go about their practice. Often their theories will be implicit rather than explicit, and it may take time for them to think through their assumptions about how their work will lead to the effects they seek. Nevertheless, the feasibility of theory-based evaluation rests on their ability to articulate their assumptions (or to assent to someone else's formulation), and it is important to see how well this phase of the task can be done.

Then it will be useful to assemble the available evidence from prior evaluation and research studies. Perhaps, where the weight of the evidence casts doubt on the efficacy of particular strategies and lines of work, practitioners may feel impelled to find alternative ways to think and to act. Even before the evaluation gets under way, the process of subjecting assumptions to the test of available evidence can be a useful stimulus to re-thinking and re-tooling.

Another advantage of looking at past studies comes when an initiative has many ideas and assumptions that are worth studying and, because of inevitable limitations on resources, has to choose among them. Earlier studies can help narrow the choices. Where the overwhelming weight of existing evidence supports a theory and its associated activities, there may be less urgency to include that issue in the new evaluation. Other issues can receive priority. Similarly, it may be less important to evaluate issues where firm evidence already documents the causal chains that link interventions

to early stages of progress or link early stages of progress to long-term outcomes. For example, in the evaluation of smoking-cessation programs, evaluators concentrate their efforts on studying the programs' effectiveness in getting **people** to give up cigarettes. They do not go on to study the health benefits of stopping smoking. Researchers have long since proved to everyone's satisfaction that giving up smoking yields significant decreases in morbidity and mortality. Analogously, if there is **sufficient** evidence that some indicator of intermediate progress is firmly linked to successful **long-**range outcomes, the evaluation need not proceed to **verify** the connection.

One significant point should be mentioned here. A program may operate with multiple theories. I do not mean that different actors each have their own theories, but that the program foresees several different routes by which the expected benefits of the program can materialize. To take a simple example, a counseling program may work because the counselor gives support and psychological insight that enables a young person to understand her situation and cope with it; it may work because the counselor serves as a role model for the young woman; it may work because the counselor provides practical information about jobs or money management; it may work because the counselor refers the client to other useful sources of help. All of those mechanisms are possible, and some or all of them may work simultaneously.

Similarly, a community initiative may work through a variety of different routes. There is no need to settle on one theory. In fact, until better evidence accumulates, it would probably be counterproductive to limit inquiry to a single set of assumptions. Evaluation should probably seek to follow the unfolding of several different theories about how the program leads to desired ends. It should collect data on the intermediate steps along the several chains of assumptions and abandon one route only when evidence indicates that effects along that chain have petered out.

OUTCOME INDICATORS **FOR ACCOUNTABILITY**

The aim of this paper has been to indicate a style of evaluation that comprehensive community initiatives might pursue. Evaluators could set forth a number of hypotheses that underlie the initiatives. After discussing relevant factors with program participants and reaching agreement on theories that represent the "sense of the meeting," the evaluators would

select a few of the central hypotheses and ask: To what extent are these theories borne out in these cases? What actually happens? When things go wrong, where along the train of logic and chronology do they go wrong? Why do they go wrong? When things go right, what are the conditions associated with going right? Also, the evaluation could track the unfolding of new assumptions in the crucible of practice. The intent is not so much to render judgment on the particular initiative as to understand the viability of the theories on which the initiative is based. The evaluation provides a variegated and detailed accounting of the why's and how's of obtaining the outcomes that are observed.

But sponsors and participants may also want periodic soundings on how the local program is faring and how much it is accomplishing. For purposes of accountability, they may want quantitative reports on progress toward objectives. Theory-based evaluation does not preclude-in fact, is perfectly compatible with—the measurement of interim markers and long-term outcomes, such as high school graduation rates, employment rates, or crime rates. As a matter of fact, if wisely chosen, indicators of interim and long-term effects can be incorporated into theory-based evaluation.

Indicators can cover a gamut of community conditions before, during, and after the interventions. Evaluators can collect information on:

- school attendance rates, drop-out rates, graduation rates, scores on standardized tests;

- infant mortality and low birth-weight rates;

- unmarried childbearing rates;

- overall crime rates, auto theft rates, arrests of minors, and other crime statistics;

- numbers of families receiving Aid to Families with Dependent Children (AFDC);

- numbers of families moving off welfare in a twelve-month period;

- unemployment rates for teenagers and adults;

- numbers of clubs and associations active in the community and average attendance at meetings and events;

- attendance at religious services;

- registration and voting rates;

- numbers of books borrowed from local libraries;

- usage of hospital emergency rooms; and so on.

Such data can give some indication of the state of the community before the initiatives start up, and they can be periodically updated. However, they represent gross measures of the community, not of individuals in the community. To find out about individuals (by age, race/ethnicity, income level, gender, family status, and so on), indicator data can be supplemented by survey data on a random sample of *individuals* in the community. Periodic updates can show whether changes are taking place, in what domains, and of what magnitude, and they allow comparison of those who received direct help versus those who did not, two-parent versus one-parent families, and so forth.

The shortcomings of relying only on indicator data are several-fold:

1. Data on community-based rates reflect the condition of the entire population of the community, not just those who are affected by the initiative's work. Therefore, they are likely to be "sticky"—difficult to move. Lack of change in the indicators does not necessarily mean that nothing good is happening, but if good things are happening, they are affecting too small a fraction of the community's residents to make a dent in population-based indicators.

2. Any changes that show up in the data are not necessarily due to the initiative. (This is true not only in the case of community-based indicators, but of survey data on individuals.) Many things go on in communities other than the intervention. Economic changes, new government programs or cutbacks of programs, influx of new residents, outflowofjobs, changes in the birth rate—all manner of exogenous factors can have enormous consequences in this volatile time. It would be difficult to justify giving the credit (or blame) for changes (or no changes) on outcome indicators to the initiatives.

3. We do not know *when* expected results are apt to appear. Little experience has prepared us to understand how soon change will occur. All we know is that there will be a time lag of unknown duration before the effects of CCIs are manifested. This lack of knowledge makes interpretation of indicators chancy.

4. One of the key features of CCIs is their belief that it is vital not only to help individuals but also to strengthen the community, and that strengthening the community will reciprocally work to trigger, reinforce, and sustain individual progress. CCIs tend to believe in the significance of changes at the community level, both in and of themselves and as a necessary precondition for individual advancement, just as they believe that individual improvement will support a revitalized community. But few data are systematically and routinely collected at the level of the neighborhood, and those data that are available rarely fit the boundaries of the neighborhood as defined by the CCI. It is problematic how well available indicators can characterize community-level conditions.

For a variety of reasons, then, I would propose that even if outcome-oriented data are collected on the community (and a random sample of its residents), the items selected for study be carefully chosen on the basis of program theory. Only those indicators should be studied that can be linked, in a coherent and logical way, to the expected activities of the initiatives and to the intermediary outcomes anticipated from them on the basis of thoughtful and responsible analysis.

POSSIBLE PROBLEMS WITH IMPLEMENTING
THEORY-BASED EVALUATION

Using theories of change as the basis for evaluation promises to help us avoid some of the most debilitating pitfalls of past evaluations of community-wide programs: **(1)** exclusive reliance on individual-level data, which evades questions about the role of "community" or "neighborhood" and casts no light on the effectiveness of directing program efforts at "refocusing the system," and (2) an inability to explain how and why effects (or no effects) come about in response to program interventions. Theory-based evaluation addresses such issues directly.

With all its appeal, however, the theories-of-change approach to evaluation no doubt faces serious problems in implementation. Let me mention four of them: problems of theorizing, measurement, testing, and interpretation.

Problems **of** *Theorizing*

A first problem is the inherent complexity of the effort. To surface underlying theories in as complex and multi-participative an environment as these communities represent will be a difficult task. At the first level, the level of the individual stakeholder, many program people will find the task uncongenial. It requires an analytical stance that is different from the empathetic, responsive, and intuitive stance of many practitioners. They may find it difficult to trace the mini-assumptions that underlie their practice, dislike the attempt to pull apart ideas rather than deal with them in *gestalts,* and question the utility of the approach.

The next level arrives when agreement is sought *among* participants about the theory of the whole CCI. There is likely to be a serious problem in gaining consensus among the many parties. The assumptions of different participants are likely to diverge. Unless they have had occasion before to discuss their different structures of belief, there will be a confrontation over what the real theory of the CCI is. When the confrontation surfaces widely discrepant views, it may prove to be unsettling, even threatening. I believe that in the end, the attempt to gain consensus about the theoretical assumptions will prove to have a beneficial effect on practice, because if practitioners hold different theories and aim to achieve different first- and second-order effects, they may actually be working at cross-purposes. Consensus on theories of change may in the long run be good not only for the evaluation but for the program as well. But getting to that consensus may well be painful.

There is a third level, which comes when a CCI goes public with its theoretical statement, whether formally or informally. A CCI may run political risks in making its assumptions explicit.5 Canny community actors do not always want to put all their cards on the table. Such revelation may lay them open to criticism from a variety of quarters. Particularly when racial and ethnic sensitivities are volatile, even the best-meaning of assumptions may call forth heated attacks from those who feel slighted or disparaged as well as from those who dispute the analytical reasoning of the theories proposed.

Before we reach conclusions about adopting theory-based evaluation, it will be important to try it out with engaged actors in communities undergoing significant interventions. Their willingness and ability to work through the concept are necessary conditions for effective conduct of this kind of evaluation.

Politics can also inhibit theorizing. Observers of evaluation and other policy-oriented research have suggested that the urge to be "policy-relevant" impels evaluators to take their research questions and their measures of success from the political sphere and to concentrate on issues and options that fit the current political agenda. To the extent that evaluators focus narrowly on issues that are politically acceptable, they fail to articulate and test "alternative sets of assumptions-or alternative causal stories. ... [This omission] effectively creates conditions in which we are likely to 'know' more but 'understand' less" (Brodkin, Hass, and Kaufman 1993, 25). Analysts like Brodkin suggest that evaluation of government policies is so embedded in politics that it is fruitless to hope for the necessary attention to causal theory.

Perhaps the same limitation would hold for evaluation of foundation-supported activities. Organizational politics may call for a blurring of outcomes and alternatives. On the other hand, foundation initiatives operate at some remove from the turbulent politics of Washington, and they may allow greater scope for rational evaluation.

Problems of Measurement

Once consensual theories of change are in place, evaluators have to develop techniques for measuring the extent to which each step has taken place. Have agencies adapted their procedures in ways that enable them to function in a multi-agency system? Have practitioners reinterpreted their roles to be advocates for clients rather than enforcers of agency rules? Some of the mini-steps in the theories of change will be easy to measure, but some-like these-are complicated and pose measurement problems. Whether they will all lend themselves to quantitative measurement is not clear. My hunch is that some will and some will not.

Whether exclusively quantitative measurement is desirable is also not clear. To the extent that theory-based evaluation represents a search "for precise and decomposable causal structures" (Rockman 1994,148) through quantitative measurement and statistical analysis, it may be taking too positivistic a stance. The logic of qualitative analysis may be more compel-

ling, since it allows not only for rich narrative but also for the modification of causal assumptions as things happen in the field. But since sponsors often find quantitative data more credible than narrative accounts, efforts should probably be made to construct measures of key items.

Problems of Testing Theories

Under the best conditions of theory, design, and measurement, will it be possible to *test* (that is, to support or disconfirm) theoretical assumptions? It is possible that statements of theories of change will be too general and loosely constructed to allow for clear-cut testing. Data collected may be susceptible to alternative interpretations. Unless statements about the theoretical assumptions of the CCI expressly articulate what is *not* meant, what is *not* assumed, as well as what is, it may be difficult to formulate decision rules about the conditions under which a phase of theory is supported or rejected.

Problems of Interpretation

Even if we should find theories that tend to explain the success of particular initiatives in particular places, it is uncertain how generalizable they will be. Will interventions in another community follow the same logic and bring about the same outcomes? On one level, this is a question of how sufficient the theories are. It is possible that even when available data seem to support a theory, unmeasured conditions and attributes in each local case actually were in part responsible for the success observed. Unless other CCIs reproduce the same (unmeasured and unknown) conditions, they will be unable to reproduce the success. Only with time will enough knowledge accrue to identify all the operative conditions.

On a deeper level, the question involves the generalizability of any theory in the social sciences. Postmodern critics have voiced disquieting doubts on this score. But this subject gets us into deeper waters rhan we can navigate here.

CONCLUSION

For all its potential problems, theory-based evaluation offers hope for greater knowledge than past evaluations have generally **produced.** I believe that the current comprehensive community initiatives should try out its possibilities. If we are to make progress in aiding children and families, the

nation needs to know and understand the effects of major interventions. These initiatives represent a potent opportunity not only to *do* good but, perhaps more important, to *understand how,* when, and why the good is being done. Only with greater understanding of the processes of change will it be possible to build on successes in demonstration communities, to "go to scale" and bring benefits to children and families all over the country.

NOTES

1. I wish to thank Penny Feldman, Ron Register, Gary Walker, and Jo Birckmayer for their helpful comments on an earlier version of this paper, as well as the participants in the Evaluation Steering Committee Workshop in Aspen in August 1994. I'd also like to acknowledge the originator of the title; it was, of course, Kurt Lewin who said that there is nothing as practical as a good theory.

2. Some people are concerned that without experimental design (or some close approximation to), evaluations will not yield valid conclusions. Others worry that good data are not available at the community level to use as markers of success, and that evaluators will settle for small-area data of doubtful quality and unknown reliability. Another worry is that the selection of indicators can distort the work of CCIs. Just as teachers can "teach to the test," CCIs can work on those issues that will be measured, rather than on issues that would yield greater benefit to the community. Still other observers wonder whether local residents and service providers are having adequate say in the definition of the outcomes (and the measures) that will render judgment on their efforts. Some people recommend an emphasis on qualitative evaluation, which has the advantages of enabling the evaluator to follow the dynamics of program development and to understand the perspectives of the participants and the meanings they attach to events. However, qualitative evaluation of large-scale CCIs is time-consuming and expensive, and to be feasible, it would have to be highly selective in focus. Moreover, qualitative reports might not have the immediate credibility that quantitative reports command among decision-making audiences. The discussion about appropriate evaluation methods goes on.

3. From time to time in this inventory of assumptions, I interject a contrary note, as in the reference to conflicting pressures on service staff. This is not to express my own beliefs (heaven forfend) but to recognize the status of these assumptions as hypotheses. While I try to represent the beliefs of CCI advocates fairly as I read and heard them, caution seems to be in order before we let the beautiful rhetoric sweep aside our sense of reality.

4. I thank Ron Register for suggesting this point.

5. I thank Martin Gerry for reminding me of this point.

REFERENCES

Arizona Department of Economic Security. 1989. *Arizona Community Services Integration Project: Final Evaluation Report.* December. Phoenix, Ariz.: Arizona Department of Economic Security.

Bane, Mary Jo, and D. T. Ellwood. 1994. *Welfare Realities: From Rhetoric to Reform.* Cambridge: Harvard University Press.

Brodkin, Evelyn Z., Debra Hass, and Alexander Kaufman. 1993. "The Paradox of the Half-Empty Glass: Speaking Analysis to Poverty Politics." Paper prepared for annual meeting of Association for Public Policy and Management, Washington, DC, October 1993.

Chaskin, Robert J. 1992. *The Ford Foundation2 Neighborhood and Family Initiative: Toward a Model of Comprehensive Neighborhood-Based Development.* April. Chicago: Chapin Hall Center for Children at the University of Chicago.

Chen, Huey-Tsyh. 1990. *Theory-Driven Evaluations.* Newbury Park, Calif.: Sage Publications.

Chen, Huey-Tsyh, and Peter H. Rossi. 1987. "The Theory-Driven Approach to Validity." *Evaluation and Program Planning* 10: *95-103.*

Enterprise Foundation. 1993. **Community** *Building in Partnership: Neighborhood* Transformation *Demonstration, Sandtown-Winchester, Baltimore.* Progress Report, March 1993, typescript. Baltimore: Enterprise Foundation.

Friedlander, Daniel, and Gayle Hamilton. 1993. *The Saturation Work Initiative Model in San Diego: A Five- Year Follow-up Study.* July. New York: Manpower Demonstration Research Corporation.

Hamilton, Gayle, and Daniel Friedlander. 1989. *Final Report on the Saturation Work Initiative Model in San Diego. New* York: Manpower Demonstration Research Corporation.

Haveman, Robert H., ed. 1977. *A Decade of Federal Antipoverty Programs: Achievements, Failures, and* Lessons. New York: Academic Press.

Haveman, Robert, and Barbara Wolfe. 1994. *Succeeding* Generations: On *the Effects of Investments in Children. New* York: Russell Sage Foundation.

Jencks, Christopher. 1992. *Rethinking Social Policy: Race, Poverty, and the Underclass.* Cambridge: Harvard University Press.

Lipsey, Mark W. 1993. "Theory as Method: Small Theories of Treatments." In *Understanding Causes and Generalizing About Them,* ed. L. B. Sechrest and A. G. Scott. New Directions for Program Evaluation, No. 57, pp. 5-38.

Pew Charitable Trusts. n.d. *The Children 's Initiative: Making Systems Work, A Program of The Pew* Charitable *Trusts.* Typescript. Philadelphia: Pew Charitable Trusts.

Rockman, Bert A. 1994. "The New Institutionahsm and the Old Institutions." In New *Perspectives on American Politics* ed. L. C. Dodd and C. Jillson, pp. 143-61. Washington, DC: CQ Press.

Rostow, W. W. 1993. "The Austin Project, 1989-1993: An Innovational Exercise in Comprehensive Urban Development." Paper prepared for Seminar on Inner City Poverty, Yale University Institution for Social and Policy Studies, October 1993.

Schorr, Lisbeth B. 1988. *Within Our Reach: Breaking the Cycle of Disadvantage. New* York: Anchor Press/Doubleday.

_____. 1991. "Attributes of Effective Services for Young Children: A Brief Survey of Current Knowledge and its Implications for Program and Policy Development." In *Effective Services for Young* Children, ed. L. B. Schorr, D. Both, and C. Copple. Washington, DC: National Academy Press.

_____. 1994. Personal communication, August 9.

Shadish, W. R., Jr. 1987. "Program Micro- and Macrotheories: A Guide for Social Change." In *Using Program Theory in Evaluation,* ed. Leonard Bickman. New Directions for Program Evaluation, No. 33, pp. 93-108.

State Reorganization Commission (South Carolina). 1989. *An Evaluation of the Human Service Integration Project, 1985-1988* December. Columbia, S.C.: State Reorganization Commission.

Stephens, S. A., S. A. Leiderman, W. C. Wolf, and P. T. McCarthy. 1994. *Building Capacity for System Reform.* October. Bala Cynwyd, Pa.: Center for Assessment and Policy Development.

Walker, Gary, and Frances Vilella-Velez. 1992. *Anatomy of a Demonstration: The Summer Training and Education Program (STEP) from Pilot through Replication and Postprogram Impacts.* Philadelphia, Pa.: Public/Private Ventures.

Weiss, Carol H. 1972. *Evaluation Research: Methods of Assessing Program Effectiveness.* Englewood Cliffs, N.J.: Prentice-Hall.

Wilner, Daniel, and others. **1962.** *The* Housing *Environment and Family Life: A Longitudinal Study of the Effects of Housing on Morbidity and Mental Health.* Baltimore: Johns Hopkins Press.

Wilson, William Julius. 1987. *The Truly Disadvantaged The Inner City, the Underclass,* and *Public Policy.* Chicago: University of Chicago Press.