

---

---

---

# Evaluation of School-linked Services

Deanna S. Gomby  
Carol S. Larson

## Abstract

Evaluation of school-linked service initiatives, which are characterized by great flexibility and variability, is challenging but also possible and desirable. Indeed, every school-linked service effort should undergo some level of evaluation, whether for the purpose of honing an existing program or for providing evidence of its effectiveness. Evaluations of previous school-linked service programs offer limited support for the school-linked service movement and indicate how complex programs placed in the schools can be evaluated with sensitivity and rigor. Evaluation can serve as a useful tool to program providers, policymakers, and funders, but each group must make significant commitments to ensure a meaningful and high-quality evaluation. Expensive, unevaluated programs that are continued year after year and that are based only on hunches or political winds can represent a waste of millions of dollars as well as lost opportunities to try what could well be more effective approaches.

*Systematic evaluation is increasingly sought to guide operations, to assure legislators and planners that they are proceeding on sound lines, and to make services responsive to their publics. Evaluation has thus become the liveliest frontier of American social science.<sup>1</sup>*

*Two additional articles on evaluation of school-linked services are available upon request from the Center. The first is a literature review of previous studies of school-linked service programs. The second is a primer on evaluation, with a special focus on conducting evaluations of school-linked service programs.*

*Deanna S. Gomby, Ph.D., is director of research and grants for child development at the Center for the Future of Children.*

*Carol S. Larson, J.D., is director of research and grants for law and public policy at the Center for the Future of Children.*

**O**ver 10 years have passed since these words were written, and program evaluation remains a frontier that is both lively and less than fully explored. On that frontier, school-linked service efforts are outposts in need of improved evaluation—evaluation that will furnish information to service planners, providers, beneficiaries, and funders.

## The Importance of Evaluation

As many of the articles in this journal issue discuss, school-linked service programs are designed to offer multiple and flexible services to children and their families through collaborative partnerships among school, health, and social services agencies. Ideally, each effort is shaped according to the needs and resources of the community as well as the needs of the individuals served. (See the articles by Gardner, by Levy



and Shepardson, and by Jehl and Kirst in this journal issue.) Thus, by definition, each school-linked service effort differs, both in terms of services offered and administrative structure. Evaluating programs that are characterized by such flexibility and variability is certainly a challenge, but it is not impossible.

The premise of this article is that evaluation of school-linked service efforts needs more attention. Indeed, every effort—regardless of size—should include some level of evaluation. As discussed below, for some programs, this evaluation will consist primarily of descriptive data collection; for others, a carefully designed outcome evaluation will be possible.

Evaluation of school-linked services is important for at least two reasons:

1. Evaluation can provide information about whether the school-linked service approach is effective and/or worth the investment. Implementation of school-linked service efforts requires significant changes in funding, utilization of personnel, and services. Such change is rarely without cost, and backers of school-linked services should be able to demonstrate that the direct and indirect costs inherent in the new programs are warranted on the basis of demonstrated outcomes. A well-crafted evaluation can help determine if a new school-linked service initiative generates outcomes for children, families, and communities that are better than those generated by either existing service delivery systems or new service arrangements (for example, by a community-based, integrated service program). Information from evaluations is important not only for policymakers but also for all those involved in providing services. Teachers and other staff want to know whether the changes they have made are improving the lives of children.

2. Evaluation can also provide information about how best to implement a program. Apart from comparing the outcomes of a school-linked service program with those of some other service system, an evaluation can also be structured to help those implementing a program identify areas where they are meeting or ex-

ceeding goals or where the program should be modified to improve service to students and families. In a field like school-linked services, where most programs are new or just getting under way, evaluation can play an invaluable role in honing programs.

---

*Every school-linked service effort—regardless of size—should include some level of evaluation.*

---

This article discusses the need for the evaluation—indeed, for a range of evaluation strategies—of school-linked service initiatives. Previous evaluation efforts are described briefly. Although previous efforts suggest limited support for some key tenets of new school-linked service initiatives, they demonstrate the difficulty even the best programs have in changing people's behavior. They also illustrate how school-linked service programs can be evaluated. Finally, this article discusses the considerable commitment program providers, policymakers, and funders must make to conduct meaningful evaluations.

### **Types of Evaluation**

One of the first steps in launching an evaluation of school-linked services is to determine the goals of the evaluation. Is the purpose of the evaluation to describe the services the program provided and

who received them? Or to determine if and for whom the services made a difference? Once those goals are selected, a number of strategies of varying appropriateness, comprehensiveness, and cost are available to accomplish them.

Evaluations are often divided into two general types, both of which can be useful in assessing a school-linked service initiative: process (formative) evaluation and outcome (summative) evaluation.

#### Process Evaluation

A process evaluation focuses on what services were provided to whom and how. Its purpose is to describe how the program was implemented—who was involved and what problems were experienced. A process evaluation is useful for monitoring program implementation; for identifying changes to make the program operate as

planned; and, generally, for program improvement. Figure 1 further describes this type of evaluation and its goals. (A more thorough discussion of this and other aspects of evaluation is available upon request from the Center.)<sup>2</sup>

In a complex school-linked service collaboration among a number of child-serving agencies, a process evaluation may also examine how the agencies interacted during implementation. Such an assessment might document changes in relationships among social and health service providers, educators, and client families. A process evaluation might document system change as evidenced by new intake procedures, new forms, memoranda of understanding, or interagency linkage agreements. In some cases, changes in relationships and systems may be specifically planned goals of the program. In other

### Figure 1. BASICS OF EVALUATION: Purpose of Evaluation

**Intuitively, program planners and policymakers want to learn at least three things from a program evaluation:**

- What services did the program provide?**
- Who received the services?**
- Did the services make a difference?**

In fact, depending upon the purposes of the evaluation, answering only one or two of these questions may be sufficient. If the primary purpose of the evaluation is to help program staff hone a new program, then answering the first two questions may be enough. Answers to those questions will determine if the program is delivering the services it intended to deliver and if those services are reaching the intended audience. If, on the other hand, the purpose of the evaluation is to persuade funders to continue funding a program, to expand services, or to fund a spin-off of an existing program in a new site, then the third question may become increasingly important.

This division is roughly analogous to the oft-cited division between *process* (or formative) and *outcome* (or summative) evaluations.

**Process evaluations** focus on descriptions of the “activities of service programs and their relation to program goals and objectives.”\*

**Outcome evaluations** are studies to determine the effectiveness or outcomes of service programs. They must be designed to

- 1) measure whether anticipated changes occurred, and
- 2) prove that the changes were caused by the program and not something else. Designing an evaluation that is rigorous enough to rule out all the alternative causative explanations requires considerable time, effort, expense, and commitment from all concerned. (Figure 3 illustrates some of the techniques that are used to demonstrate causation.)

The best outcome evaluations will include and be informed by components of process evaluations. Even evaluations whose primary goals are to determine if services were effective should include enough measures of process to help explain whatever results are eventually obtained.

\*J. Pawl, K. Barnard, A. Korner, et al., *Charting Change in Infants, Families, and Services: A Guide to Program Evaluation for Administrators and Practitioners*, National Center for Clinical Infant Programs, 1987, p. 5.

cases, such changes may only be intermediate outcomes accomplished en route to the outcome of most interest to program planners: changes in the students.

#### **Outcome Evaluation**

In most cases, the primary motivation for creating and funding school-linked service efforts is to accomplish specific student-related goals. Examples of these goals include improvement of academic performance or reduction of problems such as drug use or teen pregnancy. In contrast to process evaluation, an outcome, or summative, evaluation determines whether the services that were provided to and used by the students led to the desired changes in the participating students. Typically more complex and expensive than a process evaluation, an outcome evaluation can be undertaken only if (1) there is a clear statement about what changes are expected, (2) appropriate measures are selected for tracking such change, and (3) a mechanism is established to collect reliable data about these outcomes. Figure 1 discusses outcome evaluations. Figure 2 lists the issues involved in selecting and measuring outcomes.

The purpose of an outcome evaluation is not only to measure changes in outcomes but to establish that the intervention (that is, the school-linked service program) caused the changes. Collecting data about targeted outcomes does not necessarily provide information about causation. For example, data may show that test scores of students increased or drug use by students decreased after school-linked services were introduced at a school, but this information alone cannot prove that the services provided caused the change. Other factors may have been the real causes—for example, a change in the composition of the student body or a community-wide change in attitude caused by the drug-overdose death of a celebrity.

The most crucial decisions in an evaluation to establish causation involve its design—that is, the decisions about what will be measured and when those assessments will occur. Only an evaluation with a rigorous design can establish causal links between the intervention and the observed outcomes.

Evaluators employ a variety of research designs to structure outcome evaluations, ranging from simple pretest/posttest designs with one group to complicated

multigroup and multisite designs. Some of the more generic designs are discussed in figure 3.

As a general rule, to help establish a causal link between the provided services and observed changes and to eliminate alternative explanations of outcomes, evaluators must compare students or families who did not receive the services with similar students or families who did. Researchers often use one of two techniques to construct a comparison group: Researchers (1) match the group (such as students, classrooms, schools, or communities) receiving services with a similar group that does not receive program services or (2) they use random assignment to decide which members of the target group receive program services and which do not.

Random assignment is the most reliable technique available to guarantee that the group receiving program services (the intervention group) and the group not receiving services (the control group) are initially equivalent. The effects of the program are then measured by comparing the amount of change displayed in each of the two groups. Figure 4 presents further discussion of random assignment, reasons why opposition to it sometimes runs high, and suggestions for implementation of research designs employing it.

#### **Selection of Appropriate Evaluation Strategies for School-linked Services**

As stated earlier, no school-linked service effort, no matter the size, should be undertaken or funded without including some level of evaluation. Of course, it is unrealistic to expect that all efforts—or even many—will have sufficient funding and be large enough to embrace a high-quality evaluation with random assignment, control groups, and well-validated measures of outcome to establish causal links between the intervention and outcomes.

However, it is not unrealistic to expect that every school-linked service initiative adopt clearly stated goals both for process and outcomes, have the means to collect some amount of data related to these goals, and undertake some form of process evaluation. The National Center for Clinical Infant Programs offered advice regarding the evaluation of service programs for infants, toddlers, and their families, and that advice applies to the evaluation of school-linked service efforts in general:

*Continued on page 75.*

## Figure 2. BASICS OF EVALUATION: Choosing and Measuring Outcomes

Selecting the outcomes to document is an early step in conducting an evaluation.

**Should changes in the individual participant, the family, the service system, and the community be assessed? Should cost savings be monitored?**

Most programs are guided by implicit or explicit theories of how the planned interventions will lead to changes in knowledge, attitudes, or behavior in individuals, families, or communities. In some cases, program developers may regard changes in knowledge or attitudes as the endpoints of interest. In most cases, however, program developers, funders, and policymakers are interested in those types of outcomes only if they are predictive of subsequent changes in behavior; it is the behavior changes that are the ultimate goals.

Before evaluators assess outcomes of a program, they frequently try to determine if the intended services (the intervention) actually were delivered to the intended audience. If the intervention was not implemented as intended, then any changes in knowledge, attitudes, or behavior that are observed later cannot be attributed to the intervention—at least not to the originally proposed intervention. After evaluators determine that the planned program operated as it was intended to operate, evaluators usually turn to assessments of outcomes.

For example, consider a program to prevent school drop-out by decreasing class size and providing a counselor for each student. At a minimum, the measured outcomes should be class size and provision of counseling, to ensure that the intended intervention occurred. Other outcomes might include the attitudes of the students and teachers toward one another, actual dropout rates after one or more years of the program, and, perhaps, rates of employment in the community. Depending upon the scope and intensity of the intervention, one or more of these outcomes may be expected.

Most programs, including most school-linked service programs that have been evaluated to date, have focused on changes in the student participants. Most evaluations have sought to determine if the services led to changes in student knowledge, attitudes, and/or behavior. Change in knowledge and attitude is easiest to document through the use of paper-and-pencil questionnaires but is not necessarily predictive of behavior change. Programs have therefore increasingly tried to assess changes in behavior itself. Furthermore, programs have begun to use objective measures of behavior rather than self-reports from program participants.

The following is a list of possible outcomes—that have been used in previous studies and/or might be appropriate to use to assess new school-linked service initiatives:

### Examples of Outcomes Used to Assess School-linked Service Initiatives

General Outcome	Specific Outcome
<b>THE STUDENT</b>	
School Achievement	Grade point average Standardized test scores Absenteeism Dropout rates School re-entry rates for initial dropouts Student attitudes toward school Teacher ratings of student performance Observation of in-class behavior
Teen Pregnancy	Student attitudes toward sex and contraception Student knowledge about sex and contraception Rates of contraceptive use Rates of pregnancy (self-report; clinic records) Rates of repeat pregnancies Live birth rates

**Figure 2. (continued)**

<b>General Outcome</b>	<b>Specific Outcome</b>
Social Skills	Ability to withstand peer pressure (in role playing, or in observations of behavior in class or on playground)
Cardiovascular Fitness	Diet (self-report; observation of school lunches or food in refrigerator at home; analysis of salt/fat content of school lunches)
	Cholesterol level
	Blood pressure level
	Percent overweight
	Exercise levels (self-reported or observed)
Mental Health	Suicide rates
	Scores on depression scales
	Scores on self-esteem scales
Drug, Alcohol, and Cigarette Use	Rates (self-reported; breath, saliva, urine tests)
	Student attitudes toward use
	Student knowledge about effects of drugs
Job Placement	Job rates while enrolled in school
	Job rates after exit (post-graduation; post-drop-out)
<b>THE SERVICE SYSTEM</b>	
Utilization of Services	Services offered by collaborating agencies
	Services used by participants
	Referrals to other agencies
	Services delivered by other agencies
	Patterns of utilization across different groups of students/parents
Costs	Sources of funding
	Amount of funding
	Direct costs of services/personnel
	Indirect costs of services/personnel
Interagency Collaboration	Existence of memoranda of understanding
	Frequency of meetings among participating agencies
	Existence of steering committee with representation from collaborating agencies
	Existence of waivers to document changes in funding streams
Streamlined Procedures	Existence of new, simpler forms
	Number of contacts families have with multiple agencies
	Time spent waiting for services
<b>THE FAMILY</b>	
Child Abuse or Neglect	Rates for program participants (reports to child protective agencies; court decisions; removal of child from biological family)
Home Environment	Ability to promote child development (self-report, observation by home visitor)
	Safety (self-report; observation by home visitor)
Parent-Child Interaction	Style and content (self-report; observation)
Connection with Community Institutions	Parental attitudes toward schools and collaborating agencies
	Parental knowledge about community services available
Cardiovascular Fitness (parental)	Blood pressure
	Cholesterol level
	Pulse rate
	Percent overweight
	Exercise levels

**Figure 2. (continued)**

General Outcome	Specific Outcome
<b>THE COMMUNITY</b>	
Media	Number of stories on school-linked initiative Number of stories on issues addressed by the initiative
Teen Pregnancy	Community-wide rates
Child Abuse and Neglect	Community-wide rates
School Drop-Out	Community-wide rates
Unemployment	Community-wide rates
Advocacy for Children and Families	Rates of volunteerism Amount of funding for children's programs

**Figure 3. BASICS OF EVALUATION: Designing an Evaluation**

Once outcomes and measures are selected, the overall design of the evaluation must be determined. The choices about what will be measured and when those assessments will occur are perhaps the most critical in ensuring the quality of the evaluation.

The strength of the design will determine the extent to which it will be known that the intervention (the services that were provided), and not other factors, caused the observed outcomes. The following illustrates some typical evaluation designs. The most rigorous designs utilize random assignment to create control and experimental groups. Random assignment is discussed further in figure 4.

**X** = Intervention (the program services)  
**O** = Observation, either a pretest, posttest, or follow-up assessment, using observation of behavior, paper-and-pencil questionnaires, or physiological assessments  
**O<sub>1</sub>** = Observation at time 1  
**O<sub>2</sub>** = Observation at time 2  
**Group 1** = Intervention group  
**Group 2** = Control or comparison group

Notation style based on D.T. Campbell and J.C. Stanley, *Experimental and Quasi-experimental Designs for Research*, 1963.

### 1. One-Group, Posttest-Only Design

Group 1: X O

In this design program participants are tested after receiving services.

*Comments:* Observed changes cannot be attributed to the intervention (X) because (1) the initial starting point for the group is unknown; and (2) it is possible that the group would have changed anyway.

### 2. One-Group, Pretest and Posttest Design

Group 1: O<sub>1</sub> X O<sub>2</sub>

In this design program participants are tested both before and after receiving services.

*Comments:* Any observed changes could be due to something else that happened at the same time as the intervention. Group scores on a posttest could have changed from pretest levels even without the intervention. This design is especially risky when used to assess children, in whom developmental changes occur as part of normal maturation.

**Figure 3. (continued)**

**3. Two-Group (Matched), Pretest and Posttest Design**

Group 1:  $O_1$  X  $O_2$   
 Group 2:  $O_1$   $O_2$

In this design a comparison group is constructed. Both groups are tested at two points in time, but only one group receives the services.

*Comments:* The two groups are matched initially on a set of characteristics thought to be important (typically, ethnicity, socioeconomic status, pretest scores). Nevertheless, it is always possible that the groups differ on some important, but unknown and unmeasured, characteristic. If that characteristic is associated with change in the targeted outcome, then the groups might look different after the intervention, not because of the intervention, but because of the influence of the unmeasured characteristic.

**4. Two-Group (Randomly Assigned), Posttest-Only Design**

Group 1:  $O_1$  X  $O_2$   
 Group 2:  $O_2$

In this design a pool of participants are randomly assigned to two groups, one of which receives program services.

*Comments:* Random assignment (assigning potential program participants to groups by utilizing techniques akin to flipping a coin) is the best way to ensure that groups are equivalent prior to the intervention. Because the intervention and control groups are assumed to be equivalent initially, no pretest for the control group is necessary.

More elaborate designs using random assignment can include pretests for the control group and/or multiple intervention and control groups. Groups may consist of individuals, classrooms, schools, or communities. In some projects, members of the control group are placed on a waiting list until the second measurement ( $O_2$ ) occurs. They then receive program services.

*Continued from page 71.*

All service providers should be engaging in some form of ongoing process evaluation, whether or not they are ready for a formal outcome evaluation. A program's own staff can design and implement ongoing record-keeping activities that not only keep the program accountable to families and sponsors but that also suggest avenues for program improvement.<sup>3</sup>

In addition to such process evaluations, every school-linked service effort, regardless of size, should at least take beginning steps toward collecting information about behavioral outcomes and systems change. At a minimum, this requires planners and providers to identify clearly which changes they intend to produce in the behavior of students or in the procedures of cooperating agencies and how these changes can be measured. Furthermore, before services begin each program should collect baseline data on these measures and update the data each year.

As part of this ongoing monitoring, some school-linked service efforts may also be able to conduct a pre- and postservice

assessment of students who received services. Other efforts may even be able to construct a comparison group that did not receive the intervention against which to compare results.

But to determine the effectiveness of school-linked services requires more than these less formal attempts at outcome evaluation. There must also be comprehensive, well-funded, and well-designed outcome evaluations that use random assignment or appropriate comparison groups. In the long run, the best evidence for the effectiveness of school-linked services will be produced by a convergence of results from a number of evaluations using different designs.<sup>4</sup>

Statewide or multisite efforts at school-linked services afford a particularly rich opportunity for learning. This opportunity should not be squandered. Although each program site differs, policymakers, funders, and service providers should actively develop a comprehensive evaluation plan to take advantage of their diversity and answer a wide range of important questions. For example, when sites offer different services, evaluation can help dis-

cover which services are most effective. When sites serve different types of students, evaluation can help discover how best to deliver services to students of different backgrounds. When sites differ in the amount of initial training of program staff or in the amount of training offered by the program, evaluation can reveal how much training is necessary for the most effective implementation. When pilot programs move from individual sites to statewide efforts and scarce dollars must be spread thinner than ever, evaluation can help detail the effect service dilution or diminished funding per site has on program effectiveness.

Multisite evaluations also afford the opportunity to compare alternative service delivery systems such as the fragmented existing system; comprehensive, school-linked service programs; or some other community-based, integrated service delivery system. Even if the existing system is inadequate, changing the system will not necessarily lead to improvement. And, since change typically involves cost, advocates for change must show that the costs of system-wide reform are justified.

There are several important considerations and obligations for funders, policymakers, and evaluators when designing

#### Figure 4. BASICS OF EVALUATION: Random Assignment

Random assignment is the most reliable technique available to guarantee that the group receiving program services (the intervention group) and the group not receiving services (the control group) are initially equivalent. Despite the methodological advantages, using a comparison or control group can engender resistance for several reasons.

First, including any comparison group adds to the expense of the evaluation because measuring two or more groups is more costly than assessing only one. Funders must acknowledge the extra expense and be willing to cover it.

Second, using random assignment to create a control group can generate ethical and political dilemmas for service providers who may be reluctant to withhold services from any students who they believe need them. There are strategies, however, to reduce problems: (1) Establish criteria for entrance into the pool of subjects who will be randomly assigned to receive services. For example, the criteria could exclude the most needy students (such as pregnant teens under the age of 14) from random assignment to ensure that they will always receive services; (2) Randomly assign all eligible students into an intervention group and a wait-listed control group. Rather than never receiving services, members of the control group are assured that they will eventually participate in the program after the impact of the intervention is measured; or (3) Provide different or less intensive services to individuals in the control group.

##### Random assignment at the school or classroom level

Sometimes, rather than randomly assigning individuals to experimental conditions, evaluators will suggest that whole schools or classrooms be assigned to either an intervention or control group. This approach is used when it seems impossible to keep individuals in one group from learning about and being affected by the services being delivered to individuals in another group. This is often necessary, for example, when the intervention being evaluated involves school- or community-wide changes. With such interventions, evaluators may need to employ random assignment at the level of the school or community.

When random assignment occurs at the level of the institution, planners and providers may be concerned that a rigorous experimental design will tie their hands in terms of the services that they will be able to provide to students and their families. That is, they may fear that the experimental design will require a specific and consistent set of services to be delivered to every student in the experimental group. Such an approach would conflict with the basic tenet of the school-linked service approach of giving line workers more discretion to tailor services to the particular needs of the child or family.

Such flexibility does not make an outcome evaluation impossible, but it does make a better than average record-keeping system necessary, and it does change what questions the evaluation can answer. If positive changes are noted in the experimental group and not in the untreated control group, one knows that some combination of services made a difference, although one usually cannot tell which services caused the change. With the addition of an excellent record-keeping system, it may be possible to note some tentative associations between services and observed changes, but those associations are probably more likely to remain tentative than conclusive. In other words, the evaluation will be able to answer the global question, Did the whole new package of services lead to change? but not the more specific question, Which of the provided services were most important?

evaluations for school-linked services. However, before discussing these, we turn to a discussion of the considerable history of evaluating school-linked services. This literature provides some evidence that school-linked service initiatives hold promise and that evaluating the initiatives is indeed possible and desirable. The literature also indicates, however, that the difficulty of changing human behavior should not be underestimated.

## Previous Evaluations of School-linked Service Programs

This section briefly summarizes the lessons offered by the rich existing literature on school-linked services. The summary is based upon a literature review (available upon request from the Center)<sup>5</sup> of approximately 10 years of published studies on school-linked service programs.

Most of the hundreds of such studies concern single-focus, single-service programs based at schools. Some programs focused on children from low-income neighborhoods, but most research projects were conducted in white, middle-class areas. Although some programs focused on children of a particular age range, the literature shows that school-based services as a whole have been offered to children from grades K-12. By far the vast majority of published research literature chronicles the classroom-based, curriculum-driven approach. Except for these similarities, however, the literature documents a great variety of programs. These diverse efforts

---

*It is difficult and perhaps impossible to examine the literature about school-linked services through a single conceptual prism.*

---

addressed problems as varied as teen pregnancy; dropping out of school; use of alcohol, cigarettes, and illicit drugs; teen suicide and depression; reactions to parental divorce; child abuse; obesity; cardiovascular disease; and inadequate access to medical care. The programs utilized many strategies, including classroom-based curricular interventions (some emphasizing factual information, others teaching life

skills), new curricula as well as provision of direct services such as health care or counseling to students, direct services alone, and the coupling of curricula with services to students and families throughout a community.

The providers and participants in school-linked service programs have also varied greatly. Some programs were administered by schools, others by university researchers or community agencies working in conjunction with schools, and others by public agencies. The services were delivered by school personnel and by outsiders, adults, or students' peers, and by combinations of these.

These diverse efforts were, for the most part, isolated projects. They were not designed as part of a comprehensive strategy to determine the most effective method of serving students. Thus, it is difficult and perhaps impossible to examine the literature about school-linked services through a single conceptual prism. Nevertheless, the history of school-linked service evaluation does suggest some lessons that may help guide current initiatives and the evaluation of them. These lessons are discussed briefly below.

### The Promise of Comprehensive, School-linked Services

Proponents of school-linked services contend that achieving better outcomes for children and families requires at least three elements: comprehensive services, increased involvement of parents, and changes to make schools and agencies more responsive to children and families. In two ways, the published literature on single-focus, school-based programs provides some support for this contention. First, the literature indicates how hard it is to change behavior through traditional classroom-based approaches that do not include the three elements. Second, the literature shows that programs that do employ the elements can lead to a change in student behavior.

For example, the literature indicates that a fact-based curriculum alone is not sufficient to change student behavior. Initial efforts to prevent risky health behaviors—smoking, alcohol or drug abuse, or teen sex, for example—employed heavily information-based or fear-laden messages. Gradually, research discredited these curricula because they tended to influence student knowledge but not behavior.<sup>6-9</sup> Today, the most up-to-date programs in-

volve curricula that are based on social learning theories. Such curricula provide students with information as well as training in social skills to resist pressures from peers, family, or the media; with skills to make thoughtful decisions about health behaviors; and with opportunities to role-play and practice their skills.<sup>10</sup>

Recently, however, even these more comprehensive programs have been combined with efforts to enlist parental, community, and media support, and there is evidence that these broader initiatives are effective. For example, parental and/or community involvement has led to positive results in programs designed to improve cardiovascular fitness,<sup>11</sup> to help obese elementary school students lose weight,<sup>12</sup> to deter young children from smoking,<sup>13</sup> and to prevent teen pregnancy.<sup>14</sup>

Similarly, programs that change the school structure to make it more student-focused (for example, to place classes close to one another, to provide a teacher who acts as a resource for a student and meets regularly with him or her, and to keep students in small groups that foster comfortable student-peer and student-teacher relationships) demonstrate promising decreases in absenteeism, lower dropout rates, better academic performance, and improved self-concept among students.<sup>15-17</sup>

### **The Difficulty of Changing Human Behavior**

As mentioned earlier, a purely information- or fear-based approach is largely ineffective in altering behavior though it may increase knowledge. Indeed, in most domains (for example, use of drugs, alcohol, and cigarettes; exercise; diet; teen pregnancy prevention),<sup>7,9,18-20</sup> the literature indicates only a fairly modest relationship among changes in knowledge, attitudes, and behavior. When programs led to behavior change, the changes were often small and usually involved delaying the onset of a behavior such as smoking rather than preventing it altogether.<sup>21</sup>

Similarly modest behavior changes were also achieved in school-based health clinics, often the most comprehensive school-based services. After more than a decade of experimentation and some fairly rigorous evaluation, the results achieved by school-based health clinics are mixed. Although these clinics have increased access to and utilization of basic physical and mental health services, their

ability to change students' reproductive behavior or substance use has been minor.

For example, the most comprehensive study to date of school-based clinics found that, although clinics did not hasten the initiation or increase the frequency of sex among students, neither did they promote greater use of contraception or lead to decreased pregnancy rates.<sup>22,23</sup> Clinics were associated with increased contraceptive use only when schools or the surrounding community made the prevention of pregnancy or AIDS a special focus. Despite predictions that school-based clinics might prevent some mild problems from becoming serious, clinics did not have a significant impact upon the use of emergency rooms, the number of nights spent in the hospital, or school absenteeism. The clinics also had mixed effects on students' self-reported use of cigarettes, alcohol, or illicit drugs.

Although the evaluations of school-based clinics may seem discouraging, the results are consistent with the other findings discussed above and may be consistent with the contentions of the proponents of school-linked services. It is very hard to change behavior, but the most promising opportunities to do so may involve school-wide and community-wide mobilization.

### **Elements of Effective School-linked Service Programs**

The published literature about curriculum-based programs suggests some key elements of successful program implementation. These elements include adequate training of program providers; sufficient intensity of services; and, in some cases, peer rather than adult providers. Programs in which providers received poor training (such as a half-day workshop), for example, were usually not as successful as those in which providers received more extensive training (such as a day-long session with periodic review meetings).<sup>10</sup> Some apparent program failures were later traced to inadequate implementation (for example, not covering core topics, changing the curriculum, or spending less time on the program than planned).<sup>24,25</sup>

Generally, in classroom-based alcohol-use and smoking prevention programs, intensive services (for example, 10 to 15 sessions in 1 year with booster sessions continuing over 2 to 3 years) were more effective than less intensive services (for

instance, four class lectures without follow-up).<sup>10,26</sup> The required intensity of services for other domains and other service strategies is unknown.

Some studies of curricula considered whether efforts were more effective if led by teachers or by outside experts, by adults or by peers.<sup>27-30</sup> No clear results emerged, but many programs now use teachers or adults to guide the program and incorporate in-class peers to add information.

Research indicates that training, service intensity, and program leadership are important factors in the success of curriculum-based, school-linked service programs; these three elements may also be relevant to the new school-linked service efforts, which are not primarily in-class, curriculum-based efforts. To verify that these relationships hold for other sorts of school-linked service initiatives, future evaluations must investigate these elements directly.

### The Evolution of Quality Evaluation

The existing school-linked service evaluation literature varies greatly in quality. Over time, the methodology used became more rigorous, especially for curriculum-based efforts designed to prevent smoking or drug and alcohol use or to promote cardiovascular fitness. (Perhaps, because those efforts received the largest amounts of federal funding, researchers could afford the time and commitment required for rigorous evaluation.) Rather than using one-group pretest and immediate posttest designs, for example, researchers increasingly relied on more sophisticated designs involving random assignment of whole schools to groups and follow-ups over longer periods of time. Increasingly, researchers used advanced measurement techniques (such as actually observing children's lunches<sup>31</sup> or measuring blood cholesterol levels<sup>32</sup> rather than relying on self-reports about diets or actually measuring biochemical markers for drug or cigarette use rather than relying on student self-reports<sup>11,33,34</sup>). Nevertheless, critics noted that many studies of curriculum-based approaches were plagued with one or more of the following methodological problems: weak pre-post or nonrandomized designs; small, convenient samples rather than large, representative groups; poor measurement instruments; assessments of knowledge or attitudes rather than actual behavior; use of self-reports with limited or no physiological validation;

high attrition rates; and poor implementation of experimental intervention.<sup>18,35,36</sup>

The evolution of evaluation has also included broader focus on questions other than the overall effectiveness of a program. For example, some recent investigations explored implementation questions such as how intensively services should be offered or how best to train providers.<sup>37</sup> Other efforts investigated questions of generalizability—that is, how well programs can be transferred to different settings. For example, studies indicated that the same general curriculum approaches are not equally effective at deterring the use of cigarettes versus alcohol<sup>9,21,38</sup> and that the same program may have very different rates of effectiveness with different groups of children<sup>38-41</sup> or if presented by different providers (for instance, adults versus peers, schoolteachers versus members of the outside research team).<sup>10,28,30</sup>

Finally, the literature demonstrates the importance of a longitudinal approach. Few studies have followed children over a number of years to see if program effects are maintained over time. Those that did indicate that results of fairly intensive curriculum programs may be maintained for as long as 4 to 5 years, although differences between control and intervention groups disappear about 6 years after the intervention.<sup>41,42</sup>

## Current and Future Evaluations of School-linked Services

These lessons from previous evaluation efforts are important to keep in mind when planning current and future evaluations of school-linked services. They provide helpful insight into how difficult it can be to change human behavior and to design high-quality evaluations.

### Evaluation of Current Multifocus, School-linked Service Programs

Current efforts at providing school-linked services differ in significant ways from those described in the published literature. Today, multiagency collaborations offer or coordinate multiple education, health, and social services at or near the school site. These efforts usually have several goals that include not only improving student outcomes (such as reduced dropout rates, improved academic performance, and decreased substance abuse), but also family outcomes (such as improved

parent-child relationships) and systems outcomes (such as better working relationships among education, health, and social services agencies).

Given the more complex nature and goals of current efforts compared to earlier ones, evaluating them is also more complex. Appendix B lists 16 current initiatives to deliver school-linked services to school-age children and briefly describes their evaluations. As the appendix shows, evaluation of these efforts is in a prelimi-

---

*Given the more complex nature and goals of current efforts compared to earlier ones, evaluating them is also more complex.*

---

nary stage, just as evaluation of single-focus interventions was during the late 1970s and early 1980s. Some programs have not undertaken any formal evaluation at all; some have begun to compare data about students served with data about other groups of similar students; and some are currently planning more formal outcome evaluations. Given the scope and complexity of the initiatives, it is not surprising that methodologically rigorous evaluations have not yet been conducted. Nevertheless, it is best to incorporate evaluation considerations in initial planning for any school-linked service initiative, even if the evaluation is not launched at the same time the services begin. That the most recent school-linked service programs have planned more rigorous evaluation is, therefore, an encouraging trend.

#### **Future Evaluation Efforts**

To develop effective evaluations, planners and providers must increase their commitment to broadening the scope of evaluations while paying more attention to the methodology of analysis.

#### **The Commitment Necessary for Quality Evaluation**

To date, evaluation has not received sufficient attention from advocates for or participants in school-linked service programs. In part, this lack of attention reflects the distrust that can exist between service providers and evaluators. For example, in the preparation of this journal issue, the director of a school-linked service initiative was asked what evaluation had occurred. She replied, "We have been

fortunate not to have been evaluated." She went on to explain that, in her experience, evaluators often did not understand the complex and competing needs of these programs and were not creative enough to design sound evaluations that were responsive to those needs.

This is just one example of the rift that can exist between the planners and providers of school-linked services and those who evaluate them. Providers worry that they will be held accountable for meeting unreasonable expectations regarding long-term, multifaceted problems. They fear that evaluators will focus on a narrow outcome (such as test scores) and overlook a less tangible but equally important change (such as improved cooperation among service providers). Finally, service planners and providers fear that evaluators will get in the way by imposing, for the sake of the evaluation, such rigidity that the main goal of serving children in a flexible and effective way will be thwarted.

These fears are real, and not without some justification. But they are too often not challenged or explored. The long history of program evaluation indicates that evaluations of school-linked service efforts can be structured to provide useful, objective information while remaining sensitive to the needs of the programs they examine.

To generate the most useful and reliable information, evaluators must adhere to more rigorous standards; settling for less will delay the discovery of how best to deliver services to children and families.

Achieving higher-quality evaluation will require a significant commitment from policymakers and funders. Top-quality evaluation requires substantial amounts of both money and time. Evidence from the past indicates that the best outcome evaluations of multisite programs require state or federal funding to allow the necessary depth, breadth, and duration (for example, funding from the National Cancer Institute for anti-smoking initiatives or from the National Heart Lung and Blood Institute for programs to prevent cardiovascular disease).

Top-quality evaluation requires years. In many instances, complex outcome evaluations should not even be undertaken until a year or two after the school-linked service effort begins. Funders must understand that this time is necessary to

work out the problems that occur in implementing any new program. Although the eventual goal may be to institute an outcome evaluation, for the first couple of years the program is in operation developers should focus on creating the best program possible—not trying to measure definitively whether the new program is having an effect on the students or families it serves. During this time, however, process evaluation is necessary (that is, documenting who receives which services) and plans for the eventual outcome evaluation should be completed. This period can also be a time for pilot-testing the methods and measures to be used in the outcome evaluation.

In planning an outcome evaluation, funders should be sensitive to program staff's concerns about being measured with an unfair yardstick. In selecting outcome measures, evaluators should try to translate the impressions of program staff (from line workers to administrators) and program participants about what changes the program is creating. Undoubtedly, there will be some outcomes that policymakers believe must be measured in most programs, such as school dropout rates, teen pregnancy, licit and illicit substance abuse, and school attendance; changes in system procedures; and costs. There may be other outcomes, however, that only program staff and participants know should be measured—changes in the family or the community, for example. As Bruner writes: "judgments of effectiveness should be comprehensive and interdisciplinary, rather than narrowly defined or single-agency focus."<sup>43</sup> Program staff and participants usually have very good insights about making such judgments.

Although objective, behavioral outcomes are an indispensable priority in a program evaluation, it is important to acknowledge that softer measures, such as participant or staff satisfaction with a program or a new system, are also meaningful. For example, when no other measures show improvement, but costs are equivalent and students, families, and the staff feel happier as a result of a new program, the program may well be judged a success.

Finally, funders must understand that outcome evaluations should continue over a number of years. In appraising programs that try to intervene early in children's lives to prevent later problems, taking the long view is essential. By definition, such pro-

grams seek to forestall the development of problem behaviors or conditions that either do not yet exist or only exist at such low levels that traditional statistical tests will not be able to demonstrate significant differences between experimental and control groups.

For example, consider a school-linked service initiative for children in elementary school that seeks to prevent future drug use. At baseline, very few children in either the experimental or the control group are likely to be drug users. Even 1 year after initiation of the program, the number of children in either group likely to be experimenting with drugs is too low to be able to demonstrate statistically significant differences between the two groups. Upon the basis of those results, it would be foolish to conclude, however, that the apparent equivalence between groups means that the program is ineffective. Instead, one must follow the children in both groups for at least 2 or 3 years. At that point, enough children in both groups will have begun to experiment with drugs that group differences in the rate at which that experimentation occurs will be discernable through traditional statistical tests.

#### Methodological Issues in Outcome Evaluation

Effective evaluations of school-linked services require clear definitions of program goals; evaluation plans that reflect those goals; selection of objective outcome measures; and, when possible, research designs to demonstrate causation.

In addition, such evaluations require careful consideration and resolution of very difficult methodological issues. For example, how will the evaluation address the inevitable movement of students in and out of participating school districts? Some of the students and families who begin a program will not complete it or be available for follow-up afterward. Suppose a program that sought to prevent smoking assessed smoking only in those who remained in school, even though those who dropped out were more likely to be smokers. The results of the program would appear more positive than they actually were. The literature on single-focus, school-linked service programs can provide useful suggestions about how to deal with attrition from a study.<sup>36,44,45</sup>

Methodologically sound studies must also be attentive to commonly accepted rules of statistical analysis. Such rules gov-

ern how many children, classrooms, or schools must be involved to detect the level of change in outcome expected by program developers as a result of intervention. In addition, rules for statistical tests govern how large an effect must be present before the outcome can be labeled as due to the intervention rather than to chance. Sound evaluations must also make sure that the unit of analysis (school, classroom, or individual student) is appropriate.<sup>2,36,46,47</sup>

## Conclusions

Policymakers, funders, evaluators, and service providers should not accept statements concerning the benefits of a particular program without some evidence provided by evaluation, and they should not pass up opportunities to learn more from their new projects. Even if one question seems to have been settled, other questions usually remain. Suppose, for example, that the evaluation of Project X indicated that the project "worked." Did it work better than some other alternative? Would the program work for other groups of students and

families? Would it work in other schools? What aspects of the project were most effective or important? What were the costs of the program?

Evaluation never has and perhaps never will be the ultimate arbiter of whether or not a program is funded. Successful programs die because of budget deficits or changes in public values. Unevaluated programs or evaluated programs with inconclusive or even negative results continue year after year because of politics or entrenched constituencies. Other values can override any guidance provided by evaluation. Nevertheless, evaluation is too important a tool to overlook. "Although policy decisions must ultimately rest on value preferences, evidence about the costs and benefits of alternative policies are critical in making value choices."<sup>48</sup> Expensive, unevaluated programs that are continued year after year and that are based only on hunches or political winds can represent a waste of millions of dollars as well as lost opportunities to try what could well be more effective approaches.

1. Cronbach, L., Ambron, S., Dornbusch, S., et al. *Toward reform of program evaluation*. San Francisco: Jossey-Bass, 1980, pp. 12-13.
2. Gomby, D. Basics of evaluation. Unpublished manuscript. Available upon request from the Center for the Future of Children, The David and Lucile Packard Foundation, Los Altos, CA. 1992.
3. Pawl, J., Barnard, K., Korner, A., et al. *Charting change in infants, families and services: A guide to program evaluation for administrators and practitioners*. Washington, DC: National Center for Clinical Infant Programs, 1987, p. 8.
4. Weiss, H.B., Jacobs, F.H. *Evaluating family programs*. Hawthorne, NY: Aldine De Gruyter, 1988.
5. Gomby, D. School-linked services programs: Lessons from the past. Unpublished manuscript. Available upon request from the Center for the Future of Children, The David and Lucile Packard Foundation, Los Altos, CA. 1992.
6. Forman, S.G., Linney, J.A. School-based prevention of adolescent substance abuse: Programs, implementation and future directions. *School Psychology Review* (1988) 17,4:550-58.
7. Goodstadt, M.S. Alcohol and drug education: Models and outcomes. *Health Education Monographs* (1978) 6,3:263-79.
8. Hofferth, S.L. The effects of programs and policies on adolescent pregnancy and childbearing. In *Risking the Future: Adolescent sexuality, pregnancy, and childbearing*. S.L. Hofferth and C.D. Hayes, eds. Washington, DC: National Academy Press, 1987, pp. 207-63.
9. Moskowitz, J.M. The primary prevention of alcohol problems: A critical review of the research literature. *Journal of Studies on Alcohol* (1989) 50,1:54-88.
10. Glynn, T.J. Essential elements of school-based smoking prevention programs. *Journal of School Health* (1989) 59,5:181-88.
11. Walter, H.J. Primary prevention of chronic disease among children: The school-based "Know Your Body" intervention trials. *Health Education Quarterly* (Summer 1989) 16,2:201-214.

12. Brownell, K.D., Kaye, F.S. A school-based behavior modification, nutrition education, and physical activity program for obese children. *The American Journal of Clinical Nutrition* (1982) 35:277-83.
13. Pentz, M.A., Dwyer, J.H., MacKinnon, D.P., et al. A multicomunity trial for primary prevention of adolescent drug abuse. *Journal of the American Medical Association* (1989) 281,22:3259-66.
14. Vincent, M.L., Clearie, A.F., Schluchter, M.D. Reducing adolescent pregnancy through school and community-based education. *Journal of the American Medical Association* (1987) 257,24:3382-86.
15. Felner, R.D., Gintner, M., Primavera, J. Primary prevention during school transitions: Social support and environmental structure. *American Journal of Community Psychology* (1982) 10,3:277-90.
16. Felner, R.D., Adan, A.M. The school transitional environment project: An ecological intervention and evaluation. In *14 ounces of prevention*. R.H. Price, E.L. Cowen, R.P. Lorion, J. Ramos-McKay, eds. Washington, DC: American Psychological Association, 1988, pp. 111-22.
17. Felner, R.D., Felner, T.Y. Primary prevention programs in the educational context: A transactional-ecological framework and analysis. In *Primary prevention and promotion in the schools*. L.A. Bond and B.E. Compas, eds. Newbury Park, CA: Sage Publications, 1989.
18. Schaps, E., DiBartolo, R., Moskowitz, J.M., et al. A review of 127 drug abuse prevention program evaluations. *Journal of Drug Issues* (Winter 1981): 17-43.
19. Kirby, D. Sexuality education: A more realistic view of its effects. *Journal of School Health* (1985) 55,10:421-24.
20. Bartlett, E.D. The contribution of school health education to community health promotion: What can we reasonably expect? *American Journal of Public Health* (1981) 71,12:1384-91.
21. Rundall, T.G., Bruvold, W.H. A meta-analysis of school-based smoking and alcohol use prevention programs. *Health Education Quarterly* (Fall 1988) 15,3:317-34.
22. Kirby, D., Waszak, C.S., Ziegler, J. *An assessment of six school-based clinics: Services, impact and potential*. Houston, TX: Center for Population Options, October 1989.
23. Kirby, D., Waszak, C., Ziegler, J. Six school-based clinics: Their reproductive health services and impact on sexual behavior. *Family Planning Perspectives* (1991) 23,1:6-16.
24. Conte, J.R., Rosen, C., Saperstein, L., Shermack, R. An evaluation of a program to prevent the sexual victimization of young children. *Child Abuse & Neglect* (1985) 9:319-28.
25. Taggart, V.S., Bush, P.J., Zuckerman, A.E., Theiss, P.K. A process evaluation of the District of Columbia "Know Your Body" project. *Journal of School Health* (1990) 60,2:60-66.
26. Goodstadt, M.S. School-based drug education in North America: What is wrong? What can be done? *Journal of School Health* (1986) 56,7:278-81.
27. Cohen, R.Y., Felix, M.R.J., Brownell, K.D. The role of parents and older peers in school-based cardiovascular prevention programs: Implications for program development. *Health Education Quarterly* (Summer 1989) 16,2:245-53.
28. Ellickson, P.L., Bell, R.M. Drug prevention in junior high: A multi-site longitudinal test. *Science* (1990) 247:1299-1305.
29. Ellickson, P.L., Bell, R.M. *Prospects for preventing drug use among young adolescents*. Santa Monica, CA: The Rand Corporation (R-3896-CHF), March 1990.
30. Tobler, N.S. Meta-analysis of 143 adolescent drug prevention programs: Quantitative outcome results of program participants compared to a control or comparison group. *Journal of Drug Issues* (1986) 16,4:537-67.
31. Coates, T.J., Jeffrey, R.W., Slinkard, L.A. Heart healthy eating and exercise: Introducing and maintaining changes in health behaviors. *American Journal of Public Health* (1981) 71,1:15-23.
32. Resnicow, K., Orlandi, M.A., Vaccaro, D., Wynder, E. Implementation of a pilot school-site cholesterol reduction intervention. *Journal of School Health* (1989) 59,2:74-78.
33. Killen, J.D., Robinson, T.N., Telch, M.J., et al. The Stanford Adolescent Heart Health Program. *Health Education Quarterly* (Summer 1989) 16,2:263-83.
34. Luepker, R.V., Johnson, C.A., Murray, D.M., Pechacek, T.F. Prevention of cigarette smoking: Three-year follow-up of an education program for youth. *Journal of Behavioral Medicine* (1983) 6,1:53-62.

35. Flay, B.R. Psychosocial approaches to smoking prevention: A review of findings. *Health Psychology* (1985) 4,5:449-88.
36. McKinlay, S.J., Stone, E.J., Zucker, D.M. Research design and analysis issues. *Health Education Quarterly* (Summer 1989) 16,2:307-13.
37. Best, J.A., Thomson, S.J., Santi, S.M., et al. Preventing cigarette smoking among school children. *Annual Review of Public Health* (1988) 9:161-201.
38. Hansen, W.B., Malotte, C.K., Fielding, J.E. Evaluation of a tobacco and alcohol abuse prevention curriculum for adolescents. *Health Education Quarterly* (Spring 1988) 15,1: 93-114.
39. Overholzer, J.C., Hemstreet, A.H., Spirito, A., Vyse, S. Suicide awareness programs in the schools: Effects of gender and personal experience. *Journal of the American Academy of Child Adolescent Psychiatry* (1989) 28,6:925-30.
40. Shaffer, D., Vieland, V., Garland, A., et al. Adolescent suicide attempters. *Journal of the American Medical Association* (1990) 284,24:3151-55.
41. Murray, D.M., Pirie, P., Luepker, R.V., Pallonen, U. Five- and six-year follow-up results from four seventh-grade smoking prevention strategies. *Journal of Behavioral Medicine* (1989) 12,2:207-18.
42. Flay, B.R., Koepke, D., Thomson, S.J., et al. Six-year follow-up of the first Waterloo school smoking prevention trial. *American Journal of Public Health* (1989) 79,10:1371-76.
43. Bruner, C. *Thinking collaboratively: Ten questions and answers to help policy makers improve children's services*. Washington, DC: Education and Human Services Consortium, 1991.
44. Biglan, A., Severson, H., Ary, D., et al. Do smoking prevention programs really work? Attrition and the internal and external validity of an evaluation of a refusal skills training program. *Journal of Behavioral Medicine* (1987) 10,2:159-71.
45. Hansen, W.B., Collins, L.M., Malotte, C.K., et al. Attrition in prevention research. *Journal of Behavioral Medicine* (1985) 8,3:261-75.
46. Linney, J.A. Optimizing research strategies in the schools. In *Primary prevention and promotion in the schools*. L.A. Bond and B.E. Compas, eds. Newbury Park, CA, Sage Publications, 1989, pp. 50-76.
47. Murray, D.M., Hannan, P.J., Zucker, D.M. Analysis issues in school-based health promotion studies. *Health Education Quarterly* (Summer 1989) 16,2:315-20.
48. Wald, M.S., Carlsmith, J.M., Leiderman, P.H. *Protecting abused and neglected children*. Stanford, CA: Stanford University Press, 1988, p. 181.