PROCESS EVALUATION FOR PUBLIC HEALTH INTERVENTIONS AND RESEARCH

Allan Steckler

Laura Linnan

Editors

Foreword by Barbara A. Israel

CHAPTER FOUR

Process Evaluation of the Adolescent Social Action Program in New Mexico

Deborah L. Helitzer and Soo-Jin Yoon

This chapter describes the process evaluation of a primary prevention program, the Adolescent Social Action Program (ASAP), which was aimed at reducing alcohol- and drug-related morbidity and mortality rates among New Mexico's youth. The program began in 1982 with supervised youth visits to the Emergency Room of the University of New Mexico (UNM) Hospital. During these visits, young people interviewed patients who had problems related to alcohol and drug use. Over the years since its inception, ASAP was successful in gaining various funding for its implementation and improvement. A curriculum was written around the core experience of the patient interviews, and, over time, ASAP broadened its focus to include tobacco, all types of drugs and substances, interpersonal violence, gangs, and other issues relevant to young people. With its evolution, the program's original name, Alcohol and Substance Abuse Prevention, was changed to Adolescent Social Action Program, to reflect its overall philosophy of empowering young people to become capable of promoting change in their communities.

In 1994, ASAP received a five-year research grant (mid-1994 to mid-1999) from the National Institute on Alcohol Abuse and Alcoholism (NIAAA). Although

previous funding had produced extensive outcome evaluations for the program, the NIAAA grant provided ASAP with its first opportunity to conduct a systematic process evaluation. The process evaluation was designed at the beginning of the NIAAA grant period, along with the development of the outcome evaluation instruments, curriculum revision, and planning for program implementation. Because the quasi-experimental design of the NIAAA-funded outcome evaluation required ASAP to implement its intervention over a period of three and a half years (seven semesters between Spring 1995 and Spring 1998), the process evaluation was designed to capture information during this same period.

Process evaluation has been shown to (1) help determine if Type III error occurred (Steckler and others, 1992)—that is, if there was the belief that changes in program outcomes were due to the program when, in fact, there was an absence of intervention, and (2) enhance the understanding of program effects by linking exposure to outcome (McGraw and others, 1994). The purpose of this process evaluation was to document fidelity to the program (tracking students and documenting participation, attendance, and exposure), thereby enhancing the interpretation of the program's results. This check for *fidelity* (to ensure that the program was implemented as it was intended) was designed not only for the summative purpose of determining if Type III error had occurred but also for the formative purpose of reducing the possibility of such error through the program's constant monitoring of its implementation as well as appropriate activities to correct insufficient implementation. Evidence of program fidelity enhances investigators' ability to attribute changes in outcomes to the program's intervention. For the formative purpose of fine-tuning the program, this process evaluation was designed to provide immediate feedback to program personnel with data that would have to be, in a timely manner, analyzed, interpreted, and given back to the staff members who implemented the program.

The process evaluation design evolved slowly. The original design was based on literature available at the time, and subsequent modifications were made after the onset of the NIAAA grant. For example, focus groups consisting of young people were added to gain information about barriers and facilitators to participation. Observations of program sessions were added to check for the presence of theoretical constructs, which additionally enabled triangulation (Miles and Huberman, 1994) with other data sources. These modifications were based on a prior process evaluation study conducted by Deborah Helitzer for an obesity prevention program for Native American school children, in which twenty-seven sets of data collection instruments were developed, each set having between three and five instruments, from 1994 to 1999 (Helitzer and others, 1999).

A classic distinction in evaluation is whether the evaluator is internal or external to the program. External evaluators are assumed to be more objective,

whereas internal evaluators are assumed to have better insight into program function and meaning (Weiss, 1998). The process evaluation of ASAP was conducted, for the most part, internally, which, in this case, involved the handling of design, instruments, data collection/documentation, and data management, as well as analysis of the process evaluation—functions considered to be part of the role of program implementers, which included two co-principal investigators, a full-time project director, a full-time program manager, and a small team of part-time staff members (site coordinators and research assistants). An external consultant was brought in for the process evaluation after the start of the grant, to increase the objectivity of the evaluation and to attempt to achieve some separation between program implementation and evaluation.

The Adolescent Social Action Program

Program Summary

The Adolescent Social Action Program had two primary long-term goals. The first was to prevent morbidity and mortality from risky behaviors—specifically tobacco and alcohol abuse—among low-income minority youth in high-risk communities. The second was to empower the young people to become leaders who were capable of promoting changes in their communities' behaviors, social values, and environmental policies and norms. The theoretical foundation of the program was based on protection motivation theory (Rogers, 1983) and Freire's empowerment dialogue method (Wallerstein and Bernstein, 1988). Over the course of sixteen years, ASAP operated in over thirty multiethnic (predominantly Hispanic and Native American) communities in New Mexico, served nearly sixteen hundred middle- and high school students, and trained over four hundred adult facilitators. The program was eventually expanded to serve twenty-seven hundred elementary school children through the program's peer education activities with younger students. ASAP attracted national and international attention and received an outstanding program award from the U.S. Department of Education's Safe and Drug-Free Schools Program for two consecutive years.

Participating Schools

In 1994, upon receipt of NIAAA funding, eight Albuquerque middle schools were invited to participate in the program and were informed that each of them would be randomly allocated to either the intervention or the comparison condition. Hence, four schools were randomly selected to receive the intervention during

the research study period, whereas the other four schools received the delayed intervention (comparison condition). Before the onset of the research activities, ASAP gained Institutional Review Board approval from the Human Research Review Committee at UNM.

Youth Recruitment

Site coordinators, who were part-time paid personnel, had the responsibility of recruiting young people to participate in ASAP. The ASAP program manager and site coordinators consulted with the school administration to identify at each of the eight schools a seventh-grade teacher who facilitated access to his or her students and was willing to help in the recruitment process. These teachers were provided with a nominal stipend for their time spent on such tasks as arranging meeting spaces at the school for ASAP whenever needed, collecting consent forms, and reminding students of upcoming ASAP sessions. ASAP site coordinators entered the teachers' classrooms to recruit students, and ASAP research assistants accessed students for data collection (using questionnaires, forms, and saliva samples [for a cotinine assessment]). Those students who volunteered to participate were enrolled on a first-come, first-served basis. Between eighteen and twenty-one students were recruited each semester at each school. Site coordinators were responsible for orienting parents of the students and for coordinating transportation of the students to the program sites (either UNM hospital or Bernalillo County Detention Center).

ASAP Facilitators

An important part of the ASAP program was the group of adult facilitators who worked with the young people during the six-week hospital-detention center sessions (a more detailed explanation of the centers is given later in this chapter). Most facilitators were recruited through UNM classes and received university credit for their involvement, and some facilitators were community members who volunteered. (See the Results section for specific demographic information on facilitators.) All facilitators were required to attend a two-day training course at the beginning of the semester. During the semester, facilitators were required to attend a weekly seminar session, which included class readings of current literature on health issues related to young people and discussions about the progress of the hospital-detention center sessions. These sessions allowed the opportunity for facilitators to learn from each other's experiences and share lessons learned about how to keep the sessions on track.

The Core Curriculum: Hospital-Detention Center Sessions

Students who chose to participate in ASAP were assigned to a small group by school. ASAP attempted to recruit seven to nine students for each group. Due to attrition and the need to make up the numbers (for the appropriate sample size for the outcome evaluation), the number of students per group actually ranged from two to eleven. Each group of students was assigned two facilitators, who worked together to implement the core curriculum sessions held once a week for each group for six consecutive weeks.

The UNM hospital could handle only a moderate number of students without disrupting its normal operations, so the groups were staggered by day and by waves. In other words, the length of each semester allowed ASAP to schedule two waves of the six-week curriculum, and each group held a session one night a week (on a Monday, Tuesday, Wednesday, or Thursday). Hence, during a semester, between five and twelve groups of students went through the curriculum.

The first session of the six-week curriculum was a group-building session, followed by three supervised visits to the University-affiliated hospital and one visit to the county detention center. On these site visits, the students interviewed patients, their families, and jail residents who were affected by alcohol or substance abuse problems. After interviews, students and facilitators were debriefed on their experiences, using the SHOWED model (Helitzer, Yoon, Wallerstein, and Garcia-Velarde, 2000) to generate critical thinking and dialogue. This model was derived from ASAP's theoretical foundation. The acronym SHOWED stands for S—What did the students *see* and observe? H—What is *happening* in their stories? O—How do the stories relate to *our* lives? W—Why is this a problem? E—What would *empower* this person or us to change? and D—What can we *do* to improve our lives or the lives of others? A variety of other participatory learning exercises, designed to trigger topics relevant to young people, augmented each session. For example, for a media literacy activity, students constructed a mosaic of cigarette and alcohol advertisements from magazines and they discussed how advertisers target young people. The sixth session was a potluck dinner with family members, and it provided the opportunity for students to brainstorm about community projects in which they could further engage after the curriculum sessions.

Social Action Projects and Booster Sessions

After the core hospital-detention center experience, the site coordinators worked with the youth groups on social action projects. In alignment with the philosophy of the program of empowering young people, the student groups were encouraged

to choose their own projects. These projects ranged from no project at all to establishing ASAP clubs at schools to painting murals on a school wall to creating educational music videos.

In addition, a booster session was conducted by ASAP site coordinators six months after the students received the core curriculum. This booster session consisted of a visit to a courtroom, where students witnessed live hearings and interviewed judges and lawyers to gain insight into the legal and financial ramifications of risky behaviors, such as drinking and driving drunk. As in the first six sessions, the booster included a thirty-minute debriefing that used the SHOWED model.

Outcome Evaluation

For the outcome evaluation, a questionnaire was developed to assess whether the program influenced students' attainment of constructs pertinent to protection motivation theory. The questionnaire appraised (1) *threat*—students' perception of their susceptibility and vulnerability to risky behaviors, (2) *coping*—students' perception of their self-efficacies and intentions to engage in self- or social protective behaviors, and (3) *empathy*—students' social relationships and their ability to experience vicariously the feelings and thoughts of others. Figure 4.1 illustrates the ASAP program theory model, which is explained in more detail in the Theoretical Framework section of this chapter (Wallerstein and Sanchez-Merki, 1994).

The outcome evaluation questionnaire was administered to each student in both the intervention and the comparison groups at pretest (immediately before they received the core curriculum), immediately posttest, and six weeks, eight months, and fifteen months posttest. In addition, a direct measure of tobacco use was collected in both groups through saliva samples at pretest and fifteen months posttest. During the course of seven school semesters, sixty-three groups of students participated in the project. Among the students who provided basic demographic information, 547 (63.2 percent) were female and 318 (36.8 percent) were male. Sixty four percent of the students self-identified as "Hispanic" (n = 557), 23.6 percent as "White/Anglo" (n = 204), 8.3 percent as "American Indian" (n = 72), 6.6 percent as "Black" (n = 57), 2.1 percent as "Asian American" (n = 18) and 7.2 percent as "Other" (n = 62). (These percentages add up to over 100 percent because nearly 10 percent of the students self-identified as belonging to more than one group. For example, forty-six students indicated that they were both Hispanic and white/Anglo). Follow-up rates at fifteen months were between 76 and 86 percent for both intervention and comparison students.

FIGURE 4.1. ASAP PROGRAM THEORY MODEL.

*Protection Motivation Theory
Motivation comes from cues to act
Threat Appraisal: severity/susceptibility
Coping Appraisal: self and response efficiency
Protective Behaviors

+Freirian Theory
Listening
Dialogue
Action
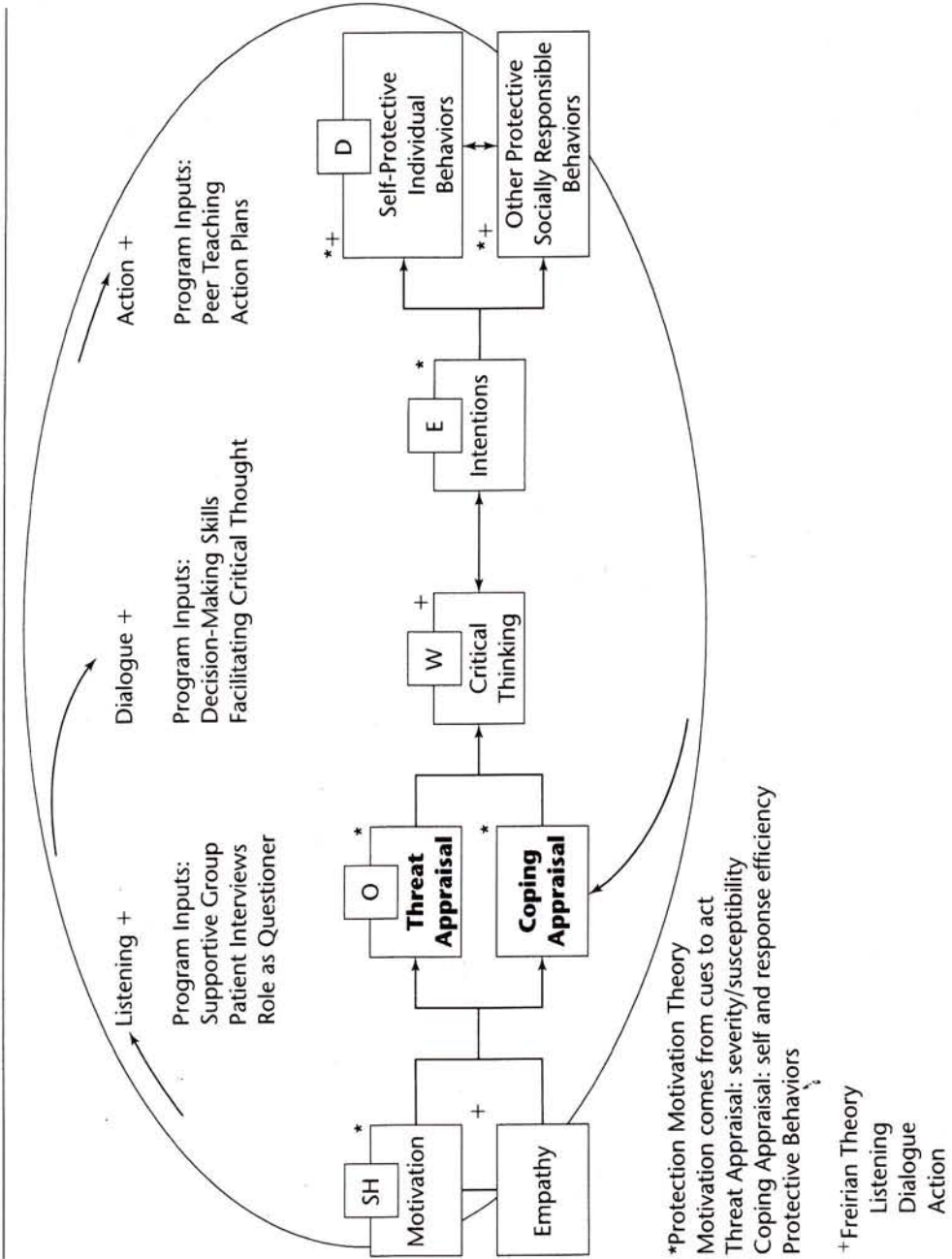
Source: Wallerstein and Sanchez-Merki, 1994.

Theoretical Framework

The hypothesis of the ASAP program was that the experiential aspect of the program (for example, patient and inmate interviews), in combination with the facilitator-led dialogical method, would lead students to empathize with the patients and jail residents, in hearing their stories. Students were encouraged to talk about whether or how they identified with the issues raised in interviews and to critically analyze the influences that contributed to risky behaviors. For example, the specific targeting of young people by the liquor and tobacco industries is responsible for increasing levels of youth consumption of alcohol and cigarettes, and young people, being engaged in such risky behaviors, become negative role models for their friends and family members. The expectation was that the ability to analyze their own experiences would empower these young ASAP participants to take action.

The ASAP program focused on reducing risky behaviors and encouraging socially responsible behaviors that would protect others. The curriculum was based on two complementary theoretical perspectives that were woven together in the project curriculum: Freire's empowerment theory (Wallerstein and Bernstein, 1988; Wallerstein and Sanchez-Merki, 1994) and Rogers's protection motivation theory (Floyd, Prentice-Dunn, and Rogers, 2000; Rogers, 1983). Freire's approach was used in designing the structured dialogue component of the ASAP intervention. The theories used in ASAP were hypothesized to provide psychological empowerment, including personal efficacy, recognition of the need for group participation, critical consciousness, and the willingness to participate in collective action (Rissel and others, 1996). Figure 4.1 shows how the two theories and the SHOWED model were expected to move participants to empowerment.

The first theoretical framework was the Freirian dialogue method, which was used to foster critical thinking about adolescents' perception of alcohol, drug, or tobacco consumption and the meaning of such behavior to adolescents trying to achieve or maintain a certain self-image. For example, a teenage girl feels that smoking makes her look more grown up. The Freirian empowerment dialogue method is based on a continuous listening-dialogue-action cycle, in which program participants identify cues to action by listening to their own issues of emotional and social import, engaging in dialogues about these issues, and developing strategies for addressing them. The SHOWED model was derived from this theoretical foundation. This critical thinking process has been shown to influence youth engagement in such risky behaviors as drinking, smoking, and unexcused absence from school (Chassin, Presson, Sherman, and Steinberg, 1989; Jacobson, Atkins, and Hacker, 1983).

Rogers's protection motivation theory was a second theoretical foundation of the ASAP curriculum. Protection motivation theory hypothesizes that the

decision to act is initiated through a range of informational sources and is mediated through a nonlinear cognitive perceptual process (Rogers, 1983; Rogers, Deckner, and Mewborn, 1978). A health-seeking response is expected when a health threat increases one's vulnerability and susceptibility, when self-efficacy and *response efficacy* (the belief that one's actions can have an effect and that one can respond correctly to a health threat) increase, and when the rewards for engaging in a maladaptive behavior decline (Rippetoe and Rogers, 1987; Rogers, 1983). The jail and hospital visits were expected to heighten the students' threat appraisal (or risk perception) of the seriousness of their own susceptibility to the consequences of substance abuse. The dialogue with patients and jail residents was expected to encourage the students to develop the ability to think through a situation and make an assessment that would lead to an appropriate behavioral response, as well as enhance their empathy, knowledge, and intention to change their own behaviors. It was anticipated that their coping skills and *self-efficacy* (personal responsibility) would be improved by their participation in peer resistance and decision-making exercises. For example, students were asked to enact scripted and improvised role-playing exercises in which they responded to common peer pressure situations.

Overall Program Results

Prior outcome evaluations demonstrated that the curriculum increased the students' social skills, competence, critical consciousness, knowledge, and self-efficacy for their own behaviors, as well as their social responsibility and prosocial behaviors, including their self-efficacy related to helping others, their recognition of the need for group participation, and their willingness to participate in collective action (Rissel and others, 1996).

However, the outcome results of the NIAAA-funded study of ASAP showed no significant differences in alcohol, tobacco, or other substance use among the young people—between the intervention groups and the comparison groups (Wallerstein and Woodall, 2000). In addition, no differences in other measures of threat appraisal or coping appraisal were found. The major pattern observed in the data was the regular increase over time in the proportion of students reporting ever having had an alcoholic drink. This percentage increased from 61.2 percent pretreatment to 64.0 percent posttreatment, to 66.3 percent at six months posttreatment, to 74.4 percent at the fifteen-month follow-up. In fact, compared with the outcomes of other published studies, the ASAP study participants reflected higher levels of alcohol initiation and exposure—higher than the national average for this age group. This was not surprising, given that the risk level of the study population was higher than the national average. All ethnic groups and both

genders, in both the intervention and the control groups, showed, over the course of the study, an increased tendency to drink alcohol. The rate for females increased from 58.6 percent pretreatment to 75.7 percent at the fifteen-month follow-up.

Program Evaluation Design and Methods

The NIAAA-funded process evaluation of the ASAP program was designed to measure the following: (1) the fidelity of the implemented curriculum to the designed curriculum, (2) curriculum implementation consistency across groups of young people, (3) the level of exposure and participation of the students, (4) barriers and facilitators to participation, (5) competing or intervening influences on participation and exposure, (6) the existence of other health-related programs going on concurrently in both intervention and comparison populations, and (7) the characteristics of facilitators and students in the program. Figure 4.2 lists the actual evaluation questions.

The Role of Theory

The critical theoretical elements previously described were included in the design of several process evaluation instruments. Examples of theoretical constructs measured include the SHOWED model, social analysis, critical thinking, group process, facilitation style, cognitive dimensions, the continuous listening-dialogue-action cycle, cognitive appraisals of threat, and coping abilities.

Process Evaluation Design

The design of the process evaluation was an intervention/comparison group design with continuous measurement throughout the training and the seven semesters of intervention group implementation. Most of the data collection occurred

FIGURE 4.2. PROCESS EVALUATION QUESTIONS.

1. Was the curriculum implemented with fidelity to the original design?
2. Was the curriculum implemented similarly in each youth group?
3. What was the level of exposure/participation by students?
4. What were the barriers and facilitators to participation?
5. What were the competing or intervening influences on participation and exposure?
6. What were the characteristics of facilitators and students in the program?

at the intervention sites, as is appropriate for process evaluation. Some of the instruments were designed to contribute to a formative evaluation, some contributed to summative evaluation, and others were useful for both types of evaluation. Many of the instruments were designed to enable triangulation of data. Figure 4.3 shows the implementation model around which the process evaluation was designed.

Process Evaluation Resources

A small proportion of the grant funds were allocated to the process evaluation, from which the consultant was paid for 3 percent of her time for two years. A small percentage (~5 percent) of the salaries of all program staff members covered their data collection activities. In addition, one staff member conducted the process evaluation analyses, devoting to this function ~50 percent of her time for one year.

Data Collection, Sample, and Analyses

Fifteen instruments were used to collect both qualitative and quantitative information for the process evaluation data. Figure 4.3 shows the different types of information gathered and how they were related. The numbers next to each box correspond with the instruments listed in Table 4.1, which presents more details

FIGURE 4.3. IMPLEMENTATION MODEL.
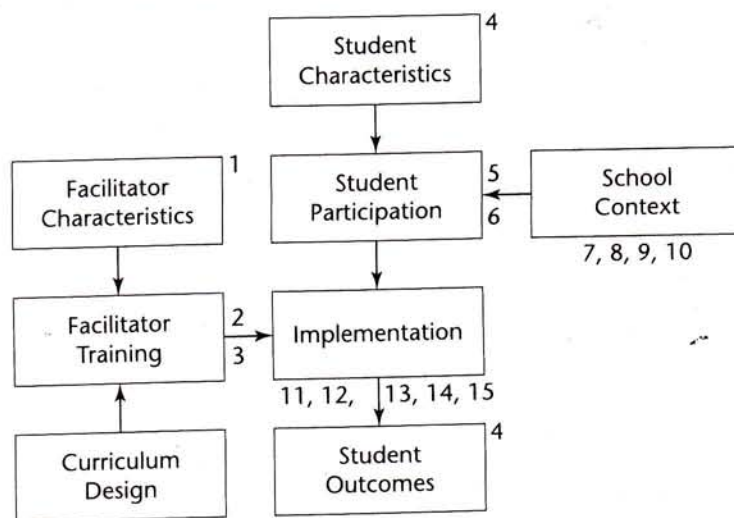
Note: The numbers in this model correspond with the list of instruments in Table 4.1.

TABLE 4.1. PROCESS EVALUATION INSTRUMENTS.

Instrument	Rationale/Purpose	Population/Sample	Analyses Method
1. Facilitator Questionnaire	• To assess facilitator characteristics, experience with drugs & alcohol, work with adolescents, self-efficacy [FORMATIVE & SUMMATIVE]	• Collected at pre- & postcurriculum	• Partially analyzed • Descriptive statistics performed on certain items [QUANTITATIVE & QUALITATIVE]
2. Facilitator Training Evaluation Form	• To assess facilitator perception of training & confidence in ability to implement curriculum [FORMATIVE]	• Forms administered in only 5 of 7 semesters • Total of 46 evaluation forms collected	• Reviewed by training director for immediate feedback • Scores for each item tallied [QUANTITATIVE & QUALITATIVE]
3. Facilitator Attendance Log of Weekly Class Sessions	• To assess participation/ exposure to knowledge and skill enhancement sessions [FORMATIVE & SUMMATIVE]	• Collected weekly each of 7 semesters	• Monitored/tracked attendance of facilitator for the semester [QUANTITATIVE]
4. Student (Outcome) Questionnaire	• To assess student characteristics [SUMMATIVE]	• Collected at pre- & postcurriculum, 8-month & 15-month post [INTERVENTION & COMPARISON GROUPS]	• Descriptive statistics performed on certain items for process evaluation purposes [QUANTITATIVE]
5. 6-Session Curriculum Student Attendance Logs	• To assess participation (exposure) to curriculum • To assess whether there was a dose-response relationship [FORMATIVE & SUMMATIVE]	• One from each team of facilitators covering all 6 sessions • Total of 63 (100% response rate) [INTERVENTION GROUP]	• Attendance scores and percentages calculated by student and by group • Each student & group received a "Program Intensity" (dose) score [QUANTITATIVE]

Instrument	Purpose	Data Collection	Analysis
6. Master Attendance Log	• To assess participation (exposure) to curriculum, booster session, & social action projects • To assess whether there was a dose-response relationship [FORMATIVE & SUMMATIVE]	• One master attendance sheet for the entire program [INTERVENTION GROUP]	• Scores of attendance calculated [QUANTITATIVE]
7. Student Focus Group	• To assess barriers & facilitators to participation [FORMATIVE]	• One focus group conducted with ASAP participants from one school [INTERVENTION GROUP]	• Taped discussions were transcribed and coded for themes; themes were grouped into clusters [QUALITATIVE]
8. School Administrator & Teacher Focus Group	• To assess barriers & facilitators to participation	• School coordinators & administrators from each school [INTERVENTION SCHOOLS]	• Barriers and facilitators reviewed & summarized [QUALITATIVE]
9. Health-Related Program Inventory	• To assess competing programs in both intervention & comparison schools (school context) [SUMMATIVE]	• Collected yearly in January from each school, completed by administrator or teacher [INTERVENTION & COMPARISON SCHOOLS]	Not analyzed [QUANTITATIVE]
10. Health-Related Student Inventory	• To assess ASAP participants' involvement in other programs [SUMMATIVE]	• Collected at precurriculum at the same time as Student Outcome Questionnaire [INTERVENTION & COMPARISON GROUPS]	Not analyzed [QUANTITATIVE]
11. Facilitator Checkoff List	• To assess fidelity of curriculum implementation [FORMATIVE & SUMMATIVE]	• One from each team of facilitators covering all 6 sessions • Total of 63 (100% response rate) [INTERVENTION GROUP]	• Tally of scores by session, by content area, by group, & by semester • Reported in scores, percentages of coverage, & low, medium, or high scores of fidelity [QUANTITATIVE]

TABLE 4.1.   PROCESS EVALUATION INSTRUMENTS. (CONTINUED)

Instrument	Rationale/Purpose	Population/Sample	Analyses Method
12. Observation Checkoff List	• To assess fidelity of curriculum implementation • To assess group process • To assess student attainment of theoretical constructs [FORMATIVE & SUMMATIVE]	• 41 of 63 groups were observed during Session #5, but only 29 (46%) were done with the final version of the instrument [INTERVENTION GROUP]	• Partially analyzed • Fidelity score calculations were analyzed in the same manner as Facilitator Checkoff List & were compared with the scores reported by facilitators for the same sessions [QUANTITATIVE & QUALITATIVE]
13. Student Comment Sheet	• Check on fidelity of curriculum implementation • To assess student attainment of theoretical constructs [FORMATIVE & SUMMATIVE]	• All 6 sessions from all students • Total of 1,497 (91% response rate) [INTERVENTION GROUP]	Data quality was too poor to analyze (Students often did not answer questions or responded in only a few words) [QUALITATIVE]
14. Facilitator Log/Comment Sheet	• Check on fidelity of curriculum implementation • To provide feedback to training director on skill & confidence attainment [FORMATIVE & SUMMATIVE]	• All 6 sessions from all facilitators • Total of 548 (82% response rate) [INTERVENTION GROUP]	• Reviewed by training director for themes & areas to enhance in weekly class sessions [QUALITATIVE]
15. Participant Contact Documentation Sheet	• To document postcurriculum social action projects (extended exposure) [SUMMATIVE]	• Collected from Site Coordinators for each contact with each student [INTERVENTION GROUP]	Not analyzed [QUALITATIVE]

about the instruments—that is, the rationale and purpose of the instruments, the population sample from which the data were collected, and a brief description of the analysis methods used. The first three instruments (1, 2, and 3) were used to collect information on the facilitator characteristics and training, an outcome evaluation instrument (4) provided information on student characteristics, two instruments (5 and 6) examined student participation in and exposure to the program, four instruments (7, 8, 9, and 10) were used to examine school context, and five instruments (11, 12, 13, 14, and 15) measured the implementation of the program. As examples, the facilitator checkoff list and observation forms can be found in Appendixes A and B.

Student comment sheets and facilitator training evaluation forms were anonymous and coded by group or semester. The facilitator questionnaire, health-related student inventory, and student questionnaire did not include names but were coded to keep track of the individuals and to allow datasets to be merged by the code. This list of codes and names was maintained by ASAP research assistants and was not accessed by the outcome evaluation data analysts. The health-related program inventory and focus group were coded by school. Facilitator checkoff lists and log/comment sheets required the facilitators to include their names, as their class grade depended on their turning in these forms, and they were subsequently coded by group for analysis purposes. Student attendance information was initially tracked by student name, which program personnel used to boost attendance through follow-up, but it was eventually coded to be merged with other student-level data. Quantitative data were entered into an Excel spreadsheet (Microsoft Applications, 1995) as individual datasets by instrument and were later merged with the outcome dataset by student, group, or school code. Qualitative data were reviewed and analyzed for content where resources allowed. (Table 4.1 lists which data were collected by quantitative versus qualitative methods.)

Results

The information presented in this section describes only the results and use of data that were completely analyzed. As Table 4.1 shows, some data were found to be unreliable, whereas other data were not analyzed, because of resource constraints. Further discussion of issues related to the limitations of analyses and the use of results follows.

Facilitator Characteristics and Training

Because facilitators were so crucial to the program's success, we developed an extensive analysis of the facilitator characteristics, the training they received, their

attendance at the weekly class sessions, and their ability to implement the program (Helitzer, Yoon, Wallerstein, and Garcia-Velarde, 2000). Facilitators were primarily undergraduate students (81 percent), Caucasian (60 percent), and Hispanic (30 percent), with varied experience working with teenagers (26 percent having less than a year, 33 percent having one to three years, and 13 percent having four years or more). These data were used only to create a profile of the facilitators for progress and final reports.

Data on training effectiveness were triangulated from several sources: facilitator training evaluation, facilitator checkoff list, observation checkoff list, student comment sheet, and facilitator log/comment sheet. Training evaluations revealed that 94 percent of the facilitators were confident that the preprogram two-day training workshop prepared them well for the program. However, the observation data and facilitator self-report data suggest that the facilitators were not competent to implement all of the facilitation methods used in the program and that they were inconsistent in their implementation of role-modeling behaviors. Findings revealed that facilitators showed inconsistent and lower than desirable implementation, especially by the later (fifth and sixth) sessions. When these data suggested that specific content or theoretical constructs were not fully grasped by the facilitators, the training director addressed these issues in the weekly class sessions for the benefit of other facilitators, whose sessions were not yet complete.

Exposure: Attendance and Implementation

Student attendance scores by group for the set of six sessions ranged from 43 to 100 percent. This was determined by calculating the total possible attendance for each group (the number of students in the group multiplied by six—the number of curriculum sessions) and then determining the mean attendance attained by all students in the group. Information on attendance was designed in such a way that program personnel had immediate access to the information for constant monitoring. This enabled program personnel to troubleshoot when attendance was declining and to ensure that efforts were being made to retain students for the six sessions.

Information about implementation of the curriculum was available from the facilitator checkoff list, the student comment sheet, the observation checkoff list, the participant contact documentation sheet, and the facilitator log. The data showed that implementation, as previously noted, was inconsistent. Implementation scores were calculated by session, by activity, and by group, based on the 0- to 3-point scoring system of the facilitator checkoff list (see Appendix A) and by totaling these scores. To check whether certain sessions were less

implemented than others, actual scores from the checkoff list for each session were totaled across all sixty-three groups and checked against the total possible implementation score. There was a decline in implementation scores as sessions progressed over the six-week period, but this pattern was consistently repeated from semester to semester. Repeated elements like the interviews were highly implemented, but other, more difficult and less frequently implemented skills, such as role-playing, were often left out. Content/activity areas that received high implementation scores were those activities that were scripted versus improvised, repeated in several sessions, expected to be discussed in the seminars, and accompanied by visual materials. Poorly covered content areas (in which coverage was 50 percent or less) tended to be activities that were more abstract and more time-consuming, and which required more skill on the part of the facilitators. For example, though the patient/inmate interviews were well implemented, facilitators often failed to follow through with the next, more difficult and abstract, task of using the SHOWED model as a tool for engaging the young people in critical thinking and dialogue (Helitzer, Yoon, Wallerstein, and Garcia-Velarde, 2000).

Finally, a score for program exposure was calculated for each student. Exposure to each session was calculated by multiplying the student group's fidelity score for each session (1 = low, 2 = medium, 3 = high) by the student attendance score for that session (0 = not attended, 1 = attended). The program exposure score for the student was then calculated by totaling the exposure scores of six core sessions. Hence, the formula for program exposure was: Student Program Exposure = SUM[(Session Fidelity) × (attendance)]. The results revealed that there was inconsistent exposure to the curriculum across all of the students. With a possible score of 0 to 18, students' program exposure scores ranged from 2 to 17 ($\bar{x}$ = 10.6, n = 403).

Use of Process Evaluation Results

Some results from the process evaluation were used to inform ASAP about the process of program implementation, whereas other results were not available until after the completion of the implementation. For example, when initial data analyses suggested that the SHOWED model was being inconsistently implemented, quizzes were added to the weekly class sessions for facilitators. A final implementation score for each group (low, medium, or high) was calculated and could have been used for summative evaluation purposes for a dose response analysis, because one of the hypotheses was that there might be a positive dose-response relationship between exposure, participation, implementation, and outcome. Data

on student participation (engagement) were not analyzed. The observation form was continually revised over two semesters in an attempt to improve low interrater reliability. Despite these efforts, when all of the data were analyzed, we found that observers who had previously been facilitators were more critical of student engagement than those who had never facilitated a session. As a result, it was difficult to aggregate data that was not comparable across groups.

Another hypothesis was that student characteristics might have influenced outcomes. For example, students' school performance or previous exposure to alcohol, drugs, or tobacco may have affected their future risk of negative outcomes. Outcome data suggested that the program had no overall effect on alcohol or tobacco consumption, and additional analyses were not conducted. Other information sources, such as the participant contact sheet, facilitator log, and student comment sheet, were not used, because of the lack of resources to embark on the labor-intensive qualitative analyses that these instruments required. Information on the school context (such as that provided by the barriers focus group, the health-related program inventory, and the health-related student inventory) was not used by the program, because the outcome data suggested that the program had not achieved its intended effect.

After the end of the grant period, Helitzer, Yoon, Wallerstein, and Garcia-Velarde (2000) described the relationship between training and implementation. At the time, a new grant submission was pending approval. It was thought that the new grant would be able to take advantage of this information. However, the grant was never awarded, which rendered the process evaluation results less useful than anticipated.

Lessons Learned About Process Evaluation

LESSON 1. *Get more out of less data.*

Our experience is that process evaluation data can easily become unmanageable. If an evaluator is highly detail-oriented, he or she will have the inclination to document every aspect of the project. However, it is unethical and a waste of resources to collect data that will not be used in some way. Unless an evaluator has a plan for data collection, analysis, and use, it is highly likely that he or she will collect more than is needed. It is a valuable exercise to find ways to overlap purposes for tools and for formative and summative applications. The most frequently asked or necessary questions, including both the *whether* and the *why* questions, can be formulated in advance. Early process evaluations, such as those for the CATCH and PATHWAYS projects (McGraw and others, 1994; Helitzer and others, 1999), had hundreds of

instruments, but this is not realistic for most programs. Deciding what are the most important questions to answer will help the process evaluation be more efficient.

LESSON 2. *Use mixed methods and triangulate.*

In addition to being efficient with the collection of data, there is a need for data from different instruments, data sources, samples, and types of data. Qualitative methods, such as in-depth interviews and focus groups, are labor-intensive and perceived as subjective data because of the relationship between the data collector and the data. Furthermore, qualitative data tend to be seen as less valuable than quantitative data. Sometimes, this distinction is worded in terms of "soft" (qualitative) versus "hard" (quantitative) data. For these reasons, qualitative data can be seen as less desirable to include in evaluations than quantitative methods that are seen as more objective and less labor-intensive to collect and analyze (Helitzer and others, 1999). This debate and these biases have existed far too long in both primary and evaluation research. Using both qualitative and quantitative methods, we can better answer both the what and the why questions. Using mixed methods also provides more opportunities for triangulating data.

LESSON 3. *Use the information on time to improve the program.*

Process evaluation staff members should be included in project implementation discussions on a regular basis. In this way, these staff members can learn about the implementation process and can provide frequent feedback for midcourse corrections. This means that program staff people should expect implementation failure as a normal part of project implementation. If it is expected, then they can avoid being defensive about why it is happening, and as a result they can devote time and resources to fine-tuning and making changes based on feedback. We need to think about evaluation's purpose as program improvement rather than as a thumbs-up or thumbs-down judgment as to whether the program is effective. We need to create a culture of learning organizations as they relate to program implementation in public health.

LESSON 4. *Devote adequate attention and resources to process evaluation.*

Although evaluation has more recently been receiving a part of routine program funds, most of these funds and attention are still focused on outcome evaluation. The evidence suggests that most programs do not attain the outcomes they

are designed to attain, and for this reason it is important that more attention be focused on process evaluation. Previously, process evaluation was thought of as "bean counting" because it routinely documented information on, for example, how many trainings were held, as well as how many trainees were trained. Process evaluation has the potential to provide much more information if it is properly planned and if sufficient resources are provided for it.

LESSON 5. *Use process evaluation to answer why, not just whether.*

Program implementation failure is to be expected and planned for. Also, programs under real-life conditions rarely show the same types of results that are published in research studies. This is the difference between program *efficacy* and program *effectiveness* (Green and Lewis, 1986). Because of this, it is more important to look at why programs do not demonstrate the same intensity of intended outcomes. Programs that are designed on paper and implemented in real life face challenges that often are not anticipated or planned for. For example, if teachers implement only parts of a curriculum, we could ask them why they chose the parts they did. However, it would be far more useful to have the original curriculum designers indicate which parts of the curriculum are linked to the theoretical framework of the program.

Some participants may benefit more from an intervention than others will. Process evaluation data can be used to examine whether a dose-response relationship exists between participation and outcome changes. If such a relationship exists, it may explain why some participants show increased changes as compared with others. In addition, other characteristics of the participants, such as age, sex, and ethnicity, can be examined for relationships.

Finally, programs may work better under certain conditions than others. A retrospective analysis can be conducted to examine participants for whom outcomes were favorable to see if individual, family, school, community, or program conditions were different for those participants whose outcomes were not as positive. A strength of process evaluation is its ability to provide more depth of information and the opportunity to build hypotheses for future research.
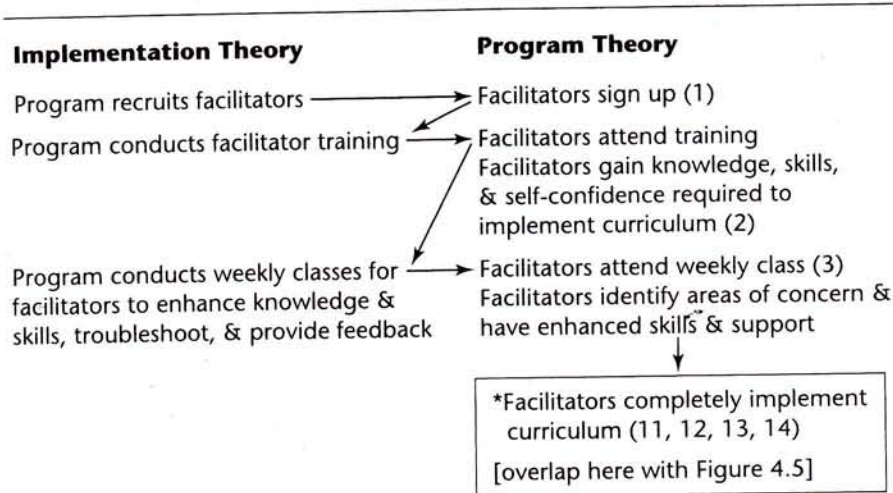
LESSON 6. *Theory, theory, theory.*

Both a strength and a weakness of the process evaluation described in this chapter was its evolving nature, reflecting new knowledge and expertise. The process evaluation of this study lacked a theoretical framework that might have examined in a comprehensive manner the assumptions upon which the program was

based. However, we now know, in 2002, more about process evaluation, which enables us to be self-reflective. We recognize that a potential for process evaluation is the examination of both the implementation and the program theory, as described by Weiss (1998) as the program's "theories of change." Articulating a program's change theories helps us understand the assumptions on which the intervention is based, and it is essential for developing appropriate evaluation questions and evaluation design. Program theories of change help evaluators plan what data to collect and from whom, as well as what types of short- and long-term effects might be expected from the intervention. Using Weiss's theories of change approach, the process evaluation can focus on detailing and systematically documenting each step of the implementation. In this way, the process evaluation can be designed to examine some of the theoretical assumptions of the program.

Figures 4.4 and 4.5 illustrate how the theory of change model could have been used for the ASAP program. We inserted the instrument numbers into the figure to demonstrate the comprehensiveness of the original design in terms of program function but not in terms of theory. Process evaluation instruments were designed to measure *what* but not *why*. They were designed to document the implementation but not to look at the context and reasons behind the variable levels of

FIGURE 4.4. THEORY OF CHANGE MODEL FOR ASAP FACILITATORS.

Implementation Theory   **Program Theory**

Program recruits facilitators ⎯⎯⎯⎯→ Facilitators sign up (1)

Program conducts facilitator training ⎯⎯→ Facilitators attend training
    Facilitators gain knowledge, skills,
    & self-confidence required to
    implement curriculum (2)

Program conducts weekly classes for ⎯⎯→ Facilitators attend weekly class (3)
facilitators to enhance knowledge &   Facilitators identify areas of concern &
skills, troubleshoot, & provide feedback   have enhanced skills & support

> *Facilitators completely implement
>   curriculum (11, 12, 13, 14)
>
> [overlap here with Figure 4.5]

Note: The numbers in this model correspond with the list of instruments in Table 4.1.

FIGURE 4.5. THEORY OF CHANGE MODEL FOR ASAP CURRICULUM IMPLEMENTATION.

Implementation Theory

Program Theory

Program gains cooperation/support from school administration & teachers

School teachers allow time in classrooms for recruitment; schools have low number of competing programs (8, 9, 10)

Site coordinators recruit students from classrooms

Students gain parental permission & sign up

Site coordinators make efforts to boost attendance with reminders, calls, transportation, etc.

Students attend 6-week curriculum sessions (5, 6, 7)

*Facilitators completely implement curriculum (11, 12, 13, 14)

[overlap here with Figure 4.4]

Patients/inmates are available for interviews (11, 12, 13, 14)

Students participate fully (12)

Students determine a social action project at the end of 6-week curriculum

Site coordinators provide support to student group for social action projects (6, 15)

Students attend, participate in, & complete social action project (6, 15)

Site coordinators provide booster session & contacts & invite students at 6-month postcurriculum; they also provide transportation & facilitate session (6, 15)

Students attend & participate in booster session (6, 15)

Students attain/increase & maintain theoretical elements (e.g., self-efficacy & coping skills) for at least 15 months (4)

Students decrease in risky behaviors (e.g., smoking, alcohol, & drug consumption) for at least 15 months (4)

Note: The numbers in this model correspond with the list of instruments in Table 4.1.

implementation. Many of the instruments measured the same component, and triangulation was desirable, but given limited resources, it would have been more effective to have more instruments or sections of instruments to measure a greater proportion of theoretical components.

Using Weiss's theoretical framework (1998) has been helpful to us in our process evaluation efforts. We recently developed a methodology for working closely with investigators and program designers in the early stages of planning and program design. We use implementation and program theory logic models, such as those shown in Figures 4.4 and 4.5, as a framework to

- Articulate the program and implementation theory
- Identify where the theory components can be found in the implementation
- Identify which components are critical to program success
- Articulate what assumptions are held about the relationship between theory and outcome
- Identify which program components require tracking

In this way, we can be more efficient about collecting data. Included in this planning process are schedules for pilot testing and feedback, as well as discussions about timing for feedback, in order to make the feedback most useful for mid-course corrections.

In summary, the lessons learned from this study are important for the future practice of process evaluation. First, evaluators need to gain from investigators and program staff members *support for* and *interest in* the process evaluation. Their support helps ensure the timely and accurate collection of process evaluation measures and it helps ensure that the evaluation continues despite preliminary outcome data findings. Second, evaluators working with staff members and investigators need to build their *capacity for understanding* the value of process evaluation. Such a process might include discussions early on in the evaluation and implementation process to help everyone involved in the project understand the important and interconnected role that all phases of evaluation—formative, process, and outcome—have in a full and insightful understanding of interventions. Evaluations are often broken into the familiar parts of formative, process, and outcome, but for evaluation to be most effective, greater emphasis must be placed on the fact that these parts are an integrated whole evaluation. Such an understanding will go a long way toward resolving the often inadequate resources that are directed toward the design, collection, and analysis of process as well as formative evaluation data. Finally, *planning* is essential in successful evaluations, including the process component, and in our experience it is extremely helpful for program designers and evaluators alike.

Conclusions

The process evaluation described in this chapter represents an ambitious undertaking for the amount of resources it was allotted, and it reflects the state of knowledge and experience about process evaluation that existed in the mid-1990s. It is through these types of projects that we can be self-reflective and take away lessons that can improve the knowledge base from which future process evaluations will be designed.

What follows in this concluding section is a list and discussion of (1) the strengths of this process evaluation and (2) the limitations and opportunities for learning.

Strengths

The existence of a process evaluation. It is significant, given the level of knowledge about process evaluation at the time, that some funding for process evaluation was provided by the NIAAA and that the principal investigators had the wisdom to plan for and allocate these resources to hire a process evaluation consultant. In addition, the design represented a praiseworthy attempt to be conceptually thorough. Program personnel did an excellent and thorough job of collecting and managing the voluminous amount of data. Response rates were extremely high (see Table 4.1) for most instruments. Data were organized and easy to access for analysis.

The multipurpose design of instruments. A second strength of this process evaluation is that instruments were designed to provide both formative and summative data. For example, the attendance logs were intended to be used to provide information on participation for formative purposes, allowing quick response to improve poor attendance by site coordinators. Attendance data also could have been used to calculate the dose response rate for each student, correlating this rate with the alcohol use behavior outcome measures.

Triangulation. A third strength of this process evaluation is the ability to triangulate data from different instruments, data sources, samples, and types of data. Some of the triangulation was accomplished by using both qualitative and quantitative data. Other triangulation was undertaken by using two or more types of informants. For example, the observation checklist was designed to estimate the fidelity of facilitators to the curriculum by an independent observer during one session. Facilitator checklist data and student comment sheets for the same session were compared with observer data to determine whether facilitators were accurately representing their coverage of curriculum components. It was the ability to triangulate the data from the checklists, comment sheets, and observations that enabled facilitators to determine that midcourse corrections were needed.

Limitations and Opportunities for Learning

Insufficient planning for the analysis process. Despite the ambitious design and allocation of some resources to process evaluation, the program did not dedicate sufficient attention to planning for the analysis process so that it would provide useful and timely feedback to the program. This could have been because human resources were mainly focused on the collection and implementation of outcome data. For example, the process evaluation consultant was not brought on board until well into the grant period. When the consultant was brought on, data were already being collected without a plan for the management and analysis of the data. No one anticipated how much data there would be, how long the data analysis would take, or how labor-intensive it would be. Therefore, much of the process evaluation data were not analyzed and synthesized in a time frame that would have enabled the program to make corrective changes to improve the program's implementation. For example, if the data regarding facilitator implementation had been analyzed throughout the project rather than just at the end, the training process could have been improved to increase fidelity to the curriculum (Helitzer, Yoon, Wallerstein, and Garcia-Velarde, 2000).

Insufficient understanding, interest, and appreciation. Although the process evaluation design was thoughtful, there was only partial appreciation for the value of process evaluation and the contribution that its results could make. This lack of knowledge about the various ways the process evaluation could be used resulted in insufficient attention being paid to it during the course of the project. For example, weekly meetings of the project staff often did not include discussions of process evaluation data, and rarely were requests made for data to be analyzed for a specific purpose. Also, the process evaluation expert was paid only 3 percent of her salary to work on the project, which provided insufficient time for the thoughtful incorporation of process evaluation results.

Process evaluation data were not being used to answer the question Why? When the outcome data suggested that there were no intervention effects, process evaluation data could have been used to understand whether or *why* there was program failure and whether or *why* the theoretical assumptions about the curriculum were not sound. For example, once we realized that the facilitators were implementing some parts of the curriculum more consistently than other parts, we could have used the qualitative data from facilitator logs to try to understand why the facilitators seemed to have more difficulty with components requiring role-playing behavior than with those involving the introduction and review of content areas.

Limited use of qualitative data. Most of the qualitative data were not analyzed. These data were a rich source of information (for example, participant contact

documentation sheets were the logs of site coordinators working on social action projects) and could have provided insight into unintended outcomes. However, the reality of qualitative data is that it can be unwieldy and that it requires a long time to analyze properly. Qualitative analysis requires not only adequate time and human resources devoted to it but also personnel who possess the appropriate research skills. In the case of ASAP, the human resources devoted to research were not adequate to be extended to most of the qualitative analyses.

References

Chassin, L., Presson, C. C., Sherman, S. J., and Steinberg, L. "Adolescent Smokeless Tobacco Use: Future Research Needs." *NCI Monograph*, 1989, *8*, 101–105.

Floyd, D. L., Prentice-Dunn, S., and Rogers, R. W. "A Meta-Analysis of Research on Protection Motivation Theory." *Journal of Applied Social Psychology*, 2000, *30*(2), 407–429.

Green, L. W., and Lewis, F. M. *Measurement and Evaluation in Health Education and Health Promotion.* Mountain View, Calif.: Mayfield, 1986.

Helitzer, D. L., Yoon, S., Wallerstein, N., and Garcia-Velarde, L. "The Role of Process Evaluation in the Training of Facilitators for an Adolescent Health Education Program." *Journal of School Health*, 2000, *7*(4), 141–147.

Helitzer, D. L., and others. "Process Evaluation in a Multisite, Primary Obesity-Prevention Trial in American Indian Schoolchildren." *American Journal of Clinical Nutrition*, 1999, *69* (supp.), 816S–824S.

Jacobson, M., Atkins, R., and Hacker, G. *The Booze Merchants: The Inebriating of America.* Washington, D.C.: Center for Science in the Public Interest, 1983.

McGraw, S. A., and others. "Design of Process Evaluation Within the Child and Adolescent Trial for Cardiovascular Health (CATCH)." *Health Education Quarterly*, 1994, *2* (supp.), S5–26.

Microsoft Applications, *Microsoft Excel*, 1995. Software.

Miles, M. B., and Huberman, A. M. *Qualitative Data Analysis: An Expanded Sourcebook.* (2nd ed.). Thousand Oaks, Calif.: Sage, 1994.

Rippetoe, P. A., and Rogers, R. W. "Effects of Components of Protection-Motivation Theory on Adaptive and Maladaptive Coping with a Health Threat." *Journal of Personality and Social Psychology*, 1987, *52*(3), 596–604.

Rissel, C. E., and others. "Empowerment, Alcohol, Eighth-Grade Students and Health Promotion." *Journal of Alcohol and Drug Education*, 1996, *41*(2), 105–119.

Rogers, R. W. "Cognitive and Physiological Processes in Attitudinal Change: A Revised Theory of Protection-Motivation." In J. Cacioppo and R. Petty (eds.), *Social Psychophysiology.* New York: Guilford Press, 1983.

Rogers, R. W. "Changing Health-Related Attitudes and Behavior: The Role of Preventive Health Psychology." In R. McGlyn, J. Maddox, C. Stoltenbery, and R. J. Harvey (eds.), *Interfaces in Psychology.* Lubbock: Texas Tech University Press, 1994.

Rogers, R. W., Deckner, C. W., and Mewborn, C. R. "An Expectancy-Value Theory Approach to the Long-Term Modification of Smoking Behavior." *Journal of Clinical Psychology*, 1978, *34*, 562–566.

Steckler, A., and others. "Toward Integrating Qualitative and Quantitative Methods: An Introduction." *Health Education Quarterly,* 1992, *19,* 1–9.

Wallerstein, N., and Bernstein, E. "Empowerment Education: Freire's Ideas Adapted to Health Education." *Health Education Quarterly,* 1988, *15*(4), 379–394.

Wallerstein, N., and Sanchez-Merki, V. "Freirian Praxis in Health Education: Research Results from an Adolescent Prevention Program." *Health Education Research,* 1994, *9*(1), 105–118.

Wallerstein, N., and Woodall, G. *ASAP Project Final Report for NIAAA.* Albuquerque: University of New Mexico, 2000.

Weiss, C. H. *Evaluation.* Englewood Cliffs, N.J.: Prentice Hall, 1998.