



**MONTCLAIR STATE**  
UNIVERSITY

Montclair State University  
**Montclair State University Digital  
Commons**

---

Department of Public Health Scholarship and  
Creative Works

Department of Public Health

---

Fall 10-14-2012

## Lessons From the Trenches: Meeting Evaluation Challenges in School Health Education

Michael Young

*New Mexico State University, myoung@nmsu.edu*

George Denny

*University of Arkansas, gdenny@uark.edu*

Joseph Donnelly

*Montclair State University, donnellyj@montclair.edu*

Follow this and additional works at: <https://digitalcommons.montclair.edu/public-health-facpubs>



Part of the [Clinical Epidemiology Commons](#), [Community Health and Preventive Medicine Commons](#), [Environmental Public Health Commons](#), [Epidemiology Commons](#), [Health and Medical Administration Commons](#), [Health and Medical Physics Commons](#), [Health Services Administration Commons](#), [Health Services Research Commons](#), [International Public Health Commons](#), [Medical Education Commons](#), [Mental and Social Health Commons](#), [Occupational Health and Industrial Hygiene Commons](#), [Other Public Health Commons](#), [Patient Safety Commons](#), and the [Public Health Education and Promotion Commons](#)

---

### MSU Digital Commons Citation

Young, Michael; Denny, George; and Donnelly, Joseph, "Lessons From the Trenches: Meeting Evaluation Challenges in School Health Education" (2012). *Department of Public Health Scholarship and Creative Works*. 163.

<https://digitalcommons.montclair.edu/public-health-facpubs/163>

This Article is brought to you for free and open access by the Department of Public Health at Montclair State University Digital Commons. It has been accepted for inclusion in Department of Public Health Scholarship and Creative Works by an authorized administrator of Montclair State University Digital Commons. For more information, please contact [digitalcommons@montclair.edu](mailto:digitalcommons@montclair.edu).

## GENERAL ARTICLE

# Lessons From the Trenches: Meeting Evaluation Challenges in School Health Education\*

MICHAEL YOUNG, PhD<sup>a</sup> GEORGE DENNY, PhD<sup>b</sup> JOSEPH DONNELLY, PhD<sup>c</sup>

---

**ABSTRACT**

---

**BACKGROUND:** Those involved in school health education programs generally believe that health-education programs can play an important role in helping young people make positive health decisions. Thus, it is to document the effects of such programs through rigorous evaluations published in peer-reviewed journals.

**METHODS:** This paper helps the reader understand the context of school health program evaluation, examines several problems and challenges, shows how problems can often be fixed, or prevented, and demonstrates ways in which challenges can be met. A number of topics are addressed, including distinguishing between curricula evaluation and evaluation of outcomes, types of evaluation, identifying stakeholders in school health evaluation, selection of a program evaluator, recruiting participants, design issues, staff training, parental consent, instrumentation, program implementation and treatment fidelity, participant retention, data collection, data analysis and interpretation, presentation of results, and manuscript preparation and submission.

**RESULTS:** Although there is a lack of health-education program evaluation, rigorous evaluations that have been conducted have, at least in some cases, led to wider dissemination of effective programs.

**CONCLUSIONS:** These suggestions will help those interested in school health education understand the importance of evaluation and will provide important guidelines for those conducting evaluations of school health-education programs.

**Keywords:** evaluation; research; school health instruction.

**Citation:** Young M, Denny G, Donnelly J. Lessons from the trenches: meeting evaluation challenges in school health education. *J Sch Health*. 2012; 82: 528-535.

Received on December 23, 2011

Accepted on February 12, 2012

---

Those involved in school health education generally believe that health-education programs can play an important role in helping young people make positive health decisions. It is important, therefore, to document the effects of such programs through rigorous evaluations published in peer-reviewed journals.

Over 50 years ago, Sellery and Bobbitt<sup>1</sup> deplored the lack of evaluation of school health programs. More recently, Stewart-Brown<sup>2</sup> noted in his review of health-education program evaluation review articles that many of the reviewers commented on the poor quality of published evaluation studies. Additionally, many evaluations are not published, and many programs are implemented without regard for evaluation.

For example, in discussing the need for evaluation, specifically evaluation of drug prevention programming, Hallfors and Godette<sup>3</sup> noted that schools often used curricula developed within the district that had not been evaluated. Additionally, many other schools use commercially available programs that have not been evaluated, curricula that have had inadequate evaluations, and those that have been evaluated but have been shown to be ineffective.

Rigorous evaluations of programs yielding positive results, published in peer-reviewed journals, can lead to wider dissemination of effective programs. For example, the National Registry of Evidence-based Programs and Practices lists programs that have been rigorously evaluated and have shown positive

---

<sup>a</sup>Professor, (myoung@nmsu.edu), CHSS-Southwest Institute for Health Disparities Research, New Mexico State University, Las Cruces, NM 88003.

<sup>b</sup>Professor, (gdenny@uark.edu), Educational Statistics and Research Methods, University of Arkansas, 231 Graduate Education Building, Fayetteville, AR 72701.

<sup>c</sup>Professor, (donnellyj@mail.montclair.edu), Health and Nutrition Sciences, Montclair State University, University Hall 4166, Montclair, NJ 07043.

Address correspondence to: Michael Young, Professor, (myoung@nmsu.edu), CHSS-Southwest Institute for Health Disparities Research, New Mexico State University, Las Cruces, NM 88003.

\*Indicates CHES and Nursing continuing education hours are available. Also available at: [http://www.ashaweb.org/continuing\\_education.html](http://www.ashaweb.org/continuing_education.html)

substance-use outcomes.<sup>4</sup> Some of the programs on this list are widely used in schools across the country.

Evaluation of school health-education programs is important. However, evaluation presents a number of different problems and challenges, not just for the evaluator, but for classroom teachers, school administrators/project directors, and funding agencies. This paper draws from real experiences in evaluating programs, helps the reader understand the context of evaluation, examines several evaluation problems and challenges, shows how problems can often be fixed, or prevented, and demonstrates ways in which challenges can be met.

## **BEFORE IMPLEMENTING THE PROGRAM**

### **Distinguishing Between Curricula Evaluation and Evaluation of Outcomes**

Curricula may be evaluated by comparing the content and teaching methods to some predetermined standard or they may be evaluated by examining the effects of the curricula on student outcomes. Wilson and her coworkers<sup>5</sup> published a review of 21 abstinence-only-until-marriage curricula. The researchers asked trained teachers to rate curricula according to how well they met 18 curriculum characteristics and 47 other factors in an overall evaluation rating. The researchers used the teacher ratings to rank the curricula from 1 to 21. This study evaluated the curricula, but did not examine outcomes, ie, the effect of a given curriculum on student knowledge, attitudes, or behaviors.

Denny and Young<sup>6</sup> examined the effects of 2 of these curricula, both from the Sex Can Wait curriculum series, and found that they produced positive student outcomes, including behavioral outcomes (upper elementary program—less likely to have had intercourse in last 30 days; middle school program—less likely to have had intercourse ever and in last 30 days) at an 18-month follow-up.

Wilson and coworkers<sup>5</sup> evaluated the curricula without examining the possible effects of the curricula on students. Denny and Young<sup>6</sup> examined the effects of curricula on student outcomes, but were not concerned with how well the curricula met standards for content or other predetermined standards or characteristics.

### **Distinguishing Among Evaluation**

#### **Types—Implementation, Process, Impact, Outcome**

Implementation evaluation has to do with the degree to which the program is actually implemented. If the program consists of 15 lessons, were all 15 taught? If the program was designed to reach high-risk youth in a particular community, are these youth participating in the program? Are the

staff members who are implementing the program adequately trained? The program may be a great one in theory, but if it is only partially implemented by teachers or staff members who are not adequately trained, and the target population is not participating, then one should not expect the target population to exhibit substantial changes in the outcome variables the program was designed to impact.

Implementation evaluation can be considered part of process evaluation. Process evaluation has to do with the conduct of the program. To a large degree it addresses the question: Did you do what you said you would do? If positive results were not achieved was it because the planned intervention was simply ineffective, or because a number of activities that were planned were never actually carried out?

Impact evaluation has to do with the immediate effects of the program on such things as knowledge, attitudes, and short-term behavior. Outcome evaluation examines longer-term effects. For example, a sexuality education program may produce immediate effects on knowledge and attitudes. It may take a little longer to identify behavioral changes. Reduction in sexually transmitted disease rates and rates of teen pregnancy are some examples of desired long-term outcomes.

### **Stakeholders and the Political Nature of Evaluation**

McDermott and Sarvela<sup>7</sup> address the political nature of evaluation and help us understand that different stakeholders may have different concerns about evaluation. For example, superintendents may be concerned that the evaluation may cast the district in a negative light, be the source of complaints, result in negative publicity for the district, and criticism for manner in which they are dealing with the issue. As one superintendent commented “All it takes is one parent dragging a reporter into a school board meeting about a question you ask on an evaluation, and I’ll be spending all of my time for at least the next week dealing with parent complaints.” Other superintendents are concerned that their district will be publicly identified and subjected to unfavorable comparisons to other districts on measures such as sexual behavior and drug use. They do not want their school to be seen as having a drug problem, or having students who are engaging in indiscriminant sexual activity.

Principals may have some of the same concerns as superintendents, but may also have concerns about disruption in the school day, extra burdens on teachers and staff, and time taken away from instruction in other “academic” areas. As long as health is not seen as an academic area, school health programs and evaluations of school health programs will not have a high priority.

Teachers may see an evaluation of student outcomes as a way for the administration or outsiders to judge their teaching. They may believe that if the curriculum does not produce positive student outcomes then they will be blamed and considered less than competent teachers.

Parents may view evaluations—especially evaluations that include student self-report of health behaviors—as a bit intrusive or as an invasion of privacy. Others may believe that questions that ask students about various health behaviors may actually encourage young people to engage in negative health behaviors.

Students may view a self-report questionnaire, used as part of the evaluation of a health-education program, as a complete waste of time. It is simply one more meaningless test to which they are subjected.

In light of all of these potential negatives, it is easy to understand why many schools may be reluctant to participate in program evaluations. However, we all engage in evaluation activities, though perhaps in a relatively informal manner, every day. We make a decision to purchase 1 product over others, to choose 1 activity over others, to decide that a particular way of performing a given task is better than others. A more formal evaluation of school health programs gives solid evidence to support the programmatic choices we make on behalf of our students.

### Selecting an Evaluator

A competent evaluator can be an asset to your project, helping you avoid evaluation problems and meet its challenges. Some funding agencies may require projects to employ an external evaluator. The ideal evaluator has expertise in applied research methods and communicates clearly to a general audience, while maintaining statistical precision. The ideal evaluator is also an effective manager who works well with those who collect data, has experience working in real-world settings, and produces reports that are accurate, clear, and on time. He/she will have a strong publication record, including published evaluations of health-education programs, in scholarly/professional journals.

### Recruiting Participants

A number of health-education grants have been awarded to non-school organizations to provide educational programs in a school-based setting. How do such organizations recruit schools to participate in the project? The organization needs approval to provide the program in a school setting and also needs permission to collect data from students. Schools should also understand that in many projects some schools will receive the program and other schools will serve as comparison schools. Recruiting is not always an

easy task, but with a good program to sell, and some school incentives, the task is not impossible. Most superintendents, principals, and teachers believe that students will benefit from a health-education program. It may be difficult, however, to convince school leaders to participate in a program in which their students complete questionnaires that ask explicit questions about personal health behaviors. To compound the problem further, it may prove more difficult to convince an administrator to allow his/her school to serve as a comparison school than to serve as an intervention.

One superintendent explained it this way. “Clearly our community has a problem with teen pregnancy. I believe our students will benefit from your program. I’ll have no problem defending the program to our parents and other community members. I know I’ll catch some heat from folks about the evaluation questionnaire. I can defend this too. I’ll say that the ‘University is providing us with a fine program, materials, training, and implementation assistance - at no cost. They need to show the funding agency the results of implementing the program. Having our students participate in the evaluation is just the ‘cost of doing business.’ So we want to participate in the project—if you ‘randomly’ assign us to the intervention group.”

He explained that his school could not participate as a control school: “Look, I can defend our participation as an intervention school to anyone, but not as a control school. Our kids will still have to complete the questionnaires. I will catch as much grief from parents for kids completing the questionnaires as I will for them participating in the program. The difference is that if we are in the intervention group I can show any critic how we benefit from the project. If we are in the control group and someone wants to know what we are getting out of it, I’ll have to say ‘I don’t know.’ We want to participate, if you can ‘randomly’ assign us to the intervention group.”

One way to respond to this concern is to use in-house control groups. Some students in a given school receive the program and other students in the same school are part of the control group. Although some administrators might prefer that all of their students be placed in the intervention group, this arrangement may be preferable to placing all students at a given school in the control group. Another way to use in-house control groups is to assign some classes to receive the intervention at the beginning of the school year and other classes to receive the intervention at the end of the school year. When the intervention is not lengthy ( $\leq 5$  weeks), the evaluators have time to conduct a pretest, implement the program, conduct a posttest, conduct a 6-month follow-up test, and still have time before the end of the school year to offer the program to the control classes. The disadvantage

is that the control group is lost for long-term follow-up.

A limitation of in-house control groups is the possibility of control group contamination, or treatment diffusion. This can reduce the size of the difference between intervention and comparison groups, making it more difficult to show statistically significant results.

Another alternative is to use intervention and control schools, but provide control schools the opportunity to implement the program in the second year. If the program is provided to eighth graders at school A, eighth graders at school B may serve as the control group. In the second year new eighth graders at school B have the opportunity to participate in the program. The original eighth graders at school B remain as the control group giving the evaluator an opportunity to follow participants over time.

### **Staff Training**

Whether the projects involve staff from community-based organizations teaching a health-education curriculum or classroom teachers teaching the program, it is important to train the people who will teach the curriculum. In our trainings we have not simply taught them the curriculum, we have helped them understand the theory on which the curriculum is based. Teachers know the points of emphasis and why the curriculum is effective. We also spend time giving teachers hands-on practice in answering questions they are likely to get from students, an especially important skill when dealing with sensitive topic areas. Implementation of a strong curriculum by trained staff/teachers with strong teaching skills and a good grasp of how to implement the curriculum will be more likely to show positive results than implementing the curriculum without providing this important training.

### **Institutional Review Board Approval and Parental Consent**

Research ethics require informed consent from all participants, and from parents or guardians of children who participate. If a program is part of the school curriculum it does not need institutional review board (IRB) approval, but if data are collected for research purposes, then parental consent for the data collection is required. Depending on the requirements of the IRB, parental consent may be either active, where parents and students both sign an informed consent form, or passive, where parents may opt out of having data collected from their child. Our experience has been that parent information nights are generally well attended and, once they understand the project, most parents and students are willing to sign consent letters, even though we make it clear that the questionnaire includes items about personal behavior, such as drug use or sexual behavior. We must admit, however,

that some parent nights are not well attended, and, on some occasions, the parents who do attend do so because they are not supportive of the program. Nevertheless, we believe that it is important to give parents an opportunity to learn about the program and know that some, though certainly not all, opponents can be won over by a clear explanation of the program.

### **Design Issues**

A common design for many programs is a pre/post design. Every student completes a pretest survey, participates in the educational program, and then completes a posttest survey. It is important to match the pretest responses of individual students to their responses at posttest. Pairing responses requires some way to uniquely identify each student at pre and post so that responses can be matched over time, while safeguarding the confidentiality of student responses. Pre-post designs without a comparison group have several flaws that make it difficult to gauge the effectiveness of the education program. It is possible that events outside the program could influence outcomes. Including a comparison group helps to distinguish the effects of the program.

Another serious threat is maturation. As young people get older they are more likely to engage in behaviors that we might prefer they avoid. For example, the percent of teens that have had sexual intercourse increases by 12% per year as they get older.

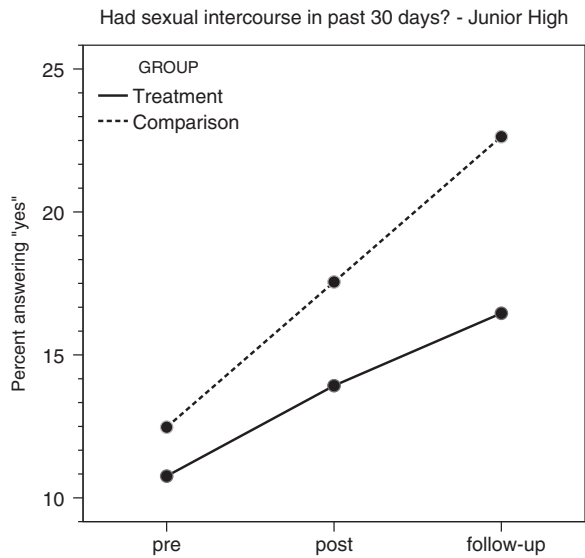
A national survey by Abma et al<sup>8</sup> found that from ages 14 to 19 the percent of teens that had ever had intercourse increased from 6% to 70% for females and from 8% to 65% for males. Because of this type of trend, pre/post designs often find higher rates of risky health behaviors after a program than before. Some have argued that these types of results are evidence of program ineffectiveness. Such findings, however, need to be compared to those of an equivalent comparison group that did not receive the education program (Figure 1).

Another deficiency in a pre/post design, even if there is a comparison group, is that it measures short-term effects on the outcomes. A better approach is to add longer-term follow-ups to the design that track whether the changes persist over time. Often program participants will show a short-term gain in attitude or knowledge, but a year or 2 after the program, these outcomes revert to baseline levels. In contrast, health-behavior changes from pretest to posttest may be too small to detect, but with a longer-term follow-up, behavioral differences between the treatment group and the comparison group can be detected.

### **Instrumentation**

Instrumentation refers to the testing instruments that will be used to elicit data. In most evaluations of

**Figure 1. Students in the Treatment Group Are Having Sex More After the Program Than Before, but the Increase Is Slower Than for the Comparison Group**



health-education programs, responses are collected via a self-report questionnaire. Several types of outcomes should be considered to assess the effectiveness of a health-education program.

**Knowledge.** Asking factual questions based on the content of the lessons can assess whether students learned from the program. Although greater knowledge does not correlate strongly with behavior, knowledge gains provide some indication of treatment fidelity. If teachers present the curricular materials clearly and thoroughly, students are expected to learn. If students do not learn, it may be an indication that the program might not have been implemented well.

**Attitude.** Attitude items ask students their personal views, eg, "There are a lot of good reasons not to have sexual intercourse while you are still a teenager." Program effectiveness may be measured by changes in the desired direction.

**Intent.** Behavioral intentions are a precursor to behavior, so students respond to items asking how likely it is that they will engage in a behavior over a specific time period. For example, students might be asked to rate the chance they will start (or continue) smoking cigarettes—(a) this year, (b) while in high school, and (c) after high school graduation.

**Behavior.** Health educators are interested in helping young people make healthy behavioral choices, but many programs are reluctant to ask young people about behavior. Questions about substance use or sexual behaviors are often seen as too sensitive to be included in a program evaluation. Knowledge, attitudes, and intentions are important variables to address as part of an evaluation. They are not, however, the same as behavior. Give serious consideration to the

inclusion of behavior items as part of any evaluation of school health-education programs.

## DURING THE PROGRAM

### Program Implementation and Treatment Fidelity

If the project is designed to test the effects of a given curriculum on student outcomes, it is important for the curriculum to actually be implemented as it was intended. A number of years ago we provided training for law enforcement officers to implement a school-based drug education program. The instructional specialist at 1 of the project schools called to indicate that the officer we had trained was doing a fantastic job. She offered to videotape each session he presented and send us a copy, thinking perhaps we could use portions of the tape as a training video at future workshops. Because there was little money to actually visit schools to monitor program implementation and because we were always looking for ways to improve the training we provided, we gladly accepted her offer. Upon review of the tapes she sent, we found the officer to be well organized and at ease in the classroom. His presentation was entertaining and informative. He told a few jokes, and even did some magic tricks. None of what he presented, however, was a part of the curriculum we thought he was teaching. Had we not seen the videos we would have assumed he was implementing the curriculum we provided.

In more recent projects travel funds have been available to support trips to schools to monitor curriculum implementation. A project staff member sits in classrooms to observe what teachers are teaching.

Some teachers are better than others at presenting information, facilitating discussion groups, etc. All teachers, however, should present a lesson that looks like the lesson from the curriculum they are supposed to be teaching. If the lesson taught when staff members are there to observe does not look much like the curriculum, then it is likely that this is also the case when staff members are not there. In addition to classroom visits, teachers are also asked to complete check sheets to indicate which lessons from the curriculum they actually taught, the activities within each lesson that were covered, and provide comments as to how the lesson was received by the students. Many times schools participate in a project as part of a comparison/current practice group, ie, they are not implementing the intervention curriculum but continue to teach the program they were already using. They may address some of the same topics covered in the intervention curriculum. Thus, teachers in comparison classrooms are asked to keep track of the material they teach by checking off topics taught. They also indicate the time spent on given topics, and provide comments as to how well the students received the material.

### **Participant Retention**

Part of the problem facing the evaluators is to have a sufficient number of students participating in the evaluation, either as students receiving the intervention or as students in the comparison group. Another part is retaining them. Retention/attrition is a problem, in part, because students who are not retained may differ, in ways that may influence the outcome of the evaluation, from those who stay in the study.

When project staff members travel to schools to collect data, not all students will be available. Some students will have moved. Other students are absent. Sometimes a special event will take students out of school. Individual teachers may have something special planned and do not want to release their students for an hour to complete a questionnaire. Some of our project schools have been as much as 6 hours away. It is not always easy to go to a school, and then go back again another day. Make sure to schedule dates and times with school principals, or other designated contact people, well in advance, and call a day or 2 before you go, to make sure everything is still okay. Even with advance preparation there will still be some glitches.

Another issue that can be a problem is a change in school administrations. In 1 of our projects, 2 school administrators who had been supportive of the project moved on to other schools. Their successors were not interested in having their schools participate. Even though our university's general counsel indicated this was a clear breach of contract, the university was not going to get involved in a legal squabble with a school district so we could collect evaluation data for our project.

What lessons were learned? First, keep projects a manageable size. Second, communicate with participants frequently. We had far too many school districts, often 15 or more at a time, spread over far too great a geographic area, for our small staff to communicate as frequently and to be on-site as frequently as we would have liked. In a current school-based project we have 6 different schools, all within 7 miles of the university. I visited 1 school 4 different times in less than 2 weeks. I talked with students in 7 different classes at 3 different schools in 1 day. A contact person at each school, usually the principal, knows who I am. We stay in frequent contact about the project. There are still problems, but with fewer schools and close proximity, problems can be more easily managed.

### **Data Collection**

The most frequently used method of data collection is use of a self-report questionnaire. Although there are limitations associated with this method, evaluators can do much to maximize the validity of self-report

data. For example, (1) administer questionnaires in the students' regular classroom setting. This avoids inevitable distractions that occur when large groups of students are herded into the school cafeteria or gymnasium. (2) Read the questionnaire items aloud. Many students may not be good readers. (3) Let students know that while you hope they will answer all questions, it is okay to leave a question blank if it makes them uncomfortable, or if they do not want to answer it. (4) Encourage students to answer questions honestly, to not visit with others, or to look at others' answers while they are completing the questionnaire.

One concern some students have concerns confidentiality of the data; will other people find out how they answered the questions? One way to deal with this concern is to say "Yes, people will probably find out—but because you will probably tell them." They may be concerned that some "hacker" can find their answers. Help students realize that it would be highly unlikely that someone could, or would, do this. Why go to all of the trouble of trying to hack into a student's responses, when all he/she has to do is ask the student?

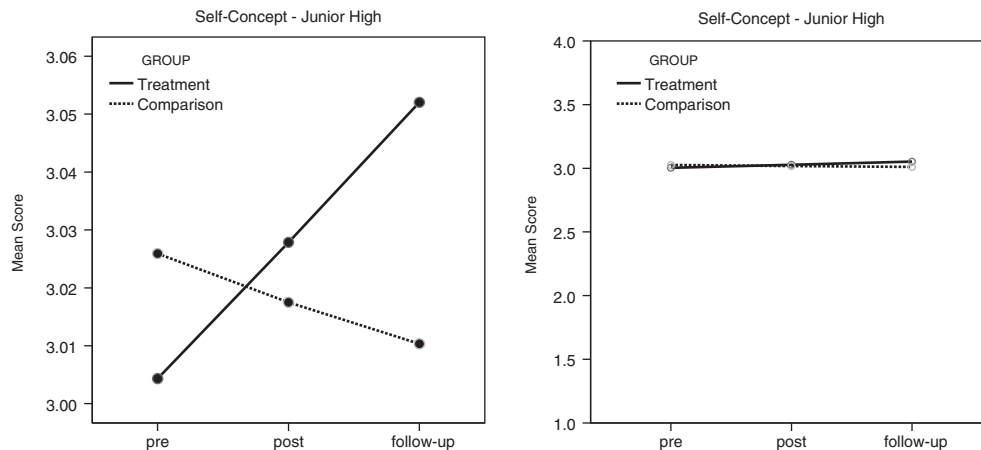
### **AFTER THE PROGRAM**

#### **Data Analysis and Interpretation**

Conduct data analysis as soon as possible after the data are collected. One advantage to quick turnaround is prompt feedback to the program. Know that data cleaning—getting data from the student response sheets into a format ready for computer software analysis—usually takes longer than the data analysis itself. If data have been collected at several different time points, getting student identification numbers and other variables to match longitudinally can be difficult, especially with the "messy data" collected from young teens.

The fundamental question in evaluating the program is whether it made a difference in the outcome variables. A simple *t*-test for comparing means or a chi-square test for comparing percentages is not always the best approach. A stronger approach is to account for initial differences in individuals, as measured at pretest, when testing for differences between groups at posttest. The pretest should include measures likely to correlate with posttest outcomes. For example, if the posttest outcome is intent to practice abstinence, as measured on a continuous scale, an analysis of covariance could control for differences at pretest on relevant variables including age, sex, virginity status, and certainly the pretest measure of intent to practice abstinence. For a dichotomous outcome such as having smoked a cigarette in the last 30 days, a similar approach using logistic regression can find effects of the program, taking into account pretest differences on variables that correlate with recent

Figure 2. Two Graphs of the Same Nonsignificant Finding, First With the Default Scale Values of the Software, and Then With the Vertical Axis Rescaled to the Range of the 4-Point Scale



cigarette smoking such as age, sex, intent to smoke/not smoke cigarettes, and recent cigarette smoking at pretest.

This analytic approach may seem unnecessarily complicated, but including covariates gives a truer measure of the effects of the program. If 1 group were older at pretest, we would expect them to engage more in behaviors such as sexual behavior, smoking cigarettes, or drinking alcohol at posttest. By including age as a covariate, we reduce the bias of age on the outcome. Besides reducing bias, including relevant covariates increases statistical power, or the likelihood of finding statistical significance.

### Presentation of Results

In reporting results to an audience that lacks statistical training, it is helpful to include an executive summary that reports the purpose of the program, the outcomes that were addressed, and an overview of important findings of the analysis. A more detailed report including specific findings should follow the executive summary. This report includes participation or response rates, internal consistency/reliability estimates for multi-item scales used, as well as descriptive and inferential statistics—including effect sizes—for all statistical tests conducted. For the descriptive statistics on continuous outcomes, report means, standard deviations, and sample sizes for each group, and on categorical outcomes report group percentages and sample sizes. Clear descriptive statistics are important for making the magnitude and direction of findings clear, and are also necessary for any future meta-analysis that includes the study. When reporting results of statistical tests, include the value of the test statistic, the degrees of freedom, the associated probability, and a measure of effect size such as  $R^2$ .

Line graphs can be used effectively to show changes in outcomes over time, but graphs should be scaled to reflect the range of the outcome. Sometimes the default values for graphs in statistical software can exaggerate the magnitude of the findings. For example, Figure 2 shows 2 graphs, the first with the default scale assigned by the statistical software and the second with the scale modified to the 1 to 4 scale used on the self-concept measure. The first graph makes it appear as if the self-concept increased greatly for the treatment group and decreased for the comparison group. The second graph makes it clear that the changes were small, and the findings were not statistically significant.

After answering the primary question of whether the program was effective in increasing desirable outcomes, the evaluator may also want to ask whether the treatment was especially effective for certain subgroups of students. Demographic variables can serve as moderator variables, and the statistical question is whether there is a significant interaction between treatment group and a demographic variable. For example, a statistically significant group-by-sex interaction for behavior suggests that the treatment was more effective in changing behavior for 1 gender over the other.

### Manuscript Preparation and Submission

An evaluation report can be adapted for publication in a refereed journal. The introduction includes the theoretical and empirical background to the study as reported in the literature. Methods should be described in enough detail that someone else could replicate the study. Results need to be reported with few tables and figures, combining outcomes into a single table whenever possible. A discussion that includes implications of the findings for the field and suggests directions for further study concludes the article.



Many journals are appropriate for evaluations of school health-education programs. Consult the journal's guidelines for authors. Use articles previously published in the journal as examples. Be clear how your article adds to the literature. Write in a scientific, unbiased style, with clarity of expression as the primary goal.

### **Maintaining the Relationship**

A successful evaluation project can be a 1-time occurrence, or it can be the beginning of a long-term relationship between schools and the evaluation team. A university faculty member who has secured grant funds to test an intervention and has recruited schools to be a part of the evaluation project may head the evaluation team. Too often faculty members are concerned with implementing the project, collecting data, and publishing articles, but fail to establish ongoing relationships with the project schools. Evaluators can sustain a relationship by providing personal feedback to schools, articulating program benefits to parents, students, and school personnel, making sure control schools do have an opportunity to, at some point, to receive the program, and by recognizing schools as vital partners in the process. When school personnel feel valued, they are more likely to participate in subsequent projects, giving faculty members more opportunities to evaluate school-based interventions, and schools more opportunities to enhance the health-education offerings they provide to students.

### **IMPLICATIONS FOR SCHOOL HEALTH**

School health education programs can make a difference in the health of young people. It is important, however, to document the effects of school

health-education programming, through rigorous program evaluations published in peer-reviewed scholarly journals. Conducting a solid evaluation, one that can potentially yield publishable results, is not easy. We have identified some of the challenges that must be addressed—challenges prior to program implementation, during program implementation, and after programs are implemented. We hope this information is of help as you develop your evaluation plans.

### **REFERENCES**

1. Sellery CM, Bobbitt BG. Evaluation of health education and health services in the Los Angeles City Schools. *J Sch Health*. 1960;30:81-85.
2. Stewart-Brown S. What is the evidence on school health promotion in improving health or preventing disease and, specifically, what is the effectiveness of the health promoting schools approach? Copenhagen, WHO Regional Office for Europe (Health Evidence Network Report). Available at: <http://www.euro.who.int/document/e88185.pdf>. Accessed March 1, 2006.
3. Hallfors D, Godette D. Will the 'Principles of Effectiveness' improve prevention practice? Early findings from a diffusion study. *Health Educ Res*. 2002;7:461-470.
4. NREPP—SAMHSA's National Registry of Evidence-based Programs and Practices. Available at: <http://nrepp.samhsa.gov/>. Accessed February 10, 2012.
5. Wilson K, Goodson P, Buhi E, Davis-Gunnels E. A review of 21 curricula for abstinence-until-marriage-only programs. *J Sch Health*. 2005;75:90-98.
6. Denny G, Young M. An evaluation of an abstinence only sex education curriculum: an 18 month follow-up. *J Sch Health*. 2006;76:414-422.
7. McDermott RJ, Sarvela PD. *Health Education Evaluation and Measurement: A Practitioner's Perspective*. Madison, WI: Brown & Benchmark; 1999.
8. Abma JC, Martinez GM, Mosher WD, Dawson BS. Teenagers in the United States: sexual activity, contraceptive use, and childbearing, 2002. National Center for Health Statistics. *Vital Health Stat*. 2004;23:24.