Evaluating HIV Education Programs

BOOKLET 1

DIVISION OF ADOLESCENT AND SCHOOL HEALTH
NATIONAL CENTER FOR CHRONIC DISEASE PREVENTION AND HEALTH PROMOTION
CENTERS FOR DISEASE CONTROL

Evaluating HIV Education Programs

W. James Popham

This booklet was prepared by IOX Assessment Associates under Contract No. 200-88-0683 with the Division of Adolescent and School Health, Centers for Disease Control. Preliminary draft: May 1992.

TABLE OF CONTENTS

Page

INTRODUCTION 1
What is HIV education? 1
Why evaluate HIV education? 1

GUIDELINES FOR HIV EDUCATION EVALUATORS 2

Guideline 1: Focus on a manageable number of important program-related decisions. 3
Two kinds of decisions 3
The role of program objectives 5
Focusing on major decisions 6
Must all decisions be linked to attaining objectives? 7
Final thoughts about Guideline 1 7

Guideline 2: Select and administer suitable assessment instruments. 8
An emphasis on student outcome data 8
How to acquire suitable assessment devices 10
Securing permission to gather data 11
Confidentiality considerations 12
What about qualitative data? 13
Final thoughts about Guideline 2 13

Guideline 3: Use a data-gathering design consistent with the orientation of the evaluation. 14
A data-gathering design for program-improvement evaluations 14
A data-gathering design for program-continuation evaluations 17
Sampling 19
Final thoughts about Guideline 3 20

Guideline 4: Use data-analysis procedures that yield understandable results. 20
Practical versus statistical significance 20
Comprehensibility of results 21
Final thoughts about Guideline 4 22

Guideline 5: Report your results using a multilevel reporting scheme featuring written and oral reports. 22
An appropriate level of detail 23
Oral reporting 23
Making recommendations 24
Final thoughts about Guideline 5 24

CONCLUSION 25

ADDITIONAL READING 27

Introduction

What is HIV education?

AIDS (acquired immunodeficiency syndrome) was identified as a new clinical condition in 1981, with HIV (human immunodeficiency virus) discovered as its cause soon after. Since that time, policymakers have responded in various ways to the crisis. Many educational policymakers have agreed that schools should provide HIV-related programs to educate students and help them eliminate, or at least greatly reduce, their likelihood of becoming infected with HIV. Such HIV education programs began to be widely offered to students in our nation's schools during the late 1980s. The fact that HIV infection almost certainly results in serious illness and premature death, makes the stakes of HIV education higher than those educators commonly face and the thorough evaluation of these programs vital. This set of basic guidelines has been designed to assist in such evaluation.

Many HIV education programs are now available for students at various grade levels, most often in junior high schools. Sometimes this HIV education is part of comprehensive school health education. In other instances, special HIV education programs are inserted into existing courses, such as psychology, science, or guidance classes. In still other situations, a separate HIV education program is offered via special assemblies or minicourses.

Placement within the school curriculum is not the only difference among HIV education programs; the duration and intensity of these programs vary as well. In some settings, there is a strong commitment to prepare students to avoid behaviors that place them at risk of HIV infection. Such HIV education programs, often provided in the context of comprehensive school health education, may extend over several weeks and strive to provide students with a wide range of skills and knowledge with which to avoid HIV infection. Other HIV education programs are, unfortunately, much less substantial. These perfunctory programs, lasting no more than an hour or two, offer students little more than the most rudimentary information about HIV and preventing its transmission.

Why evaluate HIV education?

The common aim of all HIV education programs, regardless of form or fervor, is to help students avoid becoming infected with

HIV. Not every HIV education program, of course, can successfully protect all students from HIV infection. It is precisely because of the high stakes already noted, however, that thorough judgments of a program's success are particularly important. By systematically evaluating HIV education programs, we can see whether those programs have been effective.

The five guidelines provided in the following pages are intended to assist those responsible for evaluating school-based educational programs. More specifically, these guidelines address program evaluation procedures to help (1) improve HIV education and (2) determine the success of an HIV education program. For either of these purposes, program personnel will make a number of decisions concerning the HIV education program based upon the information supplied by the program evaluation.

The guidelines in this booklet are deliberately fundamental. They are intended to assist busy educators who need to evaluate their HIV education programs efficiently. These guidelines do not deal with advanced aspects of program evaluation; numerous available textbooks provide sophisticated treatments of such topics. A set of references is included at the end of this booklet for those interested in further pursuing the topic of program evaluation.

This booklet presumes that you, the reader, need to conduct or oversee the evaluation of an HIV education program. These guidelines address key procedural steps that you can follow in carrying out an appropriate evaluation. They deal specifically with fundamentals—the nuts and bolts of evaluating HIV education. Only rarely will you find discussions of possible procedural alternatives. To keep this booklet brief enough to be read and used by busy people, the guidelines more often than not embody "do this, then that" procedural suggestions.

Guidelines for HIV Education Evaluators

The topics to be addressed by the guidelines are (1) the evaluation study's focus, (2) selecting appropriate assessment devices, (3) choosing a data-gathering design, (4) analyzing the data, and (5) writing evaluation reports. Each guideline will be addressed by discussing the reasoning behind it and describing how it should be implemented. Although the guidelines are provided in a rough sequential order, you may find that you need to skip a step or repeat some steps more than once along the way.

Guideline 1: Focus on a manageable number of important program-related decisions.

When many educators hear the expression "educational evaluation study," they almost instinctively think of a study designed to determine if an educational program is *good* or *bad*. That view of educational evaluation, fortunately, is way off the mark. An educational program is evaluated for one fundamental reason: to provide information to help people make better decisions about the program.

Properly conceived educational evaluations help decision makers arrive at better decisions. The evaluator's responsibility, then, is to gather information, or evidence, for these decision makers. Sometimes, of course, the evaluator of an HIV education program will be the same person who manages the program.

Focus on a reasonable number of decisions.

The kinds of decisions that must be made by those who staff an HIV education program might deal with (1) what content to include in the program, (2) how much instructional time to allot to different topics, (3) how to organize instructional components effectively, and (4) what to do when certain parts of the program appear to be unsuccessful.

Two kinds of decisions

Decisions that relate to educational programs can be classified into two major categories. The first category includes decisions that improve the program and allow it to function more effectively. These are *program-improvement* decisions. The second category focuses on more fundamental go/no-go decisions, that is, whether to continue or discontinue the program. These decisions are *program-continuation* decisions. A decision might be made, for example, to terminate an existing HIV education program and replace it with a substantially different program. The type of decisions that need to be made directly determine the type of information you seek and the approach you will take in your evaluation.

If you are carrying out an evaluation study designed to assist with program-improvement decisions, you can be decidedly partisan. You are in every sense a "member of the team," and your chief responsibility is to boost program effectiveness. As we will see, a program-improvement evaluator can use data-gathering techniques that would be a poor choice for program-continuation evaluations.

On the other hand, when carrying out a program-continuation evaluation study, you must be completely objective and

non-partisan. Your evaluation should be governed by your need to supply accurate and credible evidence to those who will decide whether to continue the program.

Decision makers associated with an HIV education program typically fulfill one of two functions. First, they may be people who actually design and/or deliver the HIV education itself. These typically are teachers or health education curriculum specialists. Such people are mainly concerned with decisions to improve the program. Program-improvement evaluation studies are particularly appropriate in the early years of a program's existence when the instructional staff is trying to eliminate the program's deficits and make it stronger.

The second category of decision makers consists of those who authorize or fund an HIV education program. School board members at the district or state level usually set the policies that establish the programs. State or federal officials often supply funding for the programs. Both program authorizers and program funders are usually more concerned with program-continuation decisions than with program-improvement decisions. Program-continuation evaluations usually are made after a program has been in place for a few years, when it is appropriate to determine if the program is worth the money it's costing.

Most decision makers with whom you will work think of educational evaluation exclusively as a program-continuation enterprise. When many school board members or district-level administrators hear the phrase "evaluation study," they immediately imagine a study designed to determine whether or not they should continue the educational program. One of your tasks as an evaluator is to *educate* decision makers to realize that it often makes sense to evaluate programs in order to improve them, particularly in their early years of existence.

Some people mistakenly assume that once a program has been evaluated, there is no need for further evaluation. In fact, however, program evaluation should be conceived of as an ongoing enterprise. As early versions of the program are offered, they can be improved via program-improvement evaluation. Later, when the program staff believe the program is sufficiently mature, a program-continuation evaluation might be undertaken. But even mature programs can be improved. Thus, evaluation of programs for program improvement should be a continuing activity.

It is sometimes thought that evaluations of HIV education are successful only if they reveal that the program was effective. On the contrary, an evaluation that reveals a program's shortcom-

ings can point the way to program improvements and ultimately program effectiveness. Evaluations that enhance the quality of decisions are successful evaluations.

Guideline 1, to focus on a manageable number of important program-related decisions, stems from a basic human shortcoming: people can only make use of so much data. If they are given too much information, even at their own request, they are likely to become overwhelmed and, as a consequence, pay attention to none of it. The decision makers for whom you gather evidence in your evaluation will be no different. They will often want more answers than they can really use.

Regardless of whether you pursue an evaluation study aimed at program improvement or program continuation, one of your early tasks is to focus on an intellectually manageable number of decisions related to the HIV education program. Skillful evaluators focus on significant decisions, not "nice to know" information.

The role of program objectives

An HIV education program staff usually aspires to bring about worthwhile changes in students. Those changes can focus on altering either students' HIV-risk behaviors or the factors thought to contribute to such behaviors. Put most simply, an instructional objective for an HIV education program should describe the post-program knowledge, skills, attitudes, or behaviors that the program seeks to promote. This is nothing more than a classic ends/means distinction. Identifying a program's objectives can lead to identifying the decisions on which you will focus your evaluation.

A number of educators attempt to describe educational objectives as what the program itself will do rather than what it is intended to accomplish. Educational objectives have nothing to do with what the HIV education program is or how it was created. Instead, the objectives for HIV education must focus on program outcomes—that is, on what happens to students as a consequence of the program. Clearly stated, measurable objectives will provide a valuable yardstick for your evaluation of a program's effectiveness.

If you can help a program's staff identify the objectives that they hope to accomplish, and if you can help the staff define those objectives as preprogram-to-postprogram changes in students, you will have gone a long way toward clarifying the focus of your evaluation. You can then recognize and isolate evidence bearing on key program decisions.

Evaluators who wish to use an HIV education program's objectives to their advantage will need to be sure that the program is organized around only a handful of measurable objectives. Some researchers cite evidence suggesting that people have a difficult time concentrating on many more than six or seven issues at a time. Rarely permit your evaluation, therefore, to be organized around more than a half-dozen or so objectives—preferably fewer.

A program staff may have a number of fine-grained instructional objectives to use in day-to-day instruction. As an evaluator, however, your responsibility is to isolate a few educational objectives that subsume such day-to-day objectives, then gear your data-gathering toward that smaller number of important objectives.

Focusing on major decisions

One of the best ways for you to focus an evaluation study on a manageable number of decisions is to encourage decision makers to identify their most important program-related decisions. Work with decision makers to identify the one most important decision at issue, then the next most important decision to be informed by the evaluation study, and so on.

Another useful ploy to reduce the number of decisions is to group small-scope objectives into one broader yet still measurable objective. Suppose you were carrying out a program-continuation evaluation of HIV education for a district's school board. You have learned that the teachers providing the district's HIV education program have listed four different objectives each dealing with a distinctive type of knowledge students need to acquire, such as "knowledge regarding HIV infection routes" or "knowledge regarding HIV-risk behavior patterns of teenagers." It would be relatively simple to create a single objective, "increased HIV-relevant knowledge," to effectively coalesce the four small-scope objectives. The teachers can still organize day-to-day instructional activities around small-scope objectives, and you can focus your evaluation on broad-scope objectives.

In spite of your efforts to focus on a limited number of decisions, some decision makers may request information simply because it would be "interesting." Keep urging these persons to indicate how their requested information would actually make a difference in a decision dealing with the HIV education program.

One good way to verify whether a requested set of evidence really bears on a program-related decision is to present decision makers with hypothetical results and ask, "If the evidence turns out

this way, what would your decision be?" Then present a divergent set of information, asking, "If the evidence turns out the opposite way, what would your decision be?" You may discover that it makes no difference to decision makers what the results are. In such instances, of course, encourage these persons to seek other, more relevant evidence.

Must all decisions be linked to attaining objectives?

Although the decisions addressed by evaluators are often linked to the achievement of a program's objectives, decision makers face many choices that do not depend on the attainment of objectives. For example, evaluators often gather evidence as to whether an instructional program is being delivered as intended. The decision at issue in this instance would be whether the program's staff must take steps to ensure that the program is being provided as its designers intended.

Other such decisions include (1) whether community officials will permit controversial topics to be addressed in instructional activities, (2) whether students will regard HIV-related information as more believable if provided by peer counselors rather than teachers, and (3) whether the program's objectives are appropriate. There are also instances in which unanticipated effects of the program, that is, effects not foreseen in the program's objectives, might be significant in judging a program's effectiveness.

In short, although the degree to which an HIV education program's objectives have been achieved can illuminate certain kinds of decisions, other kinds of decisions will demand that the evaluator adopt alternative approaches.

Final thoughts about Guideline 1

The purpose of evaluating HIV education programs is to help those who must make decisions about the program do so in an appropriate manner. You will discover that evidence gathered in an evaluation study often plays only a minor role in the decisions ultimately made about a program. For many decision makers, a series of political and personal factors play a far more prominent role than any evidence of program effectiveness provided by an evaluator. As an evaluator, therefore, you will need to structure your evidence-gathering efforts to yield information that has at least a reasonable chance of affecting the decisions to be made.

Guideline 2: Select and administer suitable assessment instruments.

Secure and use assessment devices.

One of the evaluator's most important tasks is choosing which information to assemble for decision makers. Guideline 2 deals with the instruments you will use to gather decision-relevant data.

Evidence regarding changes in student behavior, which is the outcome typically sought by educational programs, can be described as *outcome data.* Outcome data represent the effects of an educational program. Evidence regarding the nature of the educational process itself, in contrast, is referred to as *process data.* Typically, process data are gathered when the evaluator wants to determine whether an instructional program is being provided as intended.

A variety of "process instruments" are provided in this handbook regarding the quality and implementation of HIV-related policy, curriculum, and staff development programs. In addition, CDC recommends using the *HIV Education Survey*, developed cooperatively with state and local education agencies, to collect data on HIV education programs. The *HIV Education Survey* collects information on the number and percentage of schools providing and the number and percentage of students receiving HIV education, as well as on teacher training, curriculum, content, scheduling, provision within special education, and barriers to instruction. A handbook for conducting this survey and software to assist with selecting schools and summarizing data are available from CDC (404/488-5330) or Westat, Inc. (800/937-8287).

An emphasis on student outcome data

Students receiving the HIV education program supply the bulk of the data the evaluator typically gathers. One method of gathering such data might be to have students participating in the program fill out anonymous questionnaires. Because evaluators in most cases will be interested in the changes in student behavior resulting from HIV education, a questionnaire will typically be given to students both before and after the program.

Evaluators of HIV education programs should attempt to secure four types of student outcome data:

- Evidence of the extent to which students engage in HIV-risk *behaviors*

- Evidence of students' ability to display key *skills* needed to reduce their likelihood of being infected with HIV
- Evidence of students' *attitudes* that are likely to influence their HIV-related behaviors
- Evidence of students' *knowledge* regarding those aspects of HIV and AIDS apt to influence their HIV-related behaviors.

As we see, the four categories of student outcome measures are behavior, skills, attitudes, and knowledge.

The same four general categories of outcome data can be used to evaluate HIV staff development programs for teachers who will deliver the instructional program for students. In these staff development activities, of course, teachers are the "students." Although the nature of these outcomes will be different, the categories remain essentially the same. Table 1 presents illustrations of the sorts of outcome evidence that might be sought when evaluating (a) an HIV education program for students and (b) a staff development program for teachers who will provide HIV education. (Although the guidelines provided in this booklet are directed toward the evaluation of student-focused HIV education, in most instances they can also be used to evaluate staff development programs for HIV educators.)

Table 1. Illustrations of Relevant Types of Evidence for Students and Teachers in HIV Education Programs

Evidence Category	For Students' HIV Education	For Teachers' HIV Staff Development
Behavior	Reported activities while in high-risk situations	Appropriate use of recommended classroom procedures
Skills	Ability to display refusal skills in simulated high-risk situations related to HIV infection	Ability to respond appropriately to students' questions about sensitive topics
Attitudes	Perceptions regarding one's personal susceptibility to HIV infection	Confidence in being able to modify students' high-risk behaviors
Knowledge	Knowledge regarding the routes by which HIV is/is not transmitted	Knowledge regarding the instructional principles relevant to modifying students' attitudes

Even though the program's decision makers will ultimately decide the sorts of evaluative evidence you should collect, you should certainly encourage them to gather behavioral data in nearly all evaluations of HIV education. Many HIV education programs only attempt to influence students' knowledge regarding HIV. Yet ample evidence indicates that knowledge-only programs typically have scant influence on students' behaviors.

Ideally, you should encourage the use of an assessment strategy in which evidence is gathered about students' behavior, skills, affect, and knowledge. The nature of the HIV education program itself will prominently determine which outcomes you should measure.

Of the four types of outcome data from students, the most important is behavioral data. Strive to collect student behavioral data if at all possible. This will sometimes oblige you to provide education and information to local community groups to overcome obstacles on the collection of sensitive behavioral data. Try as hard as possible to assemble evidence of the HIV education program's impact on students' HIV-risk behaviors. Without such evidence, a misleading picture of the program's effectiveness can emerge. HIV education programs that enhance only students' knowledge or attitudes may be judged effective when behavioral data would indicate otherwise.

It is difficult, of course, to demonstrate that an HIV education program has produced genuine changes in young people's behaviors. Part of the difficulty lies in having a sufficient period of time to discern changes in behavior. Program effectiveness may not be apparent for six or more months following the intervention, particularly if rates of sexual intercourse are low. Nevertheless, the program's staff has a responsibility to judge its efforts according to the changes that take place in students' behaviors.

How to acquire suitable assessment devices

Once you have decided to measure the four types of student outcomes we have been discussing, where do you get your assessment instruments? There are two possible ways to proceed. You can either construct the instruments yourself or use (perhaps adapt) existing instruments. In selecting, adapting, or constructing your assessment devices, it is extremely important to ensure that they match your program objectives. Measuring an attitude or behavior change that was not sought as part of your instructional program may set up the program for failure.

The problem with creating your own assessment devices is that wording questions to assess behavior, skills, and attitudes is exceedingly tricky. Most educators have substantial experience in developing knowledge tests, but those sorts of assessment instruments are far easier to create than the other three types. Unless you have training and experience in the development of assessment instruments, it makes much more sense to use existing ones.

A set of assessment instruments designed to evaluate HIV education programs for students in grades 5-7 and 7-12 is provided in other booklets contained in this handbook. Start first by carefully considering whether some of these assessment devices will meet your needs; if not, you may need to create or adapt your own instruments. Because the development of acceptable assessment instruments for HIV education evaluation is extremely difficult, however, try to enlist the assistance of experienced test-developers to ensure the quality of your instruments.

Securing permission to gather data

Asking students questions about their sexual activities is considerably different from asking them about the Civil War. Because sexual activity is the most common way HIV is transmitted, your assessment instruments will often contain questions about students' sexual behaviors. It is essential that you clear your intended assessment instruments with appropriate school-district authorities. A special review group consisting of educators, parents, and other citizens will often have been established to judge the acceptability of HIV education materials and data-gathering instruments. The assessment instruments included in this handbook should be cleared by a local review group.

You should follow established district procedures in the use of assessment instruments dealing with sensitive subjects such as sexual conduct or drug use. Some districts require that either *active informed consent* or *passive informed consent* be secured from parents of students prior to the administration of such assessment devices. With active informed consent, a letter is sent to a student's parents or guardians describing the general nature of the intended data-gathering and asking permission for the student to complete the assessment instruments described. This letter must be signed by parents or guardians indicating their permission to have the data-gathering instruments administered to the student. With passive informed consent, a similar descriptive letter is sent to the student's parents or guardians. They are required to sign and

return it, however, only if they do *not* wish the student to complete the assessment instruments. Obviously, because active informed consent requires the receipt of a signed authorization letter from parents, it is the more difficult to implement. Most school districts already have policies in place regarding whether active or passive informed consent is required for data gathering.

The sorts of assessment instruments that might offend local citizens varies greatly among communities. This is an opportunity for you to play a significant educational role with local officials. If fears of citizen disapproval lead to the elimination of questions dealing with key HIV-risk behaviors (such as sexual behavior), you will be unable to discern whether the HIV education program is accomplishing some of the outcomes that it ought to accomplish. You may need to apprise local officials of the deadly threat to students engaging in HIV-risk behaviors and of the consequent peril to those students if educational programs to reduce HIV-risk behaviors are ineffective. When local officials, parents, and guardians are made aware of this serious potential, they will usually allow reasonable questioning about high-risk behaviors.

Confidentiality considerations

Once you have secured approval to administer suitable assessment instruments as part of your HIV education evaluation, you must structure the data gathering to increase the likelihood of getting truthful responses from students. To promote this, you should employ as many procedures as possible to enhance anonymity. Any such procedures should be announced to students in advance to assure them you do not intend to associate them with their responses.

Students should complete all instruments anonymously. Moreover, to remove the possibility that an individual's handwriting can be recognized, students should not be asked to write any words on the instruments. Instead, have students use only checkmarks or similar sorts of responses to all items. Similar, nonidentifiable pencils or pens should be used by everyone. In addition, students should place their own completed instruments in large envelopes or opaque containers that help avoid identifying the respondent. If possible, arrange seating to make it difficult for students to see each other's answers. (Several of the evaluation instruments in this handbook employ a response scheme specifically designed to prevent students from "inadvertently" seeing how others respond to

items dealing with sensitive subjects such as sexual behavior or the use of illegal drugs.)

In short, make sure that you have taken all reasonable steps to assure students of confidentiality and anonymity. Even then, of course, not all students will respond honestly to all questions. If, however, your efforts to ensure confidentiality boost the number of candid responses, your interpretations of the resulting data will obviously be more accurate. Fortunately, your quest is to evaluate program effectiveness rather than the status of individual students.

What about qualitative data?

So far we have dealt with the sorts of data gathered via fairly traditional quantitatively based assessment instruments. There are also a number of more qualitatively oriented data-gathering procedures, such as focus group interviews or one-on-one interviews with students who have received an HIV education program. These types of procedures often provide a rich source of explanatory evidence to help decision makers better understand the nature of the evidence you supply to them. For example, a few focus group sessions with students who have completed an HIV education program can prove particularly illuminating if the evaluator is trying to figure out which parts of the program worked well and which parts did not.

Final thoughts about Guideline 2

It is difficult to say that one guideline is more important than another, for all guidelines should play pivotal roles in your evaluation of an HIV education program. Guideline 2, however, leads directly to the assembly of the chief evidence you will use. To fail to identify appropriate assessment instrumentation, therefore, is to lose the whole evaluative ball game.

Few test developers are skilled enough to craft instruments that tease out subtle nuances in students' attitudes or garner honest answers to sensitive questions about sexual activities. The assessment instruments provided in this handbook were developed and field-tested by measurement experts and reviewed by specialists in the field of HIV prevention. You should review these instruments to see if they suit your needs. You should also consider the usefulness of qualitative data-gathering approaches, because schemes such as focus group interviews provide evidence that blends well with more quantitatively oriented evidence.

In addition to selecting appropriate assessment instruments, attention must be paid to obtaining needed permissions and enhancing student anonymity. Be sure that you attend to all three of the elements in this critically important guideline.

Guideline 3: Use a data-gathering design consistent with the orientation of the evaluation.

Once you have identified the assessment instruments you will use in your evaluation study, you must next determine your data-gathering design. Putting it more simply, you must decide how and when to administer the assessment instruments.

To keep these guidelines simple, we will consider one data-gathering strategy for program-improvement evaluation studies and one for program-continuation studies. If you want to explore other options, you can find a wide array of choices in almost any textbook on research methods for the behavioral sciences.

A data-gathering design for program-improvement evaluations

Choose a data-gathering design.

Let's assume you are carrying out a program-improvement evaluation of a district-level HIV education program. The chief decision makers involved are the teachers and curriculum specialists who planned and implemented the program. You must secure evidence to help these decision makers make their program more effective. As an evaluator, you are not trying to *prove* that the HIV education program works. Rather, you intend to provide your colleagues with data-based insights to help them *improve* their program. Your choice of a data-gathering design, then, should be consistent with that orientation.

The recommended data-gathering design for program-improvement evaluations of HIV education programs, presented in Figure 1, is known as the *one-group pretest-posttest design*. As seen

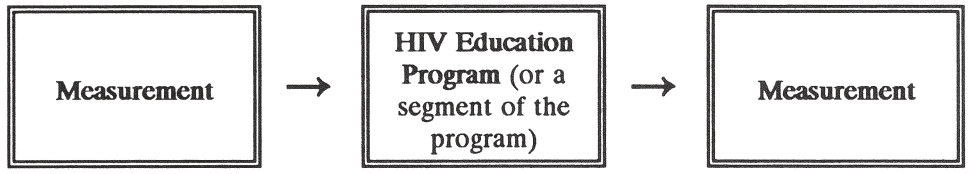

Figure 1. *A one-group pretest-posttest design*

in Figure 1, this design involves a preprogram measurement and a postprogram measurement. If one of your instruments is an anonymous questionnaire regarding students' HIV-risk behaviors, for example, you would administer that questionnaire to students before and after the program. Differences between the pretest and the posttest data would be credited to the program's effects.

The HIV education program, of course, is not the only possible reason for a change between students' pretest and posttest questionnaire responses. As students grow older, increased maturity may alter their approach to HIV-risk situations. Similarly, if they discovered that one of their classmates is infected with HIV, it will have a tremendous impact on their responses. These events, unrelated to the program, can pose interpretive problems for program-continuation evaluators, who must often prove a program's effectiveness to incredulous decision makers and must, therefore, use data-gathering designs that control for such factors. The program-improvement evaluator, however, usually has no such constraints and often needs only to point out that extraneous factors may have influenced the results.

You will note in Figure 1 that the pretest and posttest measurements may be used not only with the HIV education program in its entirety, but also with segments of the program. Suppose an HIV education program devoted three class periods to promoting students' refusal skills in situations that might involve high-risk sexual activity. If the program's staff were eager to improve this segment of the program, you could gather presegment and postsegment evidence from students to see if the three-day treatment of refusal skills led to increases in students' measured ability to apply those skills. If the presegment-to-postsegment gains were as the staff hoped, the program would not need modifying. On the other hand, if the presegment-to-postsegment gains were too small or nonexistent, alterations would be in order.

Here is a more detailed illustration. You are assigned to evaluate a school district's HIV education program for improvement purposes. Although the program has been in place for several years, the district's school board has asked administrators to ensure that the program is as effective as possible. Your job is to help teachers identify the parts of the program in need of revision.

You meet with the district's HIV education teachers and agree on five assessment instruments consistent with the program's stated objectives. The five instruments are: (1) an HIV knowledge test, (2) a test of students' refusal skills, (3) an attitudinal inventory assessing students' perceptions of their vulnerability to HIV infec-

tion, (4) an attitudinal inventory reflecting students' belief that they can take actions to reduce their likelihood of HIV infection, and (5) a questionnaire regarding the extent to which students engage in HIV-risk behaviors.

The district's HIV education program consists of fifteen hours of HIV-specific instruction during a required tenth grade health education class. You administer the five assessment instruments before and after the classes and discover that students display substantial progress on the knowledge and skill instruments but almost no change on the behavioral questionnaire, your most important instrument, or on the two attitudinal inventories. Such results would place you in a position to suggest that program alterations are warranted. Because the promotion of students' skill and knowledge appears to be successful, you might suggest that parts of the program be modified to better address the two attitudinal dimensions (students' perceived vulnerability and self-efficacy), and their behavior. If you are familiar with instructional psychology, you might suggest particular modifications in the instructional procedures used by the teachers. If you do not possess such knowledge, you could suggest that the HIV education staff rethink the dimensions on which little student progress is evident. You might also, at this point, seek qualitative data from student interviews—individual or focus group sessions—about which program components the students thought did or did not work.

One disadvantage of this design, as we have discussed, is the possibility that factors other than the HIV education program have influenced students' pretest-versus-posttest responses. You will have to be attentive to such possibilities. If other events, such as the release of a popular film about AIDS, occur during the period that the HIV education program took place, you will need to describe them in your report.

Another potential disadvantage of this data-gathering design stems from the use of the same assessment instruments before and after the program. The use of a pretest may result in a *reactive* effect by alerting students to what they are expected to get out of the program. Students may react differently to the program than they would have merely because the pretest let them know "what's important" in the program. If you are considering assessment instruments you fear would be reactive, you may wish to consider alternative data-gathering approaches such as those described in the additional readings at the end of this booklet.

A data-gathering design for program-continuation evaluations

The initial consideration in selecting a data-gathering design for program-continuation evaluations of HIV education is the confidence with which you can supply convincing evidence about the program's effectiveness. Although a data-gathering scheme such as the one-group pretest-posttest design might prove satisfactory for program-improvement purposes, it does not fill the needs of program-continuation evaluators wishing to supply evidence about whether a program really worked. You need a data-gathering design that allows you to make defensible statements about an HIV education program's success—or lack of it. And because the evaluation of school-based HIV education programs must take place in the midst of ongoing education programs, a data-gathering design must be selected that can be realistically implemented in a school setting.

The *pretest-posttest two-group design*, portrayed in Figure 2, provides the strongest basis for a program-continuation data-collection scheme to address these considerations. This design involves two groups, with only Group 1 initially receiving HIV education. Group 2 begins as an untreated control group.* This data-gathering design requires that a preprogram measurement be given to both groups. After Group 1 has completed the HIV edu-

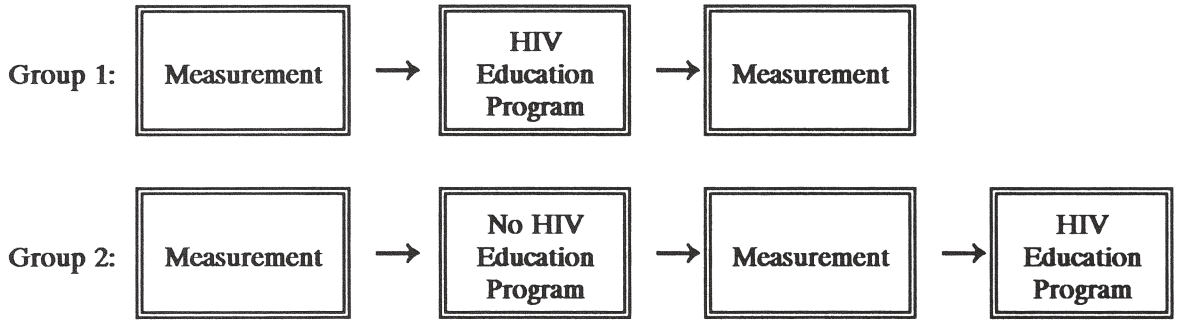

Figure 2. *A pretest-posttest two-group design*

*If Group 2 is not receiving any HIV instruction, it is termed the "control group." Sometimes, however, Group 2 is receiving a different intervention (perhaps an earlier version). Group 2 is then called the "comparison group."

cation program, both groups are posttested. Because an effective HIV education program will provide students with content that can quite literally save their lives, the prospect of employing a data-gathering design in which an "untreated control group" of students receives no HIV education runs counter to our sense of educational responsibility. Therefore, enough time must be set aside during the school year to insure that Group 2 also receives the HIV education program after the posttest.

The key comparisons in this two-group design are those between the pretest-to-posttest changes made in Group 1 (the treated group) and those made in Group 2 (the untreated group). If Group 1 outperforms Group 2 on the posttest, it would indicate that the program is effective. If there is no difference between the two groups' pretest-to-posttest changes, or if Group 2 outperforms Group 1, a lack of program effectiveness is indicated.

Interpretations of the effectiveness of the HIV education program, however, are totally dependent on the degree to which students in the two groups are similar. If the groups are essentially the same, you can draw meaningful conclusions as to whether the HIV education program worked. As the two groups become less similar, the conclusions to be drawn become less meaningful. For example, one of the concerns with classroom-based evaluations is that students in one classroom are different from students in another classroom. One reason for this is that students may be assigned to particular classes on the basis of their ability or interests. When classroom assignments are not made randomly, it is impossible to assume that students within those classrooms will be similar. Therefore, if the two groups (treated and untreated) are composed of only two different classrooms, it is nearly impossible to determine whether posttest differences are due to the intervention program or to differences among students in the individual classrooms.

One solution to this problem is to increase the number of classrooms to at least two per intervention and two per control group. The more classrooms that can be included per group, and the more randomly those classrooms can be selected from all possible classrooms, the more likely it is that the students in the intervention and control groups will be equivilant at the pretest. If a large number of classrooms (e.g., 20) can be randomly selected from the school district and randomly assigned to treatment or control situations, then a "posttest only" design may be used, in which only differences between posttest scores are examined (because we feel confident that students' scores were equivilant to begin with).

Of course, it is not always possible to study many classrooms at one time, or to select classrooms randomly from the school district. In that case, it is important to use pretest scores in the analysis to control for the potential lack of confidence about initial equality. This more common situation is analyzed in a "pretest-posttest (nonequivilant) two-group design," which is shown in Figure 2.

It may also be important to consider that the intervention could have different effects on different types of students. Age, gender, or ethnicity, for example, may be key indicators of a student's receptivity to some or all intervention components. Therefore, it may be important to analyze the results on the basis of key student characteristics. This is a somewhat more complicated design. It will require a bit more work to set up, and will need more students than the simpler designs described above. Data analysis could also be more complex. However, if particular student characteristics are responsible for different reactions to the intervention program, then it would be well worth the effort to examine those differences in the analysis which will, hopefully, lead to more meaningful results. Finally, the location of the schools within a given district may also have an effect on the results for a number of reasons. Therefore, it may be desirable to try to match classrooms from schools in similar neighborhoods, and/or with similar student populations and then randomly assign one of each pair to the treatment and the other to the control condition.

Sampling

Whenever possible, several schools should be randomly sampled from the school district for inclusion in the study. Random sampling can be as simple as pulling school names from a hat containing all school names, and then randomly selecting one or two classes from each school. These classes can be randomly assigned to treatment or control conditions by the flip of a coin.

If you are interested in matching schools on a key set of characteristics, the school district office may have relevant information on school location and student composition. You may then want to group all district schools into different types, such as urban versus suburban, and then randomly sample from within each group. Preselecting groups of schools from which to draw your random sample is known as stratified sampling. These and other sampling procedures are described in most standard research-oriented textbooks.

Final thoughts about Guideline 3

There are many more data-gathering strategies than the two basic models presented here. In the evaluation of HIV education programs, however, you will find that these two designs will satisfy almost all of your data-gathering requirements.

The one-group pretest-posttest design is recommended for program-improvement evaluations. A two-group variation of that design is recommended for program-continuation evaluations. Although it is certainly possible to use a one-group design in program-continuation evaluations, its results will not be as convincing as if a control group were used. It is equally possible to use a control-group design in program-improvement evaluations. You may find, however, that control groups often add needless complications to an evaluation focusing on program improvement.

Guideline 4: Use data-analysis procedures that yield understandable results.

Once you have gathered your data, that evidence must be summarized in such a way that is understandable to decision makers.

Practical versus statistical significance

Analyze the evaluation study's data.

Evaluators sometimes carry out data-analysis procedures that produce enough statistics to be "respectable." Such evaluators, however, must remember their audience. Unfortunately, statistical procedures that are among educational *research's* most useful tools are sometimes inappropriate for educational *evaluation*. In general, the audience for an educational researcher's efforts consists of other researchers or scholars to whom subtle, statistically significant differences may be quite important. The audience for evidence gathered by evaluators of HIV education, however, will most often be teachers, board members, or educational administrators. By and large, such decision makers are concerned with practical rather than statistical significance. A practically significant question might focus on whether a program's effect is large enough to warrant actions such as expanding the program's applications to other settings. In some cases, sophisticated statistical analyses can render an evaluation study's results virtually incomprehensible.

Comprehensibility of results

As an HIV education evaluator, you will need to analyze data in the manner most appropriate to yield easily understandable results for decision makers. This usually leads to analyses involving easy-to-read indices such as percentages and arithmetic averages or easily understood data-representation schemes such as bar graphs. In recent years, most people have become familiar with news reports of surveys having an error margin of plus or minus a certain percent. If you can analyze your data so that the results can be cast in a form accompanied by a given error percentage, most decision makers will intuitively understand what you are reporting.

If more sophisticated analysis approaches are used, make sure that results can be easily communicated to decision makers. For example, analysis of covariance is a statistical procedure often used to account for initial differences between groups of students. Were you to employ this data-analysis technique, your report to decision makers could be something along these lines: "After statistical adjustments were made for the fact that the two groups were not initially equal, the HIV education group had 13 percent fewer reported incidents of unprotected sexual intercourse."

Suppose that, prior to an HIV education program, 35 of 100 students reported that they routinely had sexual intercourse without using a condom whereas several months after the program's conclusion only 28 of 100 reported such behavior. In other words, there was 20 percent reduction in sexual intercourse without a condom among students who engaged in such a behavior. These sorts of percentage-based results are easy for decision makers to interpret. People can make sense of percentage-based differences between students' preprogram and postprogram performances because people are used to dealing with percentages in other aspects of life. Most people are not used to dealing with statistically significant differences at the .05 versus .01 probability levels.

Percentage-correct may not be a suitable descriptive scheme for all assessment instruments you choose. For example, you might use a ten-item attitudinal inventory focusing on students' perceived ability to use refusal skills that yields scores from 10 points (low perceived ability) to 50 points (high perceived ability). For such an instrument, an arithmetic average of students' scores would be more sensible than results expressed as percentages.

Because you will typically be looking at preprogram and postprogram data for your evaluations, it will be a routine matter to compare the differences between such data to discern whether the

HIV education program yielded its anticipated effects. Simple pretest-to-posttest percentage changes will usually fill the data-analysis bill satisfactorily.

Final thoughts about Guideline 4

This fourth guideline stresses the desirability of using data-analysis schemes that yield understandable results. You will discover in most instances that simple statistical procedures will take care of your data-analysis needs. In those few cases when you might need more sophisticated statistical analyses, you may wish to call on a statistical consultant to provide you with additional data-analysis guidance. Such situations might arise when it is unclear whether a difference in the performances of treated and untreated students is large enough to be meaningful.

One reason that Guideline 4 is included in this set of suggestions for HIV education evaluators is to dissuade you from believing you must carry out all sorts of complicated data analyses to make your evaluation study respectable. This is simply not the case. Your task as an evaluator of HIV education programs is to help the program's decision makers come up with better decisions. To be useful to busy decision makers, data-analysis procedures should lead to straightforward, readily interpretable information regarding program effectiveness.

Guideline 5: Report your results using a multilevel reporting scheme featuring written and oral reports.

If you design and carry out your evaluation study following the first four guidelines, you will have an intellectually manageable set of evidence—primarily student pretest and posttest data—bearing on a modest number of important program-relevant decisions. Your task at reporting time is to present that evidence to decision makers in a form most likely to influence the decisions they need to make.

Report the evaluation study's results.

An appropriate level of detail

Before reporting your evaluation study's results, you will typically find yourself in a dilemma over the suitable level of detail to include. To report concise results for busy decision makers, you would invariably need to leave out important information about such matters as the specific procedures used to assure student

anonymity. On the other hand, if you chose to report the evaluation's procedures in full detail, the resulting report would often be so lengthy that decision makers would be put off by its size.

Fortunately, there is a way out of this bind: prepare two written reports. The first report should be very brief (no more than a few pages in length) and should hit only the high points—namely, the evidence that bears most directly on the decisions at issue. This brief *executive report* should direct readers who wish more information to a *technical report* that describes the evaluation study's procedures and results in greater detail. Even with the more technical report, however, you must employ good sense regarding the level of detail acceptable to the decision makers you are attempting to serve. Too often, evaluators become caught up in the intricacies of their evaluation study's procedural nuances and tend to create excessively lengthy reports. Evaluation reports perceived as hyper-detailed will rarely be read by anyone except the evaluators who prepared them. Thus, even the technical report should be sufficiently succinct and focused so that decision makers will be inclined to use it.

In any evaluation report, try to use visual and/or graphic methods to make the results as palatable to readers as possible. Few decision makers relish the prospect of reading even three pages of single-spaced prose. Although it may be difficult, particularly in the executive report, use white space and graphic presentation techniques that stimulate the reader's interest.

Oral reporting

Increasingly, educational evaluators are being asked to supplement written evaluation reports with oral presentations to, for example, a district's school board or the teachers staffing the district's HIV education program. Such sessions provide you with an excellent opportunity to educate decision makers about the impact of your study's results on the decisions they face. The give-and-take discussion that often follows an evaluator's oral report is a wonderful forum for such educative efforts.

Be sure to devote enough preparation time to make your oral reports polished, professional, and decision relevant. If you are only asked to give a written report, encourage decision makers to allow a brief oral presentation highlighting the study's key results.

Making recommendations

Another issue you are apt to face when you make your report is whether to offer recommendations to decision makers. Suppose the pretest-posttest results of a program-improvement evaluation study regarding a one-week unit dealing with refusal skills indicate that the unit was particularly ineffective. Students' skills after the unit are essentially the same as before the unit. A logical recommendation would be that the unit be seriously overhauled. But should you make such a recommendation? Similarly, if your program-continuation evaluation study indicates that an HIV education program is having a decisively beneficial impact on reducing students' HIV-risk behaviors, should you recommend that the program be continued?

Evaluation specialists are divided over this issue. For some specialists, making decision-related recommendations is a logical extension of the evaluator's decision-facilitation role. Other specialists, however, regard evaluator-generated recommendations as intrusions on the decision maker's prerogatives. These individuals believe that the evaluator should supply evidence only and should offer no guidance regarding program decisions.

It is suggested, therefore, that you be guided by decision makers' expressed preferences. You will doubtless have met with decision makers during the early stages of designing your study, for you clearly need to find out what their important decision points really are. At that stage of the process, you can easily learn whether decision makers wish your report to include recommendations.

If you present an oral report, your recommendations will often be solicited even though you may have been directed to avoid such recommendations in writing. Be prepared to respond to such requests.

Final thoughts about Guideline 5

This final guideline may appear to involve substantial effort. After all, not only are two written reports to be authored, but an oral report is to be made as well. Any effort associated with reporting an evaluation study's results, however, will usually be well worth it. What good does it do to design and carry out a first-rate evaluation study if the results make little impact on the decision makers for whom it was originally conducted?

Reporting an evaluation's results should not be an afterthought. From the earliest days of the evaluation study, you should

continually think about how the study's results can be most effectively communicated.

Although Guideline 5 does not directly address the topic of making recommendations regarding the decisions at issue, you will find that if decision makers request your recommendations, they will typically be influenced by your views. If you offer recommendations without being asked, however, your advice may be seen as presumptuous and may be rejected. Be guided by the decision makers' preferences.

Conclusion

These five guidelines are important to consider in designing and conducting your HIV education evaluation. They will also provide a set of criteria to use in deciding whether your planned evaluation of an HIV education program is likely to be successful.

Think of these guidelines as procedural decision-points. Although you will need to make other choices as an evaluator, these guidelines can function as a framework for the procedural steps you will follow as the evaluation occurs.

As stated at the outset, this treatment of educational evaluation is decidedly modest. By consulting the references listed in the Additional Reading section, you can achieve further insight into evaluative topics.

Five guidelines cannot transform a novice evaluator into an expert. Nonetheless, if you follow this booklet's guidelines when evaluating HIV education programs, you can be confident that your resulting evaluations will be superior to evaluations departing dramatically from the guidelines. In view of the threat represented by HIV and the certainty that more effective HIV education programs will help students avoid HIV infection, improvements in the evaluation of HIV education will be well worth the effort expended.

Additional Reading

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1985). *Standards for educational and psychological tests.* Washington, DC: American Psychological Association.

Anderson, L. W. (1981). *Assessing affective characteristics in the schools.* Boston: Allyn and Bacon.

Berk, R. A. (Ed.). (1984). *A guide to criterion-referenced test construction.* Baltimore: Johns Hopkins University Press.

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand McNally.

Coyle, S. L., Boruch, R. F., & Turner, C. F. (1991). *Evaluating AIDS prevention programs.* Washington, DC: National Academy Press.

Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C., Walker, D. F., & Weiner, S. S. (1980). *Toward reform of program evaluation.* San Francisco: Jossey-Bass.

Fetterman, D. M., & Pitman, M. A. (Eds.). (1986). *Educational evaluation: Ethnography in theory, practice, and politics.* Beverly Hills, CA: Sage.

Green, L. W., & Lewis, F. M. (1986). *Measurement and evaluation in health education and health promotion.* Palo Alto, CA: Mayfield.

Joint Committee on Standards for Educational Evaluation. (1981). *Standards for evaluations of educational programs, projects, and materials.* New York: McGraw-Hill.

Levin, H. M. (1983). *Cost-effectiveness: A primer* (New Perspectives in Evaluation, Vol. 4). Beverly Hills, CA: Sage.

Linn, R. L. (Ed.). (1989). *Educational measurement.* New York: Macmillan.

McLaughlin, M. W., & Phillips, D. C. (Eds.). (1991). *Evaluation and education: At quarter century.* Chicago: National Society for Study of Education.

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage.

Popham, W. J. (1990). *Modern educational measurement: A practitioner's perspective* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Popham, W. J. (1992). *Educational evaluation* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Popham, W. J., & Sirotnik, K. A. (1992). *Understanding statistics in education* (3rd ed.). Itasch, IL: F. E. Peacock.

Rugg, D. L., O'Reilly, K. R., & Galavotti, C. (1990). AIDS prevention evaluation: Conceptual and methodological issues. *Evaluation and Program Planning, 13,* (pp. 79-89). Atlanta, GA: Division of Sexually Transmitted Diseases/HIV Prevention, Centers for Disease Control.

Stewart, D. W., & Shamdasani, P. N. (1990). *Focus groups: Theory and practice* (Vol. 20). Newbury Park, CA: Sage.

The U. S. Conference of Mayors. (December 1990). Evaluation of HIV/AIDS prevention programs: Gathering evidence to demonstrate results. *Technical Assistance Reports*. Washington, DC: U. S. Conference of Mayors.

Webb, E. J., Campbell, D. T., Schwartz, R. D., Sechrest, L., & Grove, J. B. (1981). *Nonreactive measures in the social sciences* (2nd ed.). Dallas: Houghton Mifflin.

Worthen, B. R., & Sanders, J. R. (1987). *Educational evaluation: Alternative approaches and practical guidelines.* New York: Longman.