DOCUMENT RESUME

ED 362 564 TM 020 619

TITLE	Program Assessment: A Six-Step Process to Curriculum Improvement.
INSTITUTION	Montana State Dept. of Public Instruction, Helena.
SPONS AGENCY	Centers for Disease Control (DHHS/PHS), Atlanta, GA.
PUB DATE	[93]
CONTRACT	U63/CCU803049-04
NOTE	22p.
PUB TYPE	Guides - Non-Classroom Use (055)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Accreditation (Institutions); Acquired Immune Deficiency Syndrome; *Curriculum Development; Curriculum Evaluation; Data Analysis; Data Collection; *Educational Assessment; Educational Cooperation; Educational Improvement; *Educational Objectives; Educational Planning; Elementary Secondary Education; Evaluation Methods; *Formative Evaluation; *Guidelines; School Districts; State Programs; *Summative Evaluation; Test Construction
IDENTIFIERS	*Montana

ABSTRACT

The Montana school accreditation requirements require that school districts begin a curriculum development process in 1991. A plan for student assessment is required to follow curriculum development in each program area. The guidelines in this document should facilitate the cooperative efforts of teachers, curriculum departments, administrators, and school committees of parents and community members. These guidelines provide a simple format to assess a variety of programs in a planned and orderly manner. This document was revised from the publication "Evaluating HIV Education Programs" by the Centers for Disease Control, generalizing guidelines to all program areas. The following six steps for program assessment are highlighted: (1) determine whether the evaluation is to be formative or summative; (2) focus on a manageable number of important program-related goals; (3) select or construct suitable assessment instruments; (4) use a data-gathering design consistent with the orientation of the evaluation; (5) use data-analysis procedures that yield understandable results; and (6) report and evaluate results to make recommendations and modify program as indicated. (SLD)

nt

x-step process

urriculum

rovement

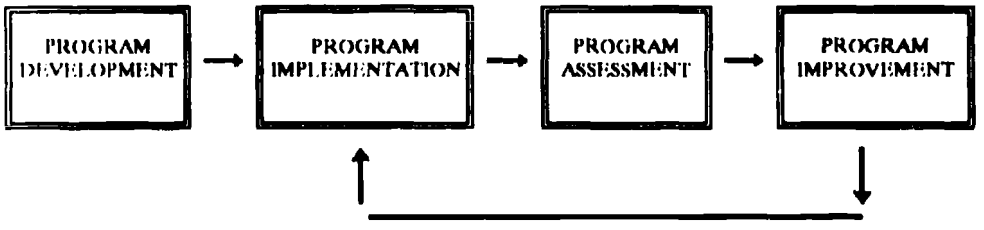
Foreword

The Montana school accreditation requirements as outlined in ontana School **Accreditation Standards and Procedures Manual**, and y the Board of Public Education, require that school districts be ılum development process in 1991. The standards further requi later than the school year immediately following the complet ritten sequential curricula in a subject area, the school shall begin opment of an assessment process for a subject area." School distric establish curriculum and assessment development processes as a coop ive effort of teachers, administrators, students, parents and communi members. In addition, curricula must be reviewed at intervals not exceeding five years. Therefore, the assessment requirement of rule 10.55.603 is twofold: a plan for student assessment must follow curriculum development in each program area; and in addition to continual program assessment, the curriculum must be formally reviewed at least every five years. The ultimate purpose of both student assessment and program assessment is to improve student achievement and success.

These guidelines should facilitate the cooperative effort of classroom teachers, curriculum departments, administrative personnel, and school committees that include parents and community members. They provide a simple format to assess a variety of programs in a planned, orderly manner. They are written with the assumption that the reader is not a trained evaluator and has limited, if any, experience in conducting formal evaluations.

This document was revised from the publication **Evaluating HIV Education Programs** by the Centers for Disease Control, Atlanta, Georgia. To generalize these guidelines for use in all program areas, modifications were made by the Office of Public Instruction with the assistance of Alex McNeill, Chair, Health and Human Development Department, Montana State University, David Puyear, Director, Golden Triangle Curriculum Cooperative, Robert Briggs, Science Specialist, Jan Cladouhos Hahn, Language Arts Specialist and Spencer Sartorius, Administrator, Health Enhancement Division, Office of Public Instruction.

A Curriculum Development and Assessment Process



Introduction

Assessment serves functions that transcend the mandate of school accreditation by helping those involved in the decision-making process improve instruction and enhance student success. The process described in this booklet is designed to help local districts with an assessment plan based on their unique programs. Assessment is an ongoing process to continually look toward program improvement. Program assessment points out strengths and weaknesses on which program modifications can be based.

As an analogy, assessment could be compared to owning a car. After the original selection of a vehicle (curriculum or program), you are continually assessing whether or not this vehicle (program) meets your needs and measures up to your identified criteria. With a car, you are listening to the engine, figuring gas mileage, assessing comfort. With a program, you are administering tests, collecting student work, asking questions. These are formative assessments.

Depending on the assessment results, you may need to perform some basic maintenance, to "tune up" the vehicle, or to upgrade or add components such as CD player, exhaust system, towing package--or for a program, computer software, print materials, lab equipment.

Suppose you have decided that every five years you will consider purchasing a new vehicle, much like a curriculum review cycle of five years. The decision to either keep the old or to select new requires a summative evaluation. The tools of a summative evaluation may be taken more seriously. To check the national norms, you may consult a consumer magazine's ratings. You may want the opinion of an expert mechanic, other drivers, and a car dealer--and you will undoubtedly focus on a few important points like the engine and safety. The gap between what you own and what you need may require a total renovation (new engine, paint job, seat replacement) or because you need all-wheel drive, an anti-lock braking system, and air bags, you may need a new car. Each program within your school's curriculum deserves no less attention and involves a similar process. If assessment shows that the program is not meeting the needs of your students within the first years of its implementation, adjustments are necessary. If, at the end of a five-year review cycle, student success cannot be documented, you may need a new program.

Guidelines for Assessing Education Programs

Program assessment can follow the step-by-step process described in the guidelines within this manual. As is common with such sequences, the guidelines don't always work in the exact order suggested. You will sometimes find that you may need to skip a step or repeat some steps more than once along the way. The guidelines represented in Figure 1 can function as a framework for the procedural steps you will follow as the assessment occurs.

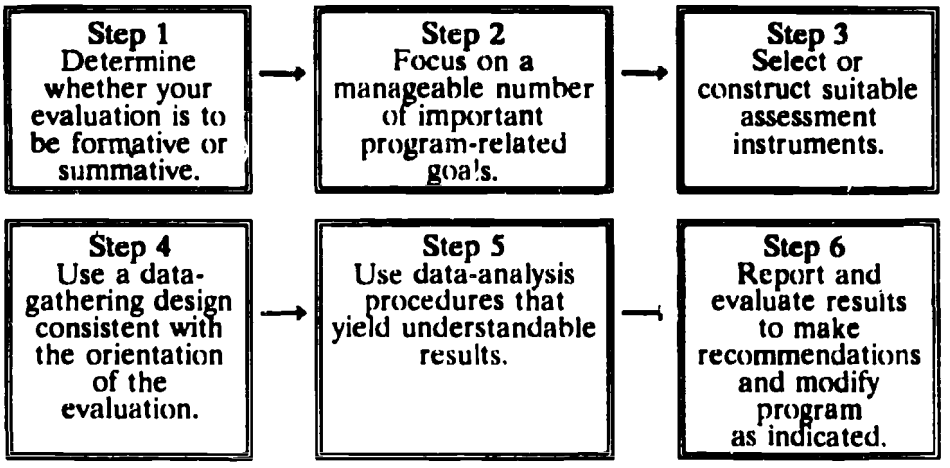


Figure 1: A sequential framework for assessing education programs.

Guideline 1: Determining the Assessment Study's Chief Function

Guideline 1: Determine whether your evaluation is to be formative or summative.

An educational program is evaluated for one fundamental reason: to provide information to help individuals make better decisions. The kinds of decisions that must be made concerning a 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. The evaluator's responsibility, then, is to gather information appropriate to the possible consequences of the decision.

Two evaluative approaches

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 or not to continue the program or the use of existing curriculum in its current form. These decisions are *program continuation* decisions.

The type of decisions needed determines the type of information you seek and the approach you will take in your evaluation. We will refer to these two evaluative approaches as shown:

Focus of Study	Type of Evaluation
Program Improvement	Formative Evaluation
Program Continuation	Summative Evaluation

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

Since core subjects, required by the accreditation standards, necessitate program improvement decisions, not continuation decisions, your evaluation will generally be formative in nature. In general, the interest for teachers is in formative data, for board members in summative data, and for administrators, both types. The possibility of moving to a radically new curriculum (from skills-based to whole language, for example) or the implementation of a program "beyond" the requirements of the standards may call for summative evaluation.

When carrying out a *summative* evaluation, you must be completely objective and nonpartisan. Your evidence will decide whether to continue or discontinue the program. Usually, summative evaluations are made after a program has been in place for a few years when it is appropriate to determine if the program is worth its time requirements and expense.

Final thoughts about Guideline 1

Although Guideline 1 appears to be simple, it will have a profound impact on your behavior during the assessment process. Regardless of whether your evaluation is dominantly summative or formative, what you choose to do, how you do it, and how you communicate what you have done--should be *decision-focused*.

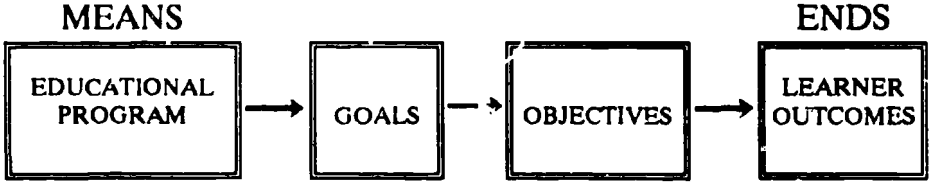
Guideline 2: Focusing on a Reasonable Number of Goals

Guideline 2: Focus on a manageable number of important program-related goals.

Educational programs in Montana must embody elements mandated by the **Montana School Accreditation Standards.** The programs must reflect the goals identified in *Sub Chapter 10, Program Area Standards.* Each goal has a series of objectives which, if achieved, will result in desired learner outcomes. Regardless of whether you pursue a formative or summative evaluation, one of your early tasks is to focus on a manageable number of goals related to the program. Remember, the purpose of an evaluation is to help make decisions that will improve your program. Because you will be trying to address only a modest number of program-relevant decisions, you will clearly need to focus on genuinely *important* goals.

The primary targets: program objectives

Teachers usually aspire to bring about worthwhile changes in students. Those changes can focus on altering either students' behaviors or the factors that contribute to such behaviors. Put most simply, an instructional objective for a program should describe the post-program knowledge, skills, attitudes, or critical thinking that the program seeks to promote. This is nothing more than a classic ends/means distinction, as illustrated below:



Identifying a program's objectives can lead to the identification of the decisions on which you will focus your assessment.

A NUMBER OF EDUCATORS ATTEMPT TO DESCRIBE INSTRUCTIONAL OBJECTIVES IN TERMS OF WHAT THE PROGRAM ITSELF WILL DO RATHER THAN WHAT IT IS INTENDED TO *ACCOMPLISH.* EDUCATIONAL OBJECTIVES HAVE NOTHING TO DO WITH WHAT THE EDUCATION PROGRAM IS OR HOW IT WAS CREATED. INSTEAD, THE OBJECTIVES FOR AN EDUCATION PROGRAM MUST FOCUS ON PROGRAM OUTCOMES, THAT IS, WHAT HAPPENS TO STUDENTS AS A CONSEQUENCE OF THE PROGRAM.

Because objectives reflect what the program intends to accomplish, the extent to which such objectives have been achieved can be helpful in determining the program's effectiveness. In order to make good evaluative use of a program objective, it should be stated in such a way that, at the end of the program, evidence can be gathered to determine if the objective has been achieved. Some evaluators refer to such objectives as *measurable* program objectives.

If you can identify the objectives that you hope to accomplish, and if you can define those objectives as pre-program to post-program changes in students, you will have gone a long way in clarifying the focus of your assessment.

Collect only information that focuses on program improvement.

Evaluators who wish to use a program's objectives to their advantage will need to be sure that the program is organized around only a handful of measurable objectives. Rarely permit your assessment, therefore, to be organized around more than a half-dozen or so objectives. (The staff may, of course, have a number of specific instructional objectives to use in day-to-day instruction.)

Gather decision-focused information. One good way to verify whether the evidence really bears on a program-related decision is to ask, "If the evidence turns out this way, what would my decision be?" Then, ask, "If the evidence turns out the opposite way, what would my decision be?"

THE EVALUATOR OF EDUCATION PROGRAMS MUST CONSTANTLY BE INFLUENCED BY THE QUESTION: "CAN THE PROGRAM BE IMPROVED IF I COLLECT THIS INFORMATION?" IF THERE'S A GOOD ANSWER TO THAT QUESTION, THE EVALUATOR SHOULD GATHER THE INFORMATION. IF THE ANSWER IS AMBIGUOUS, THE EVALUATOR SHOULD ABANDON THE QUEST FOR APPARENTLY IRRELEVANT INFORMATION.

Targets unrelated to program objectives

Although the decisions addressed by formative and summative evaluators are often linked to the achievement of a program's objectives, some choices do not depend on the attainment of objectives. Formative evaluators, for example, often gather evidence as to whether an instructional program is being delivered as intended. The decision at issue in this instance is whether changes in methodology must be made.

Other examples of decisions unrelated to objectives-attainment include (1) whether community officials will permit sensitive topics to be addressed in instructional activities, (2) whether students will regard information as more believable if provided by peers rather than teachers, and (3) whether the program's objectives are appropriate. There are also instances in which unforeseen effects of the program's objectives might be significant in judging a program's effectiveness.

In short, although the degree to which a 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 2

Collect data that will lead to appropriate and efficient decision making concerning educational programs.

Guideline 3: Securing and Using Assessment Devices

Guideline 3: Select or construct suitable assessment instruments.

As suggested earlier, the chief function of an evaluation is to assemble and make available evidence to consider when making a program-related decision. It should not be surprising, therefore, that choosing *which* information to assemble constitutes one of the most important chores. Guideline 3 deals with the instruments you will use to gather decision-relevant data.

One of the most important tasks is a careful analysis of the various forms of assessment currently available. The instruments should be valid representations of the standards students are expected to achieve. Multiple choice and standardized tests alone may be inadequate to measure many of the educational outcomes included in the 1989 Montana Accreditation Standards. Other forms of assessment that should be considered during this process are portfolios, open-ended questioning, extended reading and writing exercises, projects, exhibitions, attitudinal surveys, and skills tests. Instruments chosen should help both teachers and administrators make decisions that improve instruction and enhance student success either by assessing program segments or assessing total program effectiveness. Analytic, rather than holistic, scoring methods provide information useful for program assessment. For example, when the analysis of an oral presentation is broken into criteria for organization and delivery, evaluators can pinpoint weak areas in the speaking curriculum. The instruments should provide more than just numbers or ratings and should include information on particular abilities students have or have not developed. (See Matrix 1.)

MATRIX 1. DATA COLLECTION TECHNIQUES				
(Examples)	Content (Knowledge)	Skills (Appropriate)	Attitudes (Affect)	Thinking (Behaviors)
Tests and Quizzes	X	X		
Questionnaires	X		X	X
Personal Interviews			X	
Self-reports			X	X
Participant Interviews			X	
Observations of Participants		X		X
Observations of Behavior				X
Homework, Samples, Portfolios	X			
Oral Reports	X	X		
Labs/Problems	X	X		
Projects and Performances	X	X		X

The assessment process used to evaluate the curriculum should be multi-dimensional and collect data from students, teachers and administrators. Instruments chosen should be fair to all students: sensitive to cultural, racial, class and gender differences and to disabilities.

An emphasis on outcome data

Students supply the bulk of the data the evaluator typically gathers. One method of gathering such data might be for students to complete questionnaires, tests, or writing assignments. Because evaluators, in most cases, will be interested in the *changes* in student behavior, or thinking and reasoning skills that may contribute to changes in behavior, information will typically be collected from students before and after experience in a program or unit of a program.

Evidence regarding changes in student behavior can be described as *outcome data*. Outcome data represent the effects of an educational program. Evidence regarding the nature of the educational program itself, in contrast, is referred to as *process data*. An assessment in which the evaluator wants to determine whether an instructional program is being provided as intended is a typical situation in which process data are gathered. Checklists developed to systematically evaluate curricula, such as those available from the Office of Public Instruction, also generate process data. However, most evidence gathered in an evaluation is a form of outcome data. But what kinds of outcome data should be gathered?

Recommended categories of outcome data

There are four prominent types of outcome data that evaluators attempt to secure:

- Evidence of the extent to which students use *critical thinking* developed within the program to modify behaviors
- Evidence of students' ability to display key *skills* addressed by the education program
- Evidence of students' *attitudes* toward program goals
- Evidence of students' knowledge regarding the *content* and data included in the education program

Evidence Category	Examples
Critical Thinking	Ability to analyze a problem, to evaluate a situation, to behave accordingly
Skills	Ability to read, to conduct an experiment, to climb a rope
Attitude	Attitudes toward language diversity, environmental concerns, drug use
Content	Knowledge about literary devices, chemical properties, nutrition and fitness.

Table 1. Illustrations of Relevant Types of Evidence for Students

Data should be gathered for all four categories. Knowledge tests alone will not measure a student's attitude, nor will it measure how the new knowledge has influenced his/her critical thinking and resultant behavior. Ultimately, behavioral data may be the most important. The purpose of education is, after all, to provide the mechanisms through which behavioral change can be encouraged as a thoughtful, reasoned process.

Measuring critical thinking and behavior change can be very difficult. Some programs may not be long enough or specific behaviors may not be exhibited immediately. This does not mean a program is ineffective, but that behavior change over time should be followed through longitudinal studies.

Developing and selecting suitable assessment devices

Assessment instruments can either be developed locally, adapted from existing instruments, or secured from commercial test developers or educational resource centers and university libraries. Most educators have substantial experience in developing skills and content tests. Finding and/or developing acceptable assessment instruments for thinking and attitude are more difficult.

Paper and pencil tests

Standardized tests, which provide data that can be compared, are designed to sample what is common across typical curricula for a particular grade. As a result, there is never a perfect fit between the local objectives and those tested. Care must be taken to select a test that best matches your program goals and to use the sections relevant to your study. These scores are useful to see how well your student body can answer a specific set of questions as compared to a norming group or to some specified criterion associated with the subject matter being tested. Although basic skills and knowledge-level content are most commonly the targets of standardized tests, some do assess skills in critical thinking. If the test has not been re-normed within your targeted time period, comparisons over time can also be made. Using the Normal Curve Equivalents will allow you to compare results from different tests.

Teacher-made tests, although primarily instruments for student assessment, can also provide information for assessing a program. When developing a test, check that the curricular goals are clearly represented, that the most efficient type of question is chosen appropriate to the objective, and that a variety of cognitive levels of questions are utilized. Instructional targets and cognitive levels can be charted and then tallied to determine if the test items represent the curriculum fairly. (Such a test specification chart is available from the Northwest Regional Labs.) Teachers who have used a similar test over several years may be able to make a number of observations about the effectiveness of a program modification.

Use multiple assessment measures.

Surveys and questionnaires can be effectively used to assess attitudes, applications of skills, and curriculum implementation. A program assessment guide, such as the **Montana Assessment for Health Enhancement**, or similar questionnaires in other program areas, require that staff members answer questions about the goals and objectives, teaching strategies, materials, etc., as they evaluate curricular processes. Student surveys can be useful in determining student attitudes about a subject, materials used, technology, or whether skills learned are applied. The **Montana Youth Risk Behavior Survey** is an example.

Performance assessments can initiate program reviews. As developers design the criteria for scoring the performances, samples, or portfolios, goals and objectives must be scrutinized and achievement targets must be well understood, suggesting possible problems. Analytic scoring, in which categories such as organization, content, fluency, and conventions are scored, provide data about strengths and weaknesses in student skills and the program.

Personal communication provides more qualitatively oriented data-gathering procedures such as focus group interviews, one-on-one interviews with students who have completed a program, or conferences with students about their work. Focus group discussions with curriculum department staff often lead to useful information. These types of procedures often provide a rich source of anecdotal data that helps explain findings from quantitative assessments.

Gathering sensitive data

Some areas of the curriculum deal with socially and/or culturally sensitive subject matter. Asking questions about activities, especially in some sensitive areas, e.g., numan sexuality, environmental issues, or suicide, is much different from asking about the Civil War, sentence structure or parts of a plant. In virtually every case, you will need to clear your intended assessment instruments with appropriate school district authorities.

Follow established district procedures to review assessment instruments dealing with sensitive subjects such as sexual conduct or drug use. A tremendous diversity exists among districts regarding the sorts of assessment instruments that might offend local citizens. This is an opportunity for you to play a significant educational role with local officials.

Once you have secured approval to administer suitable assessment instruments, structure the data gathering to increase the likelihood of getting truthful responses from students. Employ as many procedures as possible to ensure anonymity.

Final thoughts about Guideline 3

It is difficult to say that one guideline is more important than another, for all guidelines should play pivotal roles in your assessment of an education program. Guideline 3, however, leads directly to the assembly of the chief evidence you will use. Using appropriate assessment instruments is crucial.

When possible, use existing assessment instruments that provide decision-focused information. Recognize, however, that knowledge tests are the most widespread form. Quality instruments designed to measure attitude, critical thinking, and the performance of skills are more difficult to develop or to find. Qualitative data-gathering approaches such as using personal communication, projects, or performances, provide evidence that complements quantitative data.

Guideline 4: Choosing a Data-gathering Design

Guideline 4: Use a data-gathering design consistent with the formative or summative orientation of the evaluation.

Once you have identified the assessment instruments you will use, you must next determine your data-gathering design. More simply, you must decide how and when to administer the assessment instruments or gather and record the assessment data.

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

A data-gathering design for formative evaluations

For a formative evaluation, you must secure evidence to help make the program more effective. As a formative evaluator, you are not trying to *prove* that the education program works. Rather, you intend to provide data-based insights to help improve the program. Your choice of data-gathering design, then, should be consistent with the formative orientation.

The recommended data-gathering design for formative evaluation of education programs, presented in Figure 2, is known as the *one-group, pretest-posttest design*. As seen in Figure 2, this data-gathering design involves a pre-program measurement and a post-program measurement. If one of your instruments is an anonymous questionnaire regarding student 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.

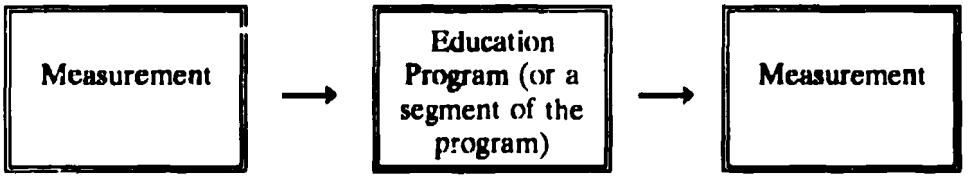


Figure 2. A data-gathering design for formative evaluation: *The one-group, pretest-posttest design*

You will note in Figure 2 that the pretest and posttest measurements may be used not only with the education program in its entirety, but also with segments of the program. Suppose a program devoted three class periods to promoting students' refusal skills in situations that might involve high-risk behaviors. If you wish to improve this segment of the program, you could gather pre-segment and post-segment evidence from students to see if the three-day treatment of refusal skills led to increases in their ability to apply those skills. To determine long-term gains, you may wish to reassess students several weeks later.

Perhaps your district has implemented a new language arts curriculum stressing the writing process. A yearly writing assessment can be used to determine if student writing skills are improving and to see if attitudes and revision skills are changing. Teachers may contribute their perceptions about the program through questionnaires. Language scores on standardized tests could also be compared.

The following is a more detailed illustration. You are assigned to formatively evaluate a school district's math education program. 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 any parts of the program in need of revision.

Administer assessment tools to gather necessary evidence.

You meet with the district's math teachers and agree on four assessment instruments consistent with the program's stated objectives. The four instruments are: (1) a math content test, (2) a test of students' critical-thinking skills, (3) an attitude inventory assessing students' perceptions of their knowledge of the mathematics included in the program, and (4) an affective self-efficacy inventory reflecting the degree to which students will be successful in using the mathematics skills and knowledge outside of the formal classroom.

Your focus is the district's math education program required in a tenth-grade class. You administer the four assessment instruments before and after the classes and discover that students display substantial progress on the content and skill instruments but almost no change on the two attitude inventories. Based on such results, you would be 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 strengthened to better address the two affective dimensions (students' perceived vulnerability and self-efficacy). 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 math education staff re-think the dimensions on which little student progress is evident. You might also, at this point, seek qualitative data from interviews, individual or focus group sessions about which parts of the program students thought did or did not work.

A data-gathering design for summative evaluations

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

Compare data for summative evaluations.

The *pretest-posttest, two-group design*, portrayed schematically in Figure 3, provides the strongest basis for a summative data collection scheme to address these considerations.

This design involves two groups, with only Group 1 initially receiving the instruction. Group 2 begins as an untreated control group. After Group 1 has completed the program, both groups are posttested. Group 2 can receive the instruction after the administration of the posttest. It is very important that the groups are comparable in terms of ability level, size, gender, etc.

To use this design and provide the program to the control group, enough time must be set aside to ensure that all students receive the program. For example, if a four-week science education unit were given to students as part of a semester-long science course, the program must be given at least eight weeks before the end of the semester in order to give the control-group students the same program during the final four weeks of the semester.

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. Conversely, 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.

Classroom teachers will notice that this is nothing more than establishing "where students are" at the beginning of school and comparing it with "where they are" at the end. It could be as simple as comparing writing samples, computation skills, physical skills or student behaviors from assignments or activities at the start to at the end of the program. There is nothing complicated in this and is typically done by many teachers with no specific evaluation thought in mind.

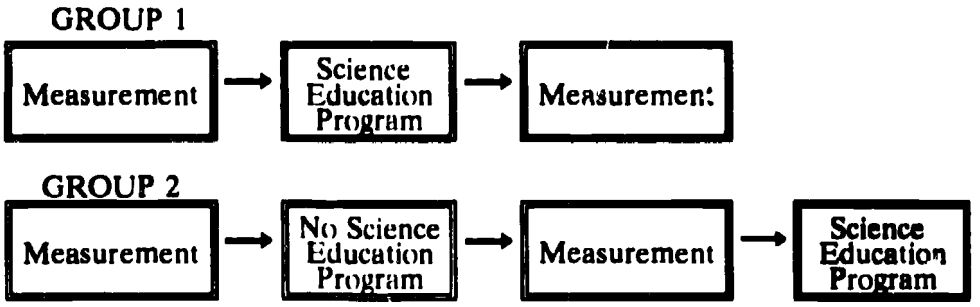


Figure 3. *A pretest-posttest, two-group design*

Final thoughts about Guideline 4

We have paid considerable attention to Guideline 4's focus on the selection of data-gathering designs because, in view of the evaluator's responsibility to present evidence relevant to program decisions, it would be foolish to gather inappropriate evidence. There are, as noted earlier, many more data-gathering strategies than the two basic models presented here. Assessing complex programs, such as the K-12 curriculum in a particular subject area, will require a variety of assessment tools, including the data-gathering designs presented here.

You must be careful when attributing outcomes to educational programs. Other external factors may be making a significant contribution. For example, a seventh-grade science class is doing a lab on bones. Because they don't have teeth, owls often swallow whole, small mammals like mice and shrew. Once a day a pellet of bones, surrounded by hair, is regurgitated under their roosting tree. These pellets are often collected for students to sort and reassemble complete skeletons. This usually successful lab was not well received in a particular class because of an external factor. The class consisted of mostly Native American students and in their culture the owl is a symbol for death, and contact with owls is usually avoided. In another instance, a science class was involved in a unit covering the solar system and showed remarkable gains on a pre-post test. Simultaneously, the television media was intensively covering a vehicular exploration of Mars. Was the spectacular gain influenced by the media coverage or the science program?

Guideline 5: Analyzing the Assessment Data

Guideline 5: Use data-analysis procedures that yield understandable results

Once you have gathered your data, the evidence must be summarized in such a way that is understandable. The audience will most often be teachers, board members, and administrators who typically are not concerned with *statistical* significance. They are more frequently concerned with *practical* significance. A practically significant question might focus on whether a program's effect is large enough to warrant actions such as altering or replacing the program.

Focus on practical analysis.

Thus, 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, arithmetic averages or easily understood data-representation schemes such as bar graphs. For example, after a reading class was completed, the students reported 13 percent more time spent in recreational reading. Or, suppose that, prior to a seat belt education program, 45 of 100 students reported that they drove without using seat belts, whereas several months after the program's conclusion only 38 reported such behavior. In other words, there was more than a 15 percent reduction in those students who drove without using seat belts. Such percentage-based results are easy for decision makers to interpret. People can make sense of percentage-based differences between students' pre-program and post-program performances because people are used to dealing with percentages in other aspects of life.

Percentage correct may not prove to be a suitable descriptive scheme for all assessment instruments you choose. For example, following a nutrition education program you might use a ten-item attitudinal inventory, focusing on students' perceived ability to select low- fat foods, 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.

For a writing assessment, the visual impact of bar graphs showing grade-level composite scores in organization, mechanics, style, and content can clarify curricular strengths and weaknesses.

When looking at pre-program and post-program data, it will be a routine matter to compare the differences between such data to discern whether the program yielded its anticipated effects. Simple pretest-to-posttest percentage changes will usually provide satisfactory data analysis. On the other hand, if much of your assessment data consists of performance assessments, surveys,

questionnaires and anecdotal records, evaluating that data may require discussion, continued research, and subjective analysis.

Final thoughts about Guideline 5

This fifth guideline stresses the desirability of using data-analysis schemes that yield *understandable* results.

Guideline 6: Evaluating Results to Make Modifications

Guideline 6: Report and evaluate results to make recommendations and program modifications as indicated.

If you design and carry out your assessment following the first five guidelines, you will have a manageable set of evidence, primarily student assessment data, bearing on a modest number of important program-relevant decisions. Your task at reporting time is to present that evidence to teachers and administrators in a form most likely to influence the decisions they need to make.

An appropriate level of detail

The report should be brief and hit only the high points, namely, the evidence that bears most directly on the decisions at issue. Try to use visual and/or graphic methods to make the results as palatable to readers as possible. Although it may be difficult, use white space and graphic presentation techniques that stimulate the reader's interest.

Evaluation

Since assessment is the process of collecting and organizing information or data in ways that make it possible for people to evaluate, reporting on the strengths as well as the weaknesses of a program is appropriate. Keep in mind that the evaluation of assessment data can be open to interpretation. Modifications to the program as a result of recommendations from personnel that gathered the data are desirable and suggestions from staff to department chairs and administrators are imperative.

Final thoughts about Guideline 6

This final step in the assessment process, evaluation, may involve decisions made by people other than yourself. You should ask yourself: who will make programmatic decisions based on this assessment? Will it be yourself, your department, principal, superintendent or school board? This will determine the scope and detail of your assessment results.

Implementing Results

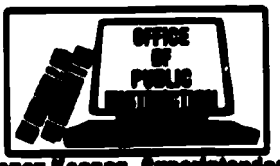
Now that you've finished your six-step assessment process, where do you go from here? Well, a logical procedure would be to look at the evaluation in relation to your program. You should now know the strengths of the program as well as weaknesses. You might see parts needing revision or enhancement as well as parts you will want to continue "as is" or even eliminate. This is where you make changes in your curriculum based on sound data.

Assessment is an ongoing process. This means you never really end your quest for curriculum improvement. Although a logical place to go now might be back to step one, you might be able to skip right to step three or four if you plan to use the same assessment instruments. If you have completed the procedure once, keeping the process in motion will be easier.

Assessment Planning Guidelines

____	1	Determine whether your evaluation is to be formative or summative.
____	2	Focus on a manageable number of important program-related goals.
____	3	Select or construct suitable assessment instruments.
____	4	Use a data-gathering design consistent with the orientation of the evaluation.
____	5	Use data-analysis procedures that yield understandable results.
____	6	Report and evaluate results to make recommendations and program modifications as indicated.

This document was printed entirely with federal funds from the HIV/AIDS Education Cooperative Agreement (No. U63/CCU803049-04) awarded to the Montana Office of Public Instruction from the Centers for Disease Control.



Nancy Keenan, Superintendent
State Capitol
Helena, Montana 59620