

# **An Introduction to Program Evaluation for Classroom Teachers**

*Howard L. Fleischman  
Laura Williams*

**Development Associates, Inc.**  
*1730 North Lynn Street  
Arlington, VA 22209-2023  
(703) 276-0677*

November 1996

# TABLE OF CONTENTS

## Foreword

### I [Introduction](#)

- A. Purposes and Uses of Evaluation
- B. Uses of Evaluation for Local Program Improvement

### II. [Evaluation Process and Plans](#)

- A. Overview of the Evaluation Process
  - Step 1: Defining the Purpose and Scope of the Evaluation
  - Step 2: Specifying the Evaluation Questions
  - Step 3: Developing the Evaluation Design and Data Collection Plan
  - Step 4: Collecting the Data
  - Step 5: Analyzing the Data and Preparing a Report
  - Step 6: Using the Evaluation Report for Program Improvement
- B. Planning the Evaluation

### III. [Evaluation Framework](#)

### IV. [Students Characteristics](#)

### V. [Documentation of Instruction](#)

### VI. [Outcomes](#)

- Alternative, Performance, and Authentic Assessments
- Portfolio Assessment

### VII. [Data Analysis and Presentation of Findings](#)

## APPENDICES

### [Appendix A: Glossary of Terms](#)

### [Appendix B: References](#)

---

## FOREWORD

---

### Why Evaluate?

Evaluation is a tool which can be used to help teachers judge whether a curriculum or instructional approach is being implemented as planned, and to assess the extent to which stated goals and objectives are being achieved. It allows teachers to answer the questions:

*Are we doing for our students what we said we would? Are students learning what we set out to teach? How can we make improvements to the curriculum and/or teaching methods?*

The goals of this document are to introduce teachers to basic concepts within evaluation. A glossary of terms is provided at the end of the document which provides definitions of terms and references to additional sources of information on the World Wide Web.

---

# I. INTRODUCTION 1

---

## A. PURPOSES AND USES OF EVALUATION

Evaluations of educational programs have expanded considerably over the past 30 years. Title I of the Elementary and Secondary Education Act (ESEA) of 1965 represented the first major piece of federal social legislation that included a mandate for evaluation (McLaughlin, 1975). The notion was controversial, and the 1965 legislation was passed with the evaluation requirement stated in very general language. Thus, state and local school systems were allowed considerable room for interpretation and discretion.

The evaluation requirement had two purposes: (1) to ensure that the funds were being used to address the needs of disadvantaged children; and (2) to provide information that would empower parents and communities to push for better education. Others saw the use of information on programs and their effectiveness as a means of upgrading schools. They thought that performance comparisons in evaluations could be used to encourage schools to improve performance. Federal staff in the U.S. Department of Health, Education, and Welfare, (HEW) welcomed the opportunity to have information about programs, populations served, and educational strategies used. The then Secretary of HEW promoted the evaluation requirement as a means of finding out "what works" as a first step to promoting the dissemination of effective practices (McLaughlin, 1975).

Thus there were several different viewpoints regarding the purposes of the evaluation requirement. One underlying similarity of these, however, was the expectation of reform and the view that evaluation was central to the development of change. However, there was also a common assumption that evaluation activities would generate objective, reliable, and useful reports, and that findings would be used as the basis of decision-making and improvement.

Such a result did not occur, however. Widespread support for evaluation did not take place at the local level. There was a concern that federal requirements for reporting would eventually lead to more federal control over schooling (McLaughlin, 1975).

In the 1970's it became clear that the evaluation requirements in federal education legislation were not generating their desired results. Thus, the reauthorization of Title I of the federal Elementary and Secondary Education Act (ESEA) in 1974 strengthened the requirement for collecting information and reporting data by local grantees. It also required the U.S. Office of Education to develop evaluation standards and models for state and local agencies, and required the Office to provide technical assistance so that comparable data would be available nationwide, exemplary programs could be identified, and evaluation results could be disseminated (Wisler & Anderson, 1979, Barnes & Ginsburg, 1979). The Title I Evaluation and Reporting System (TIERS) was part of that development effort. In 1980, the U.S. Department of Education promulgated general administration regulations known as EDGAR, which established criteria for judging evaluation components of grant applications. These various changes in legislation and regulation reflected a continuing federal interest in evaluation data. What was not in place, however, was a system at the federal, state, or local levels for using the evaluation results to effect program or project improvement. In 1988, amendments to ESEA reauthorized the Chapter 1 (formerly Title I) program, and strengthened the emphasis on evaluation and local program improvement. The legislation required that state agencies identify programs that did not show aggregate [achievement](#) gains or

which did not make substantial progress toward the goals set by the local school district. Those programs that were identified as needing improvement were required to write program improvement plans. If, after one year, improvement was not sufficient, then the state agency was required to work with the local program to develop a program improvement process to raise student achievement (Billing, 1990).

In the 1990's, there was a further call for school reform, improvement, and accountability. The National Education Goals were promulgated and formalized through the Educate America Act of 1994. The new law issued a call for "world class" standards, assessment and accountability to challenge the nation's educators, parents, and students.

## **B. USES OF EVALUATION FOR LOCAL PROGRAM IMPROVEMENT**

In much the same way that ideas concerning federal use of evaluation have evolved, so have ideas concerning local use of evaluation. One purpose of any evaluation is to examine and assess the implementation and effectiveness of specific instructional activities in order to make adjustments or changes in those activities. This type of evaluation is often labelled "process evaluation." The focus of process evaluation includes a description and assessment of the curriculum, teaching methods used, staff experience and performance, in-service training, and adequacy of equipment and facilities. The changes made as a result of process evaluation may involve immediate small adjustments (e.g., a change in how one particular curriculum unit is presented), minor changes in design (e.g., a change in how aides are assigned to classrooms), or major design changes (e.g., dropping the use of ability grouping in classrooms).

In theory, process evaluation occurs on a continuous basis. At an informal level, whenever a teacher talks to another teacher or an administrator, they may be discussing adjustments to the curriculum or teaching methods. More formally, process evaluation refers to a set of activities in which administrator and/or evaluators observe classroom activities and interact with teaching staff and/or students in order to define and communicate more effective ways of addressing curriculum goals. Process evaluation can be distinguished from outcome evaluation on the basis of the primary evaluation emphasis. Process evaluation is focused on a continuing series of decisions concerning program improvements, while outcome evaluation is focused on the effects of a program on its intended target audience (i.e., the students).

A considerable body of literature has been developed about how evaluation results can and should be used for improvement (David et al., 1989; Glickman, 1991; Meier, 1987; Miles & Louis, 1990; O'Neil, 1990). Much of this literature has taken a systems approach, in which the authors have examined decision making in school systems, and have recommended approaches for generating school improvement. This literature has identified four key factors associated with effective school reform (David et al., 1989):

Curriculum and instruction must be reformed to promote higher-order thinking by all students; Authority and decision making should be decentralized in order to allow schools to make the most educationally important decisions; New staff roles must be developed so that teachers may work together to plan and develop school reforms; and Accountability systems must clearly associate rewards and incentive to student performance at the skills-building level.

If there is an overriding themes of much of this literature, it is that there must be "ownership" of the reform process by as many of the relevant parties (district and school administrators, teachers, and students) as possible. Change must be seen as a natural and inherent part of the education process, so that individuals in the system accept and feel comfortable with new ways of

performing their functions (Meier, 1991).

---

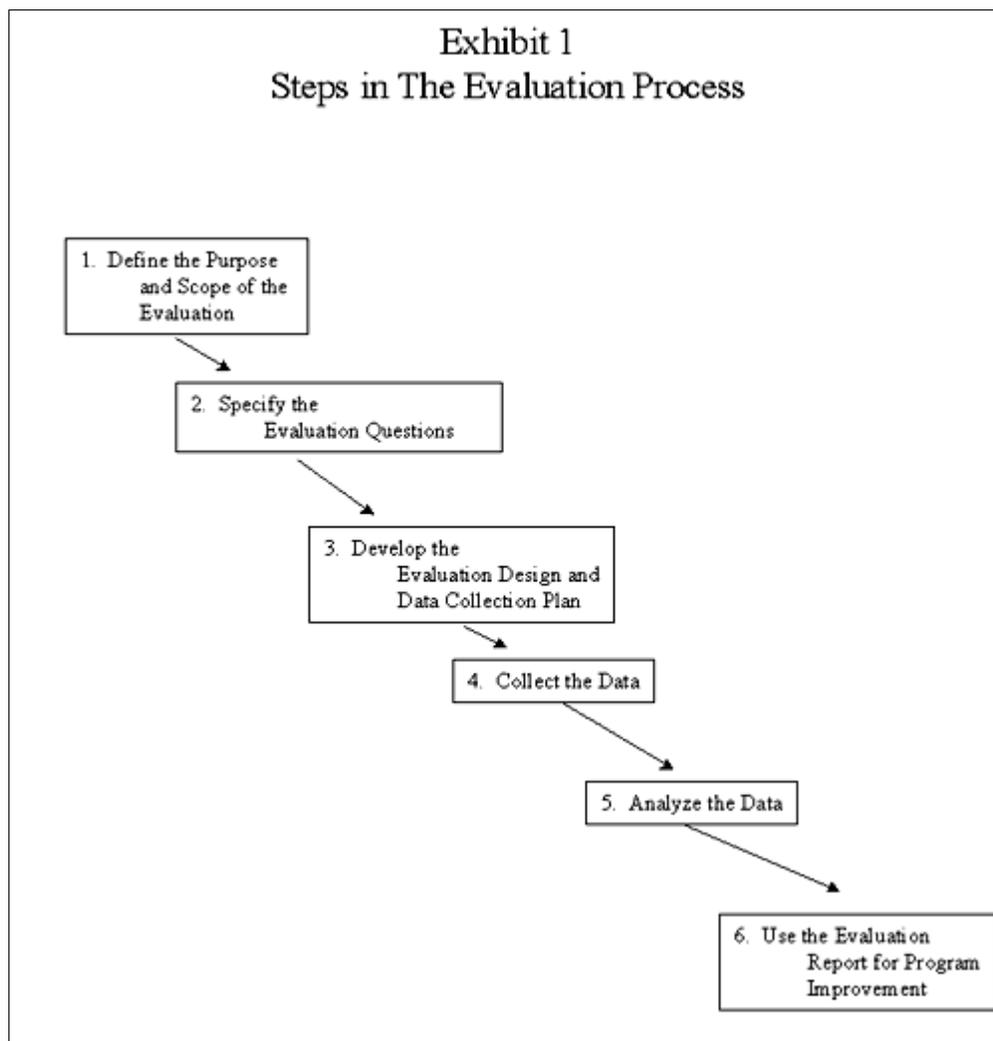
## II. EVALUATION PROCESS AND PLANS

---

As a basic tool for curriculum and instructional improvement, a well planned evaluation can help answer the following questions: How is instruction being implemented? (What is taking place?) To what extent have objectives been met? How has instruction impacted on its target population? What contributed to successes and failures? What changes and improvements should be made? Evaluation involves the systematic and objective collection, analysis, and reporting of information or data. Using the data for improvement and increased effectiveness then involves interpretation and judgement based on prior experience.

### A. Overview of the Evaluation Process

The evaluation process can be described as involving six progressive steps. These steps are shown in Exhibit 1. It is important to remember that initiating an evaluation cannot wait until developing and teaching an instructional unit is completed. An evaluation should be incorporated into overall planning, and should be initiated when instruction begins. In this manner, instructional processes and activities can be documented from their beginning, and baseline data on students can be collected before instruction begins.



### ***Step 1: Defining the Purpose and Scope of the Evaluation***

The first step in planning is to define an evaluation's purpose and scope. This helps set the limits of the evaluation, confining it to a manageable size. Defining its purpose includes deciding on the goals and objectives for the evaluation, and on the audience for the evaluation results. The evaluation goals and objectives may vary depending on whether the instructional program or curriculum being evaluated is new and is going through a try-out period for which the planning and implementation process needs to be documented, or if a curriculum has been thoroughly tested and needs documentation of its success before information is widely disseminated and adoption by others encouraged.

Depending on the purpose, the audience for evaluation may be restricted to the individual teacher, and school staff member, or may include a wider range of individuals, from school administrators to planners and decision makers at the local, state, or national level.

The scope of the evaluation depends on the evaluation's purpose and the information needs of its intended audience. These needs determine the specific components of a program which should be evaluated and on the specific project objectives which are to be addressed. If a broad evaluation of a curriculum has recently been conducted, a limited evaluation may be designed to target certain parts which have been changed, revised, or modified. Similarly, the evaluation may be designed to focus on certain objectives which were shown to be only partially achieved in the past. Costs and resources available to conduct the evaluation must also be considered in this

decision.

## ***Step 2: Specifying the Evaluation Questions***

Evaluation questions grow out of the purpose and scope specified in the previous step. They help further define the limits of the evaluation. The evaluation questions will be structured to address the needs of the specific audience to whom the evaluation is directed. Evaluation questions should be developed for each component which falls into the scope which was defined in the previous step. For example, questions may be formulated which concern the adequacy of the curriculum and the experience of the instructional staff; other questions may concern the appropriateness of the skills or information being taught; and finally, evaluation questions may relate to the extent to which students are meeting the objectives set forth by the instructional program.

A good way to begin formulating evaluation questions is to carefully examine the instructional objectives; another source of questions is to anticipate problem areas concerning teaching the curriculum. Once the evaluation questions are developed, they should be prioritized and examined in relation to the time and resources available. Once this is accomplished, the final set of evaluation questions can be selected.

## ***Step 3: Developing the Evaluation Design and Data Collection Plan***

This step involves specifying the approach to answering the evaluation questions, including how the required data will be collected. This will involve:

specifying the data sources for each evaluation question; specifying the types of data, data collection approaches, and [instruments](#) needed; specifying the specific time periods for collecting the data; specifying how the data will be collected and by whom; and specifying the resources which will be required to carry out the evaluation.

The design and data collection plan is actually a roadmap for carrying out the evaluation. An important part of the design is the development or selection of the instruments for collecting and recording the data needed to answer the evaluation questions. Data collection instruments may include record-keeping forms, questionnaires, interview guides, tests, or other assessment measures. Some of the instrumentation may already be available, i.e., standardized tests. Some will have to be modified to meet the evaluation needs. In other cases, new instruments will have to be created.

In designing the instruments, the relevance of the [items](#) to the evaluation questions and the ease or difficulty of obtaining the desired data should be considered. Thus, the instruments should be reviewed to ensure that the data can be obtained in a cost-effective manner and without causing major disruptions or inconveniences to the class.

## ***Step 4: Collecting the Data***

Data collection should follow the plans developed in the previous step. Standardized procedures need to be followed so that the data are reliable and valid. The data should be recorded carefully so they can be tabulated and summarized during the analysis stage. Proper record-keeping is similarly important so that the data are not lost or misplaced. Deviations from the data collection plan should be documented so that they can be considered in analyzing and interpreting the data.

## ***Step 5: Analyzing the Data and Preparing a Report***

This step involves tabulating, summarizing, and interpreting the collected data in such a way as to answer the evaluation questions. Appropriate descriptive measures (frequency and percentage distributions, central tendency and variability, correlation, etc.) and inferential techniques (significance of difference between [means](#) and other statistics, analysis of variance, chi-square, etc.) should be used to analyze the data. An individual with appropriate statistical skills should have responsibility for this aspect of the evaluation.

The evaluation will not be completed until a report has been written and the results communicated to the appropriate administrators and decision makers. In preparing the report, the writers should be clear about the audience for whom the report is being prepared. Two broad questions need to be considered: (1) What does the audience need to know about the evaluation results? and (2) How can these results be best presented? Different audiences need different levels of information. Administrators need general information for policy decision making, while teachers may need more detailed information which focuses on program activities and effects on participants.

The report should cover the following:

The goals of the evaluation; The procedures or methods used; The findings; and The implication of the findings, including recommendations for changes or improvements in the program.

Importantly, the report should be organized so that it clearly addresses all of the evaluation questions specified in Step 2.

## ***Step 6: Using the Evaluation Report for Program Improvement***

The evaluation should not be considered successful until its results are used to improve instruction and student success. After all, this is the ultimate reason for conducting the evaluation. The evaluation may indicate that an instructional activity is not being implemented according to plan, or it may indicate that a particular objective is not being met. If so, it is then the responsibility of teachers and administrators to make appropriate changes to remedy the situation. Schools should never be satisfied with their programs. Improvements can always be made, and evaluation is an important tool for accomplishing this purpose.

## **B. Planning the Evaluation**

An evaluation may be conducted by an independent, experienced evaluator. This individual will be able to provide the expertise for planning and carrying out an evaluation which is comprehensive, objective, and technically sound. If an independent evaluator is used, school staff should work closely with him/her beginning with the planning stage to ensure the evaluation meets the exact needs of the instructional program.

Adequate time and thought for planning an evaluation is essential, and will give the school staff and evaluator an opportunity to develop ideas about what they would like the evaluation to accomplish. The evaluation should address the goals and objectives for students specified in the curriculum. In some cases, however, one or more goals or objectives may require more attention than others. Some activities or instructional strategies may have been recently implemented, or teachers may be aware of some special problems which should be addressed. For example, there might have been a recent breakdown in communication between teachers; or the characteristics of

students might have begun to differ significantly from the past. The evaluator must then familiarize himself or herself with the special issues of concern on which the evaluation should focus.

Thus, the initial step of the evaluation process involves thinking about any special needs which will help in planning the overall evaluation. Problems identified and evaluation questions which focus on curriculum and instructional materials might suggest that an evaluator is needed with particular expertise in those areas.

In summary, defining the scope involves setting limits, identifying specific areas of inquiry, and deciding on what parts of the program and on which objectives the evaluation will focus. The scope does not answer the question of how the evaluation will be conducted. In establishing the scope, one is actually determining which components or parts of the program will be evaluated, and implies that the evaluation may not cover every aspect and activity.

---

### **III. EVALUATION FRAMEWORK**

---

This chapter presents a framework for evaluating an instructional program, combining an outcome evaluation with a process evaluation. An outcome evaluation attempts to determine the extent to which a program's specific objectives have been achieved. On the other hand, the process evaluation seeks to describe an instructional program and how it was implemented, and through this, attempt to gain an understanding of why the objectives were or were not achieved.

Evaluators have been criticized in the past for focusing on outcome evaluation and excluding the process side, or focusing on process evaluation without examining outcomes. The framework presented here incorporates both the process and outcome side. In this manner, one can determine the effect (or outcome) of a program of instruction, and also understand how the program produced that effect and how the program might be modified to produce that effect more completely and efficiently.

In order to focus on both program process and outcomes, an evaluation should be designed in which evaluation questions, and data collection and analysis, address the following:

- Students;
- Instruction; and
- Outcomes.

Descriptions are prepared of the participants and the program activities and services which are implemented. Outcomes of the program are also assessed. The description of the participants and activities and services are used to explain how the outcomes were achieved and to suggest changes which may produce these outcomes more effectively and efficiently.

Each evaluation component is described below.

## 4

**Students**

This component defines the characteristics of the students including, for example, grade level, age, socio-economic level, [aptitude](#), achievement (grades and test scores). In addition to their use for descriptive purposes, these data are useful for comparisons with other groups of students who are included in the evaluation.

**Instruction**

This component describes how the key activities of the curriculum or instructional program are implemented, including instructional objectives, hours of instruction, teacher characteristics and experience, etc. In this manner, the outcomes or results achieved by the instructional program can be attributed to what actually has taken place, rather than what was planned to occur. This component also addresses the questions of what parts of the program have been fully implemented, partially implemented, and not implemented.

**Outcomes**

This component concerns the effects that the program has on students, and to what extent the program has met its stated objectives. At the end of the instructional unit, data may be collected on what is as learned with respect to the instructional objectives and competencies.

Using the above three evaluation components, a comprehensive assessment of a program may be designed. Not only will this evaluation approach allow a teacher to determine the extent to which instructional goals and objectives are met, but will also enable them to understand how those outcomes were achieved and to make improvements in the future.

\* \* \*

The evaluation framework presented above may be implemented using the six-step process described in Chapter II. The framework describes *what* should be included in the evaluation; the six-step process describes *how* the evaluation may be planned and carried out. Guidelines for defining the scope of the evaluation, specifying evaluation questions, and developing the data collection plans for each of the three evaluation components are discussed in the following sections.

---

## IV. STUDENT CHARACTERISTICS

---

This chapter concerns that part of the evaluation related to the collection of descriptive data on students, including their number and characteristics. Most data will be descriptive in nature, such as age, gender, ethnicity, and prior educational achievement. Some data will be baseline measures related to program objectives. These data will be compared to data collected at completion of the instructional unit to determine effects.

The specific student data to be collected depends on the parts of the curriculum being addressed by the evaluation. From this, a set of evaluation questions may be developed which focus on these parts. This then leads to specification of variables, and the development of the data collection plans and instruments.

Evaluation questions concerning students are shown in Exhibit 2, along with examples of the relevant variables.

**EXHIBIT 2**  
**Program Participants: Examples of Evaluation Questions and Variables**

<b>Evaluation Questions</b>	<b>Variables</b>
1. How many students are exposed to the instructional unit?	Number of students in class.
2. What are the demographic characteristics of the students?	Grade, Age, Sex, Racial/Ethnic Group.
3. What is the level of the students' basic skills (language and mathematics ability)?	Scores on nationally standardized achievement tests.
4. What is the level of students' knowledge and skills prior to being exposed to the instructional unit being evaluated?	Scores on pre-tests of knowledge and skills related to objectives of curriculum being evaluated.

---

## V. DOCUMENTATION OF INSTRUCTION

---

This chapter focuses on documenting the objectives and scope of the instructional unit on which the evaluation is focused. The data to be collected will focus on the instruction which actually has taken place, rather than what was originally planned.

Evaluation questions need to be developed to focus the data collection requirements. Sample questions are shown in Exhibit 3.

**EXHIBIT 3**  
**Instruction: Examples of Evaluation Questions and Variables**

<b>Evaluation Questions</b>	<b>Variables</b>
1. What are the instructional objectives? Are these objectives clearly stated and measurable?	Instructional Goals and Objectives
2. What is the total number of hours of instruction provided?	Number of Hours
3. What is the instructor/student ratio?	Number of Students Per Class

4. What are the qualifications and experience of the teacher(s)? Do teachers have necessary qualifications to meet the needs of the students?	Background and Experience of Teachers
5. What kind of staff development and training are provided to teachers? Are development and training appropriate and sufficient?	Staff Development and Training Activities
6. Did the curriculum as taught follow the original plan? Is curriculum appropriate?	Description of Instruction
7. What instructional methods and materials are used? Are the methods appropriate?	Description of Instruction; Methods and Materials

---

## VI. OUTCOMES

---

Program outcome data are used to determine the extent to which the curriculum or instructional unit is meeting its goals and objectives for students.

As with the other evaluation components, the evaluation planners must define the scope of the outcome data to be collected. This should be accomplished by developing a set of evaluation questions to assess the extent to which the instructional goals and objectives are met. Examples of evaluation questions directed at the outcomes of programs is shown in Exhibit 5. Also shown are the relevant variables which relate to the questions. Additional evaluation questions may also be specified which address any special issues and concerns of the instructional unit.

**EXHIBIT 4**  
**Student Outcomes: Examples of Evaluation Questions and Variables**

<b>Evaluation Questions</b>	<b>Variables</b>
1. How many students were exposed to the instructional unit?	Number of Students
2. To what extent were instructional objectives met?	Teacher Ratings
3. To what extent did students increase their knowledge and skills?	Pre/Post Tests Related to Instructional Objectives

In addition to the traditional methods of [testing](#) to assess student outcomes, alternative methods of assessment have been developed in recent years. Some of these new approaches are discussed below.

### ALTERNATIVE, PERFORMANCE, AND AUTHENTIC ASSESSMENTS

Traditionally, most assessments of students in the United States has been accomplished through the use of formalized tests. This practice has been called into question in recent years because such tests may not be accurate indicators of what the student has learned (e.g. a student may simply guess correctly on a [multiple-choice item](#)), or even if so, students may not be learning in ways that will help them participate more fully in the "real world." Further, alternative approaches which more fully involve the student in the evaluation process have been praised for increasing student interest and motivation. In this section we provide some background for this issue and introduce three assessment concepts relevant to this new emphasis: [alternative assessment](#), [performance assessment](#), and authentic assessment.

[Standardized testing](#) was initiated to enable schools to set clear, justifiable, and consistent standards for its students and teachers. Such tests are currently used for several purposes beyond that of classroom evaluation. They are used to place students in appropriate level courses; to guide students in making decisions about various courses of study; and to hold teachers, schools, and school districts accountable for their effectiveness bases upon their students' performance.

Especially when high stakes have been placed on the results of a test (e.g. deciding the quality of the teacher or school), teachers have become more likely to "teach to the test" in order to improve their students' performance. If the test were a thorough evaluation of the desired skills and reflected mastery of the subject, this would not necessarily be a problem. However, standardized tests generally make use of multiple-choice or short answers to allow for efficient processing of large numbers of students. Such testing techniques generally make use of [lower-order cognitive skills](#), whereas students may need to use more complex skills outside of the classroom.

In order to encourage students to use higher-order cognitive skills and to evaluate them more comprehensively, several alternative assessments have been introduced. Generally, alternative assessments are any non-standardized evaluation techniques which utilize complex thought processes. Such alternatives are almost exclusively performance-based and criterion-referenced. Performance-based assessment is a form of testing which requires a student to create an answer or product or demonstrate a skill that displays his or her knowledge or abilities. Many types of performance assessment have been proposed and implemented, including: projects or group projects, [essays](#) or writing samples, [open-ended problems](#), interviews or [oral presentations](#), science experiments, [computer simulations](#), [constructed-response questions](#), and [portfolios](#). Performance assessments often closely mimic real life, in which people generally know of upcoming projects and deadlines. Further, a known challenge makes it possible to hold all students to a higher standard.

Authentic assessment is usually considered a specific kind of performance assessment, although the term is sometimes used interchangeably with performance-based or alternative assessment. The authenticity in the name derives from the focus of this evaluation technique to directly measure complex, real world, relevant tasks. Authentic assessments can include writing and revising papers, providing oral analyses of world events, collaborating with others in a debate, and conducting research. Such tasks require the student to synthesize knowledge and create polished, thorough, and justifiable answers. The increased [validity](#) of authentic assessments stems from their relevance to classroom material and applicability to real life scenarios.

Whether or not specifically performance-based or authentic, alternative assessments achieve greater [reliability](#) through the use of predetermined and specific evaluation criteria. [Rubrics](#) are often used for this purpose and can be created by one teacher or a group of teachers involved in

similar assessments. (A rubric is a set of guidelines for [scoring](#) which generally states all of the dimensions being assessed, contains a scale, and helps the grader place the given work on the scale.) Creating a rubric is often time-consuming, but it can help clarify the key features of the performance or product and allow teachers to grade more consistently.

It is widely acknowledged that performance-based assessments are more costly and time-consuming to conduct, especially on a large scale. Despite these impediments, a few states, such as California and Connecticut, have begun to implement statewide performance evaluations. Various solutions to the issue of evaluation cost are being proposed, including combining alternative assessment measures with more traditional standardized ones, conducting performance assessments on only a sample of work, and for larger contexts (school, district, or state) sampling only a fraction of the students.

Although the evaluation of performances or products may be relatively costly, the gains to teacher professional development, local assessing, student learning, and parent participation are many. For example, parents and community members are immediately provided with directly observable products and clear evidence of a student's progress, teachers are more involved with evaluation development and its relation to the curriculum, and students are more active in their own evaluation. Thus, it appears that alternative evaluations have great potential to benefit students and the larger community.

## PORTFOLIO ASSESSMENT

Portfolio assessment is one of the most popular forms of performance evaluation. Portfolios are usually files or folders that contain collections of a student's work, compiled over time. At first, they were used predominantly in the areas of art and writing, where drafts, revisions, works in progress, and final products are typically included to show student progress. However, their use in other fields, such as mathematics and science, has become somewhat more common. By keeping track of a student's progress, portfolio assessments are noted for "following a student's successes rather than failures."

Well designed portfolios contain student work on key instructional tasks, thereby representing student accomplishment on significant curriculum goals. The teachers gain an opportunity to truly understand what their students are learning. As products of significant instructional activity, portfolios reflect contextualized learning and complex thinking skills, not simply routine, low level cognitive activity. Decisions about what items to include in a portfolio should be based on the purpose of the portfolio to avoid becoming simply a folder of student work. Portfolios exist to make sense of students' work, to communicate about their work, and to relate the work to a larger context. They may be intended to motivate students, to promote learning through reflection and self-assessment, and to be used in evaluations of students' thinking and writing processes. The content of portfolios can be tailored to meet the specific needs of the student or subject area. The materials in a portfolio should be organized in chronological order. This is facilitated by dating every component of the folder. The portfolio can further be organized by curriculum area or category of development.

Portfolios can be evaluated in two general ways, depending on the intended use of the scores. The first, and perhaps most common, way is criterion-based evaluation. Student progress is compared to a [standard](#) of performance consistent with the teacher's curriculum, regardless of other students' performances. The level of achievement may be measured in terms such as "basic," "proficient," and "advanced," or it may be evaluated with several more levels to allow for very high standards or greater differentiation between levels of student achievement. The second

portfolio evaluation technique involves measuring individual student progress over a period of time. It requires the assessment of changes in students' skills or knowledge.

There are several techniques which can be used to assess a portfolio. Either portfolio evaluation method can be operationalized by using rubrics (guidelines for scoring which state all of the dimensions being assessed). The rubric may be holistic and produce a single score, or it may be analytic and produce several scores to allow for evaluation of distinctive skills and knowledge. Holistic grading, often used with portfolio assessment, is based on an overall impression of the performance; the rater matches his or her impression with a point scale, generally focussing on specific aspects of the performance.

Whether used holistically or analytically, good scoring criteria clarify instructional goals for teachers and students, enhance fairness by informing students of exactly how they will be assessed, and help teachers to be accurate and unbiased in scoring. Portfolio evaluation may also include peer and teacher conferencing as well as peer evaluation. Some instructors require that their students evaluate their own work as they compile their portfolios as a form of reflection and self-monitoring.

There are, of course, some problems associated with portfolio assessment. One of the difficulties involves large scale assessment. Portfolios can be very time consuming and costly to evaluate, especially when compared to scannable tests. Further, the reliability of the scoring is an issue. Questions arise about whether a student would receive the same grade or score if the work was evaluated at two different points in time, and grades or scores received from different teachers may not be consistent. In addition, as with any form of assessment, there may be [bias](#) present.

Various solutions for the issues of fairness and reliability have been examined. Some research has found that using a small range of possible scores or grades (e.g. A, B, C, D, and F) produces far more reliable results than using a large range of scores (e.g. a 100 point scale) when evaluating performances. Also, some teachers incorporate holistic grading to more consistently evaluate their students. When based on predetermined criteria in this manner, rater reliability is increased. Another method of increasing fairness in portfolio assessment involves the use of multiple raters. Having another qualified teacher rate the portfolios helps ensure that the initial scores given reflect the competence of the work. A third method for testing the reliability of the scoring is to re-score the portfolio after a set period of time, perhaps a few months, to compare the two sets of marks for consistency.

Bias, as previously noted, is also an issue of concern with the development and grading of portfolio tasks. To help avoid bias, portfolio tasks can be discussed with a variety of teachers from diverse cultural backgrounds. In addition, teachers can track how students from different backgrounds perform on the individual tasks and reassess the fairness if significant differences are noted.

Some recommendations for implementing alternative (e.g. portfolio) assessment activities were made by the Virginia Education Association and Appalachia Educational Laboratory. These suggestions include: start small by following someone else's example or combining one alternative activity with more traditional measures, develop rubrics to clarify standards and expectations, expect to invest a lot of time at first, develop assessments as you plan the curriculum, assign a high value to the assessment so the students will view it as important, and incorporate peer assessment into the process to relieve you of some of the grading burden.

The expectations for portfolio assessment are great. Although teachers need a lot of time to

develop, implement, and score portfolios, their use has positive consequences for both learning and teaching. Research has shown that such assessments can lead to increases in student skills, achievement, and motivation to learn.

---

## VII. DATA ANALYSIS AND PRESENTATION OF FINDINGS

---

Following data collection, the next steps in the evaluation process involve data analysis and preparation of a report. These steps may require the expertise of an experienced evaluator who is objective and independent. This is important for the acceptability of the report's findings, conclusions, and recommendations.

The evaluator will be responsible for developing and carrying out a data analysis plan which is compatible with the evaluation's goals and audience. To a large extent, data will be descriptive in nature and may be presented in narrative and tabular format. However, comparisons of pre, and post, measures may require more sophisticated techniques, depending on the nature of the data.

The data will be analyzed to answer the evaluation questions specified in the evaluation plan. Thus, the analysis will allow the evaluator to:

- describe the school and classroom environment;  
Arial
  
- describe the characteristics of the students;  
Arial
  
- describe the instructional objectives and activities;  
Arial
  
- describe the outcomes;  
Arial
  
- examine and assess the extent to which the instructional plan was followed;  
Arial
  
- examine and assess the extent to which the outcomes met the instructional goals and objectives; and  
Arial
  
- examine how the program environment, teachers, and the instructional program and methods affected the extent to which the outcomes were achieved, and how the program can be improved to achieve increased success.  
Arial

Arial

An evaluation report will then:

- describe the accomplishments of the program, identifying those instructional elements that were the most effective;

Arial

Arial

- describe instructional elements that were ineffective and problematic as well as areas that need modifications in the future; and

Arial

Arial

- describe the outcomes or the impact of the instructional unit on students.

Arial

Arial

Complete documentation will make the report useful for making decisions about improving curriculum and instructional strategies. In other words, the evaluation report is a tool supporting decision-making, program improvement, accountability, and quality control.

---

<sup>1</sup>Adapted from Hopstock, Young, and Zehler, 1993. [Back to chapter 1](#)

---

## APPENDICES

### APPENDIX A: GLOSSARY OF TERMS

### APPENDIX B: REFERENCES

---

## APPENDIX A

---

### GLOSSARY OF TERMS

For a wide variety of information about Assessment and Evaluation, see: <http://ericae.net/>

**Achievement** -- performance of a student.

**Age norms** -- values representing the average or typical achievement of individuals in a specific

age group.

**Alternative assessment** -- kind of evaluation designed to assess higher-order cognitive skills and the application of knowledge to new problem-solving situations. For more information, see: [gopher://vmsgopher.cua.edu/00gopher\\_root\\_eric\\_ae%3a%5b\\_alt%5d\\_recadm.txt](http://gopher://vmsgopher.cua.edu/00gopher_root_eric_ae%3a%5b_alt%5d_recadm.txt)

**Aptitude** -- capacity of a student to perform.

**Authentic measurement** -- assessment directly examining student performance on intellectual tasks; requires students to be effective performers with acquired knowledge. Such evaluations often simulate the conditions that students would experience in applying their knowledge or skill in the real world. For more information, see: [gopher://vmsgopher.cua.edu:70/00gopher\\_root\\_eric\\_ae%3A%5B\\_alt%5D\\_case1.txt](http://gopher://vmsgopher.cua.edu:70/00gopher_root_eric_ae%3A%5B_alt%5D_case1.txt) [gopher://vmsgopher.cua.edu:70/00gopher\\_root\\_eric\\_ae%3A%5B\\_alt%5D\\_read.txt](http://gopher://vmsgopher.cua.edu:70/00gopher_root_eric_ae%3A%5B_alt%5D_read.txt) [gopher://vmsgopher.cua.edu/00gopher\\_root\\_eric\\_ae%3a%5b\\_alt%5d\\_write.txt](http://gopher://vmsgopher.cua.edu/00gopher_root_eric_ae%3a%5b_alt%5d_write.txt)

**Bias** -- lack of objectivity, fairness, or impartiality in assessment. For more information, see: <http://www.ed.gov/pubs/IASA/newsletters/assess/pt1.html>

**Closed-ended problems** -- type of question with a short list of possible responses from which the student attempts to choose the correct answer. Common types of closed-ended problems include multiple choice and true and false.

**Competency-based** -- an approach to teaching, training, or evaluating that focuses on identifying the competencies needed by the trainee and on teaching/evaluating to [mastery level](#) on these, rather than on teaching allegedly relevant academic subjects to various subjectively determined achievement levels. (Similar to performance-based.)

**Computer-managed instruction (CMI)** -- use of a computer to track all student records. It is important for large-scale individualized instruction and may include computer diagnoses of student problems and provide recommendations for further study. For more information, see: <http://www.iksx.nl/ctc/1/cmi.htm>

**Computer simulations** -- a technique that can be used in performance assessment to enable students to manipulate variables in an experiment to explain some phenomenon. Data can be generated and graphed in multiple ways.

**Constructed-response questions** -- type of performance assessment problem which requires students to produce their own answers rather than select from an array of possible answers. Such questions may have just one correct response, or they may be more open-ended, allowing a range of responses. For more information, see: [gopher://vmsgopher.cua.edu:70/00gopher\\_root\\_eric\\_ae%3A%5B\\_alt%5D\\_techm.txt](http://gopher://vmsgopher.cua.edu:70/00gopher_root_eric_ae%3A%5B_alt%5D_techm.txt)

**Content analysis** -- process of systematically determining the characteristics of a body of material or practices, such as tests, books, or courses.

**Content standards** -- specifications of the general domains of knowledge that students should learn in various subjects. For more information, see: <http://www.ed.gov/pubs/IASA/newsletters/assess/pt1.html> <http://www.ed.gov/pubs/IASA/newsletters/standards/pt1.html> For science

see:

**Criterion-referenced** -- an approach which focuses on whether a student's performance meets a predetermined standard level, usually reflecting mastery of the skills being tested. This approach does not consider the student's performance compared that of others, as with norm-referenced approaches.

**Curriculum evaluation** -- a product assessment used to study the outcomes of those using a specific course of instruction, or a process assessment used to examine the content validity of the course. Components of curriculum evaluation may include determining the actual nature of the curriculum compared to official descriptions, evaluation of its academic quality, and assessing student learning. For more information, see: [http://www.sasked.gov.sk.ca/docs/artsed/g7arts\\_ed/g7evlae.html](http://www.sasked.gov.sk.ca/docs/artsed/g7arts_ed/g7evlae.html)

**Dimensions, traits, or subscales** -- the subcategories used in the evaluation of a performance or portfolio product.

**Essays** -- type of performance assessment used to evaluate a student's understanding of the subject through a written description, analysis, explanation, or summary. It can also be used to evaluate a student's composition skills. For more information, see:

[gopher://vmsgopher.cua.edu:70/00gopher\\_root\\_eric\\_ae%3A%5B\\_alt%5D\\_techn.txt](gopher://vmsgopher.cua.edu:70/00gopher_root_eric_ae%3A%5B_alt%5D_techn.txt)

**Grade equivalent** -- the estimated grade level that corresponds to a given score.

**Grading** -- allocating students to an ordered (usually small) set of named categories ordered by merit; also known as rating. Grades are generally criterion-referenced.

**Higher-order skills** -- abilities used for more sophisticated tasks requiring application of knowledge to solve a problem.

**Holistic scoring** -- rating based on an overall impression of a performance or portfolio product. The rater matches his or her impression with a point scale, generally focussing on specific aspects of the performance or product.

**Institutional evaluation** -- a complex assessment, typically involving the evaluation of a set of programs provided by an institution and an evaluation of the overall management of the institution. For more information, see: <http://www.ericae.net/>

**Instructional assessment** -- evaluation of class progress relevant to the curriculum.

**Instrument** -- a measuring device (e.g. test) used to determine the present value of something under observation.

**Item** -- an individual question or exercise in an assessment or evaluative instrument.

**Lower-order skills** -- abilities used for less sophisticated tasks, such as recognition, recall, or simple deductive reasoning.

**Mastery level** -- the level of performance actually needed on a criterion; sometimes the level

thought to be optimal and feasible.

**Materials assessment** -- evaluation of the effectiveness of the products used to teach students or assess student progress or levels. For more information, see: [http://www.cua.edu/www/ERIC\\_ae/intbod.htm#InstitE](http://www.cua.edu/www/ERIC_ae/intbod.htm#InstitE)

**Mean** -- the average. The mean score of a test is found by adding all of the scores and dividing that sum by the number of people taking the test.

**Measurement** -- determination of the magnitude of a quantity; includes numerical scoring.

**Median score** -- value that divides a group into two, as nearly as possible; the "middle" performance.

**Mode** -- the most frequent score or score interval. Distributions with two equally common scores are called **bimodal**.

**Multidiscipline** -- a subject that requires the methods of several branches or specialties.

**Multiple-choice test (or item)** -- a test, or question on a test, in which each question is followed by several alternative answers, one and only one of which is to be selected as the correct response.

**Norm-referenced** -- an approach which focuses on placing a student's score along a normal distribution of scores from students all taking the same test. This approach does not focus on absolute levels of mastery, as with criterion-referenced approaches.

**Open-ended problems** -- type of question used in performance assessment which allows the student to independently create a response; there may be multiple correct responses to the problem. For more information, see: [gopher://vmsgopher.cua.edu:70/00gopher\\_root\\_eric\\_ae%3A%5B\\_alt%5D\\_openread.txt](gopher://vmsgopher.cua.edu:70/00gopher_root_eric_ae%3A%5B_alt%5D_openread.txt)

**Opportunity-to-learn standards** -- criteria for evaluating whether schools are giving students the chance to learn material reflected in the [content standards](#). This may include such specifics as the availability of instructional materials or the preparation of teachers.

**Oral presentations/interviews** -- type of performance assessment which allow students to verbalize their knowledge. For more information, see: [gopher://vmsgopher.cua.edu:70/00gopher\\_root\\_eric\\_ae%3A%5B\\_alt%5D\\_techn.txt](gopher://vmsgopher.cua.edu:70/00gopher_root_eric_ae%3A%5B_alt%5D_techn.txt)

**Percentile** -- the percent of individuals in the norming sample whose scores were below a specific score. (Percentiles are based on 100 divisions or groupings, deciles are based on 10, and quartiles are based on 4 groups.)

**Percent score** -- the percent of items that are answered correctly.

**Performance assessment** -- testing that requires a student to create an answer or a product that demonstrates his or her knowledge or skills. This evaluation method emphasizes the validity of the test and is more easily scored using [criterion-referenced](#) than norm-referenced approaches. For

more information, see: <http://cresst96.cse.ucla.edu/teacher.htm#pa> <http://www.ed.gov/pubs/IASA/newsletters/assess/pt5.html> [gopher://vmsgopher.cua.edu/00gopher\\_root\\_eric\\_ae%3a%5b\\_alt%5d\\_overv.txt](mailto:gopher://vmsgopher.cua.edu/00gopher_root_eric_ae%3a%5b_alt%5d_overv.txt) [gopher://vmsgopher.cua.edu/00gopher\\_root\\_eric\\_ae%3a%5b\\_alt%5d\\_crit.txt](mailto:gopher://vmsgopher.cua.edu/00gopher_root_eric_ae%3a%5b_alt%5d_crit.txt) [gopher://vmsgopher.cua.edu/00gopher\\_root\\_eric\\_ae%3a%5b\\_alt%5d\\_urban.txt](mailto:gopher://vmsgopher.cua.edu/00gopher_root_eric_ae%3a%5b_alt%5d_urban.txt)

**Performance criteria** -- a pre-determined list of observable standards used to rate student achievement to determine student progress. Such standards should include considerations of reliability and validity.

**Performance standards** -- external criteria used to establish the degree or quality of students' performance in the subject area set out by the content standards, answering the question "How good is good enough?" For more information, see: <http://www.ed.gov/pubs/IASA/newsletters/standards/pt1.html>

**Portfolios** -- type of performance assessment which involves the ongoing evaluation of a cumulative collection of creative student works. It can include student self-reflection and monitoring. For more information, see: <http://cresst96.cse.ucla.edu/teacher.htm#pa> [gopher://vmsgopher.cua.edu:70/00gopher\\_root\\_eric\\_ae%3A%5B\\_alt%5D\\_child.txt](mailto:gopher://vmsgopher.cua.edu:70/00gopher_root_eric_ae%3A%5B_alt%5D_child.txt)

**Practice effect** -- the improved performance produced by taking a second test with the same or closely similar items, even if no additional learning has occurred between the tests.

**Program evaluation** -- assessment of the effectiveness of particular instructional interventions or programs. For more information, see: [http://www.sasked.gov.sk.ca/docs/artsed/g7arts\\_ed/g7evlae.html](http://www.sasked.gov.sk.ca/docs/artsed/g7arts_ed/g7evlae.html) [http://www.cua.edu/www/ERIC\\_ae/intbod.htm#InstitE](http://www.cua.edu/www/ERIC_ae/intbod.htm#InstitE)

**Ranking** -- placing students in an order, usually of merit, on the basis of their relative performance on a test, measurement, or observation.

**Rating** -- see **Grading**.

**Rating scales** -- a written list of performance criteria related to a specific activity or product which an observer uses to assess student performance on each criterion in terms of its quality.

**Raw score** -- the number of items that are answered correctly on a test or assignment, before being converted (e.g. to a percentile or grade equivalent).

**Reliability** -- the extent to which an assessment is dependable, stable, and consistent when administered to the same individuals on different occasions.

**Rubric** -- a set of guidelines for scoring or grading which generally states all of the dimensions being assessed, contains a scale, and helps the grader place the given work on the scale. For more information, see: <http://www.servtech.com/public/germaine/rubric.html> <http://www.nwrel.org/eval/toolkit/traits/>

**Scoring** -- use of a numerical grade to rank student performance and assessment.

**Standard** -- the performance level associated with a particular rating or grade on a given criterion

or dimension of achievements. For more information, see: <http://cresst96.cse.ucla.edu/teacher.htm#pa> <http://www.ed.gov/pubs/IASA/newsletters/standards/pt1.html> <http://www.ed.gov/pubs/IASA/newsletters/standards/pt2.html> <http://www.ed.gov/pubs/IASA/newsletters/standards/pt5.html>

**Standard deviation** -- a technical measure of dispersion which indicates how closely grouped or far apart the data points (e.g. test scores) are.

**Standardized testing** -- use of a common instrument to assess student levels, usually designed for use with a large number of people for ease of administration and scoring. Closed-ended questions, such as multiple choice, are often used. An individual student's score is often compared to the average performance by the group. For more information, see: [gopher://vmsgopher.cua.edu:70/00gopher\\_root\\_eric\\_ae%3A%5B\\_alt%5D\\_case2.txt](gopher://vmsgopher.cua.edu:70/00gopher_root_eric_ae%3A%5B_alt%5D_case2.txt)

**Standard score** -- a rating that is expressed as a deviation from a population mean.

**Stanine** -- one of the classes of a nine-point scale of normalized standard scores.

**Student assessment** -- evaluation of a student's progress through use of standardized or non-standardized measures. For more information, see: <http://www.ed.gov/pubs/IASA/newsletters/assess/pt4.html> <http://www.ed.gov/pubs/IASA/newsletters/assess/pt1.html>

**Student placement** -- selection of appropriate level of class or services for a student based upon standardized and/or non-standardized instruments, which sometimes will include an informal assessment.

**Teacher assessment** -- evaluation of a teacher's effectiveness conveying material, ideas, and new systems of thought to students. It requires evidence about the quality of what is taught, the amount that is learned, and the professionalism and ethicality of the teaching process.

**Test-driven curriculum** -- results when teachers begin to teach to the test in order to prepare students for the content of the test, particularly when there are high stakes attached to the results of the test. This is not necessarily viewed as a problem if the assessment measures the desired student skills. For more information, see: [gopher://vmsgopher.cua.edu:70/00gopher\\_root\\_eric\\_ae%3A%5B\\_alt%5D\\_case2.txt](gopher://vmsgopher.cua.edu:70/00gopher_root_eric_ae%3A%5B_alt%5D_case2.txt)

**Testing** -- measurement, in many cases any specific and explicit effort at performance or attitude evaluation, usually of students. For more information about fairness in testing, see: [http://www.cua.edu/www/eric\\_ae/Infoguide/FAIR\\_TES.HTM](http://www.cua.edu/www/eric_ae/Infoguide/FAIR_TES.HTM)

**Validity** -- the extent to which an assessment measures what it was intended to measure.

---

## APPENDIX B

---

## REFERENCES

Arter, J.A., and Spandel, V. (Spring 1992). NCME Instructional Module: Using Portfolios of Student Work in Instruction and Assessment. *Educational Measurement*, 11(1).

Aschbacher, P.R., Koency, G., and Schacter, J. (1995). Los Angeles Learning Center Alternative Assessment Guidebook. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing. Also: <http://cresst96.cse.ucla.edu/CRESST/Sample/GBTHREE.PDF>

Baker, E.L., Aschbacher, P.R., Niemi, D., and Sato, E. (1992). CRESST Performance Assessment Models: Assessing Content Area Explanations. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing. Also: <http://cresst96.cse.ucla.edu/CRESST/Sample/CMODELS.PDF>

Baker, E.L., Linn, R.L., and Herman, J.L. (Summer 1996). CRESST: A Continuing Mission to Improve Educational Assessment. *Evaluation Comment*. Los Angeles: UCLA Center for the Study of Evaluation.

Barnes, R.E., and Ginsberg, A.L. (1979). Relevance of the RMC Models for Title I Policy Concerns. *Educational Evaluation and Policy Analysis*. 1(2), 7-14.

Billig, S.H. (1990). For the Children: A Participatory Chapter I Program Improvement Process. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.

Brandt, R.S. (Ed.). (1992, May). Using Performance Assessment. [Entire issue] *Educational Leadership*, 49(8).

Cooper, W. (Ed.). (1994, Winter). [Entire issue] Portfolio News, 5(2).

Cordova, R.M., and Phelps, L.A. (1982). Identification and Assessment of LEPs in Vocational ED Programs: A Handbook. Champaign, IL: office of Career Development for Special Populations, University of Illinois.

David, J.L., Purkey, S., and White, P. (1989). Restructuring in Progress: Lessons from Pioneering Districts. Washington, DC: National Governors Association.

Glickman, C. (1991). Pretending Not to Know What We Know. *Educational Leadership*, 48(8).

Hansen, J.B., and Hathaway, W.E. (1993). A Survey of More Authentic Assessment Practices. Washington, DC: The ERIC Clearinghouse on Tests, Measurement, and Evaluation.

Herman, J.L., Aschbacher, P.R., and Winters, L. (1992). A Practical Guide to Alternative Assessment. Washington, DC: Association for Supervision and Curriculum Development.

Hopstock, P., Young, M.B., and Zehler, A.M. (1993). Serving Different Masters: Title VII Evaluation Practice and Policy. (Volume 1, Final Report). Report to the U.S. Department of Education, Office of Policy and Planning. Arlington, VA: Development Associates, Inc.

Linn, R.L., Baker, E.L., and Dunbar, S.B. (1991). Complex, Performance-Based Assessment:

Expectations and Validation Criteria. (CSE Tech. Rep. No. 331). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

McLaughlin, M.W. (1975). Evaluation and Reform: The Elementary and Secondary Education Act of 1965/Title I. Cambridge, MA: Ballinger.

Meier, D. (1987). Central Park East: An Alternative Story. *Phi Delta Kappan*, June, 1987.

Miles, M.B., and Louis, K.S. (1990). Mustering the Will and Skill for Change. *Educational Leadership*, 47(8).

Morris, L.L., Fitz-Gibbon, C.T., and Lindheim, E. (1987). How to Measure Performance and Use Tests. 11th printing. (Series: Program Evaluation Kit, Volume 7). ISBN: 0-8039-3132-8.

National Center for Research on Evaluation, Standards, and Student Testing. Assessing the Whole Child Guidebook. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing. Also: <http://cresst96.cse.ucla.edu/CRESST/Sample/WOLEKID.PDF>

National Center for Research on Evaluation, Standards, and Student Testing. Portfolio Assessment and High Technology. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing. Also: <http://cresst96.cse.ucla.edu/CRESST/Sample/HIGHTECH.PDF>

O'Neil, J. (1990). Piecing Together the Restructuring Puzzle. *Educational Leadership*, 47(7).

Pierce, L.V., and O'Malley, J.M. (Spring 1992). Performance and Portfolio Assessment for Language Minority Students. Program Information Guide Series, 9. Washington, DC: National Clearinghouse for Bilingual Education.

Popham, J.W. (1990). Modern Educational Measurement: A Practitioner's Perspective. 2nd edition. Englewood Cliffs, NJ: Prentice Hall.

Stiggins, R.J., and Conklin, N.F. (1992). In Teachers' Hands: Investigating the Practices of Classroom Assessment. Albany: State University of New York Press.

Stiggins, R.J., and Others. (1985, April). Avoiding Bias in the Assessment of Communication Skills. *Communication Education*, 34(2), p135-141.

Walberg, H.J. (1974). Evaluating Educational Performance: A Sourcebook of Methods, Instruments, and Examples. Berkeley, CA: McCutchan Publishing Corporation.

Wiggins, G.P. (1993). Assessing Student Performance: Exploring the Purpose and Limits of Assessment. Jossey-Bass education series, Vol. 1. San Francisco, CA: Jossey-Bass.

Wisler, C.E., and Anderson, J.K. (1979). Designing a Title I Evaluation System to Meet Legislative Requirements. *Educational Evaluation and Policy Analysis*, 1(2), 47-55.

Wolf, S., and Gearhart, M. (1993). Writing What You Read: A Guidebook for the Assessment of Children's Narratives. (CSE Resource Paper No. 10). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing. Also:

<http://cresst96.cse.ucla.edu/CRESST/Sample/RP10.PDF>