DOCUMENT RESUME

ED 440 980 TM 030 766

TITLE	Developing a Standards-Based Assessment System: A Handbook.
INSTITUTION	WestEd, San Francisco, CA.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.; Sacramento County Office of Education, CA.
PUB DATE	2000-03-17
NOTE	184p.
CONTRACT	RJ96006901
PUB TYPE	Guides - Non-Classroom (055)
EDRS PRICE	MF01/PC08 Plus Postage.
DESCRIPTORS	*Academic Standards; *Accountability; *Educational Assessment; Elementary Secondary Education; Guides; *Program Development; Program Implementation; School Districts; State Programs; *Student Evaluation

ABSTRACT

This handbook is intended as a resource for schools and districts interested in developing and implementing a standards-based assessment system. The chapters in the handbook introduce key steps in the assessment development process. Several types of on-demand and cumulative assessments are emphasized, including multiple-choice and written-response tests, projects, and portfolios. Chapter 1 discusses identifying standards that clearly define what students should know and be able to do. Chapters 2 through 4 consider developing or selecting a variety of effective assessments that together measure student performance in relation to the standards. Developing and refining an assessment scoring system is explored in chapter 5. Chapter 6 is concerned with reporting assessment results to key stakeholders. Supporting the overall development and implementation of the assessment system is reviewed in chapter 7. Information in this handbook is largely drawn from the experience of two interrelated career-technical assessment programs in California, Assessments in Career Education and the Career-Technical Assessment Program. The system developed from these programs uses a combination of written on-demand assessments and cumulative assessments that students shape and complete over a substantial period of time. Four appendixes contain examples of assessment practice and a list of assessment-related resources on developing and implementing a standards-based system. (Contains 45 tables and 37 references.) (SLD)

Developing a Standards-Based Assessment System

A Handbook

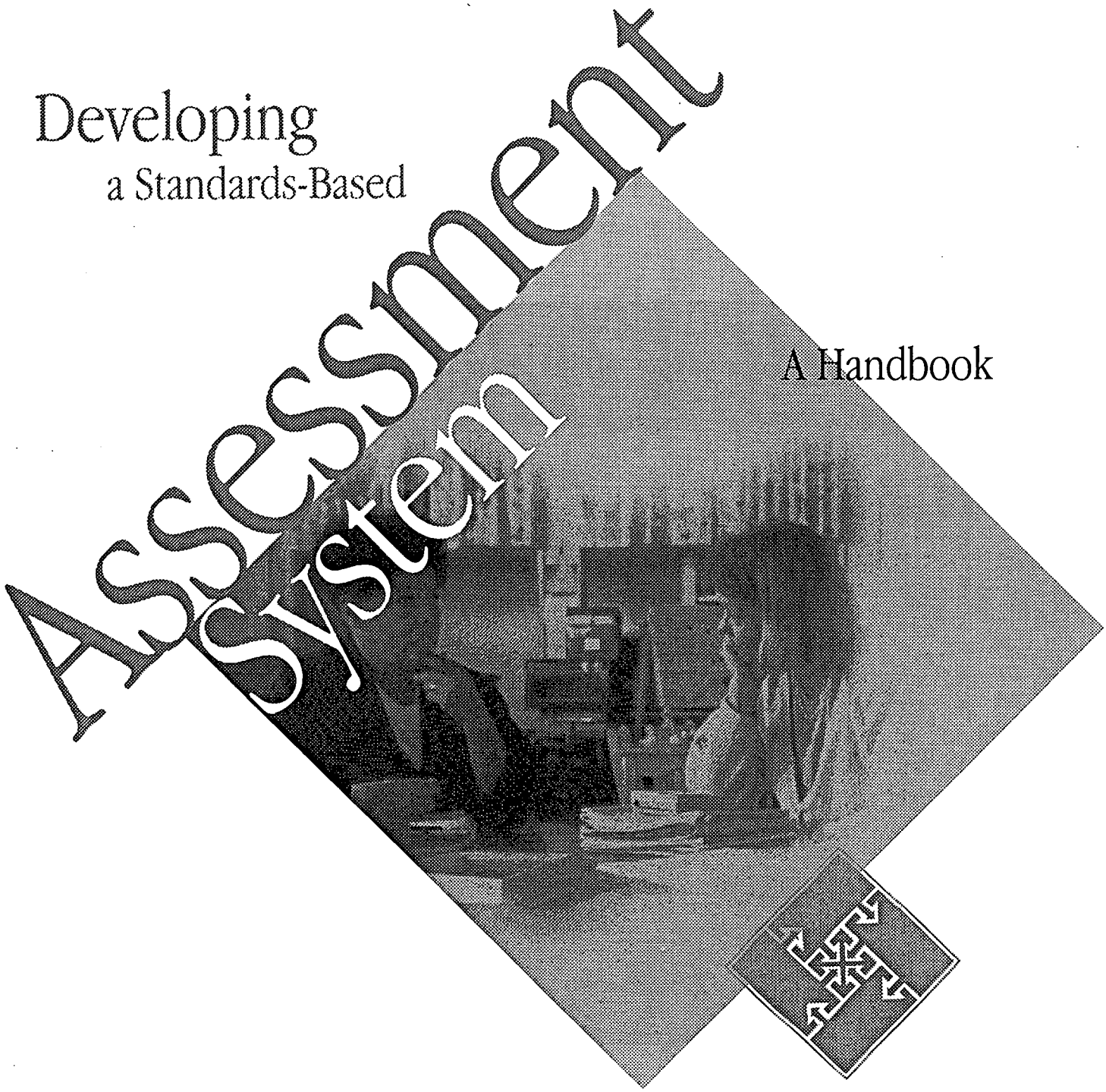
March 17, 2000

WestEd

Improving Education through Research, Development and Service

Developing a Standards-Based Assessment System

A Handbook



WestEd

Improving education through research, development, and service

The authors would like to thank the following individuals for their help in developing, reviewing, and designing this document:

Sri Ananda
Freddie Baer
Kate Jamentz
Joanne Jensen
Claudia Long
Colleen Montoya
Linda Murai
Nimfa Rueda
Mike Timms
Elise Trumbull
Burr Tyler
Joe Vera
Joy Zimmerman

WestEd is a research, development, and service agency working with education and other communities to promote excellence, achieve equity, and improve learning for children, youth, and adults. Drawing on the best knowledge from research and practice, we work with practitioners, policymakers, and others to address education's most critical issues. A non-profit agency, WestEd, whose work extends internationally, serves as one of the nation's designated regional educational laboratories — originally created by Congress in 1966 — serving the states of Arizona, California, Nevada, and Utah. With headquarters in San Francisco, WestEd has offices across the United States.

For more information about WestEd, visit our Web site: WestEd.org; call 415/565-3000 or, toll-free, (1-877) 4WestEd; or write:

WestEd
730 Harrison Street
San Francisco, CA 94107-1242

This document was developed by WestEd with the support of federal funds from the U.S. Department of Education, Office of Educational Research and Improvement, contract number RJ96006901. Additional support was provided by the Sacramento County Office of Education (SCOE). The contents of the document do not necessarily reflect the views or policies of the U.S. Department of Education or SCOE, nor does mention of trade names, commercial products, or organizations imply endorsement by either agency.

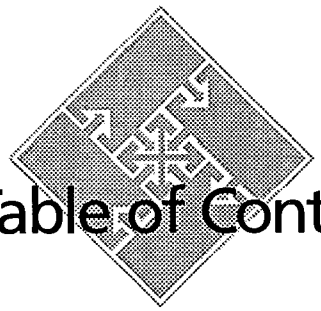


Table of Contents

Introduction ____ 1
Purpose and Goals of this Handbook ____ 2
Overview of This Handbook ____ 6

Chapter 1 Identifying Standards ____ 7
What Is a Standard? ____ 8
What Are Content and Performance Standards? ____ 9
Characteristics of Effective Standards ____ 16
Developing or Adapting Standards for Local Use ____ 18
Summary ____ 23

Chapter 2 Understanding Key Characteristics of Effective Assessments and the Importance of a Multi-Assessment System ____ 25
Characteristics of Effective Assessments ____ 26
The Importance of Using Multiple Assessments ____ 31
Summary ____ 35

Chapter 3 Developing Written On-Demand Assessments ____ 36
General Features of Written On-Demand Assessments ____ 37
Developing Written On-Demand Assessments ____ 48
Helping Students Succeed on Written On-Demand Assessments ____ 56
Summary ____ 60

Chapter 4 Developing Cumulative Assessments ____ 62
General Features of Cumulative Assessments ____ 62
Project Assessments ____ 66
Portfolio Assessments ____ 75
Challenges Associated with Developing and Implementing Cumulative Assessments ____ 82
Summary ____ 89

Chapter 5 Developing an Effective Scoring System 91
Developing an Effective Scoring System: An Overview 92
Developing a Scoring Plan 93
Drafting Scoring Scales for Assessments 102
Checking for Validity 105
Checking for Reliability 110
Choosing a Cut Score to Reflect the Performance Standard 113
Summary 114

Chapter 6 Reporting Student Achievement 117
Purposes for Reporting Student Achievement 117
Different Reporting Formats 118
Characteristics of Effective Reporting of Student Achievement 123
Combining Multiple Assessment Measures for District Reporting 127
Summary 129

Chapter 7 Supporting a Standards-Based Assessment System 130
Strengthening Organizational Support for Change 131
Developing and Implementing a New Assessment System in Phases 139
Meeting the Needs of All Students 145
Establishing Community-Wide Support 149
Coordinating Local and State Assessment Efforts 153
Summary 155

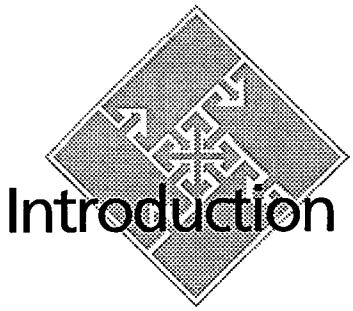
References 158

Appendix A An Example of How Student Work Can Illustrate a Performance Standard 162

Appendix B Models for Combining Multiple Measures 168

Appendix C Sample Portfolio Schedules Involving Collaboration Among Teachers 174

Appendix D Recommended Resources 177



For more than a decade, many state and local education agencies have been engaged in reform efforts aimed at ensuring that *all* students attain high levels of learning and achievement. At the core of these efforts has been a push to develop content and performance standards that clearly articulate what students should know and be able to do in various subject areas and how well they ought to perform. This new emphasis on standards is reflected in the large number of national, state, and local standards development efforts in both academic domains (e.g., National Council of Teachers of Mathematics, 1989; National Academy of Sciences, 1996) and career-related areas (e.g., 22 national skills standards projects sponsored by the U.S. Departments of Education and Labor).

Efforts to develop standards have, in turn, fueled a move toward standards-based assessment: an approach that measures students' performances against a set of common standards for learning rather than against other students' performances. Using content and performance standards as a foundation, educators in many states have been looking to develop new assessment systems that will provide a more comprehensive and valid picture of student achievement than yielded by traditional assessment systems, which have relied heavily on multiple-choice and short written-response tests. The new comprehensive assessment systems, by contrast, comprise a variety of standards-based assessments including multiple-choice/written-response tests and performance-based assessments that require students to keep portfolios of their work, solve complex hands-on problems, and plan and execute short- and long-term investigations and other projects. Together these different types of assessment aim to:

- provide informative and reliable information about student achievement vis-a-vis agreed-upon educational standards;

- engender, as well as measure, learning;
- support accurate and fair decision making about student placement in programs;
- communicate clearly to students, parents, and the community how well students are mastering the standards; and
- give teachers and districts the information needed to plan and improve curriculum, instruction, and school programs and thereby enhance student learning.

Purpose and Goals of this Handbook

This introductory handbook is intended as a resource for schools and districts interested in developing and implementing a standards-based assessment system. It is written primarily for administrators, teacher leaders, and staff developers, but may also be useful to parents, school boards, and community members who want to better understand the rationale and processes for developing and implementing a comprehensive standards-based assessment system.

The chapters in the handbook introduce key steps in the assessment development process (see Table I-1) and discuss several important issues to consider when developing and implementing a new assessment system. Several specific types of on-demand and cumulative assessments are emphasized in the chapters: multiple-choice and written-response tests, projects, and portfolios. While there are other forms of assessment that can be included in a comprehensive assessment system, this particular combination of assessments is promoted because of its ability to effectively measure the breadth and depth of students' knowledge and students' ability to apply knowledge and skills in realistic contexts.

Information in this handbook is drawn largely, although not exclusively, from experience with two interrelated career-technical assessment programs in California: Assessments in Career Education (ACE), which is administered statewide, and the Career-Technical Assessment Program (C-TAP), which is used at the local level, either districtwide, schoolwide, or in individual classrooms. The two assessment programs can be used together, as a system, to help students learn and refine important career-technical skills and to assess student readiness for entry-level jobs and postsecondary educational training.

Table I-1 Key Steps in Developing a Standards-Based Assessment System

- Identifying standards that clearly define what students should know and be able to do (Chapter 1)
- Developing or selecting a variety of effective assessments that together measure student performance in relation to the range of targeted standards (Chapters 2-4)
- Developing and refining an assessment scoring system that makes it possible to draw reliable conclusions about the degree to which students have mastered targeted standards (Chapter 5)
- Reporting assessment results to key stakeholders; for example, students, parents, teachers, school and district administrators, school board members, postsecondary admissions staff, and employers (Chapter 6)
- Supporting the overall development and implementation of the assessment system (Chapter 7)

For the purposes of this handbook, the ACE/C-TAP system offers a useful example of a comprehensive standards-based assessment system. The ACE/C-TAP system uses a combination of *written on-demand assessments* that are administered to students on specific dates under secure conditions (i.e., multiple-choice/short written-response tests and written scenarios) and *cumulative assessments* that students shape and complete over a substantial period of time (i.e., portfolio and project). All of the assessments are designed to measure student performance with respect to key standards in different career-technical programs. These standards include career preparation standards, which are common to all career-technical programs and represent general workplace readiness skills, and industry-specific content standards (i.e., industry core and career cluster standards), which identify the specific career-technical knowledge and skills to be learned in each career-technical program. The system also measures student performance with respect to the core academic skills (e.g., reading, science, mathematics) required for success in specific career-technical fields. Though the system focuses primarily on assessing career-related knowledge and skills, a close examination of each assessment within the system makes it clear that any of the assessments could be adapted for use in other content areas, including traditional academic subjects.

Table I-2 summarizes the ACE/C-TAP system. Its components will be discussed more fully throughout this handbook.

Table I-2 The ACE/C-TAP System of Assessments

ACE
(Statewide Implementation)

Multiple-Choice/Short Written-Response Exams: on-demand assessments designed to efficiently measure the breadth and some depth of students' knowledge in specific career-technical areas.

ACE tests are currently available in five career-technical content areas:

- Agricultural Core;
- Computer Science and Information Systems;
- Food Service and Hospitality;
- Health Care, Level 1; and
- Technology Core.

Additional ACE tests are under development in the following four career-technical content areas:

- Animal Science;
- Child Development and Education;
- Drafting Technology; and
- Marketing.

C-TAP
(Local Implementation)

Project: a "hands-on" cumulative assessment that requires students to plan, develop, and evaluate a product or event related to their career interests. The project gives students an opportunity to demonstrate important career-technical knowledge and skills, as well as their ability to design and create a product or event over time. The project includes four parts:

- Project Plan;
- Evidence of Progress;
- Final Product; and
- Oral Presentation.

Portfolio: a cumulative assessment requiring students to submit a collection of evidence (work) that shows important career-technical and academic knowledge and skills learned by students. It serves as a vehicle for organizing and presenting students' work for assessment purposes and for presentation to prospective employers or advanced educational training institutions. The portfolio includes four to five parts:

- Portfolio Presentation (table of contents and letter of introduction);
- Career Development Package (resume, employment or college application, and letter of recommendation);
- Work Samples (4);
- Writing Sample; and
- Supervised Practical Experience (optional).

Written Scenarios: on-demand assessments that present students with complex and realistic problems from their career-technical area to which they must respond in writing. Students are evaluated on their ability to demonstrate content knowledge, as well as on their problem-solving and communication skills.

The ACE/C-TAP system is emphasized in this handbook in part because of our in-depth understanding of this particular system, and more importantly, because the ACE/C-TAP system reflects increased attention on career preparation and the measurement of career-technical knowledge at the national, state, and local levels. Guided by overwhelming evidence that all students need more and higher quality career preparation and guidance, national, state, and local efforts are now converging to strengthen the links between school and work. Supporting these efforts is the federal School to Work Opportunities Act of 1994, which calls for school-based career education integrated with academic education; on-the-job training that is coordinated with students' school programs; and specific activities to link school- and work-based learning. In addition, the Act creates a framework and monetary incentives for state-led development of school-to-work systems that will "provide students with a foundation of academic skills and knowledge, enable them to earn portable credentials, prepare them for first jobs in high-skill/high-wage careers, and increase their opportunities for further education, including four-year colleges and universities" (California Department of Education, 1995).

In response to this Act, California created its School-to-Career Plan in 1995 and secured a large federal School-to-Career grant to help fund its efforts. California is using the federal funds to help develop a comprehensive statewide "School-to-Career" system that will help all students gain both academic and career-specific knowledge and skills, as well as the workplace readiness competencies outlined in the U.S. Department of Labor Secretary's Commission on Acquiring Necessary Skills (SCANS) report. The immediate emphasis is on helping elementary and middle school students become aware of basic career opportunities, concerns, and work attitudes, and on organizing high school instruction around career pathways that integrate academic and vocational education. A key to the success of this integrated school-to-career approach is assessment programs, like ACE and C-TAP, that can accurately capture what students are learning as they prepare for the workplace and participate in direct workplace experiences.

Overview of This Handbook

This handbook is organized into seven chapters, each of which addresses a key step in developing a standards-based assessment system.

Chapter One defines "standards," explains the difference between content and performance standards, and describes key steps to take when developing or adapting standards for local use.

Chapters Two through Four discuss the development of the individual assessments that make up an assessment system. Chapter Two explains key characteristics of effective assessments and explains the rationale for using multiple assessments to measure what students know and can do. Chapters Three and Four describe two different types of assessment, written on-demand assessments (i.e., multiple-choice and written-response tests) and cumulative assessments (i.e., projects and portfolios), highlighting examples from the ACE/C-TAP system. The advantages and limitations of these types of assessments are discussed and guidelines for development are provided.

Chapters Five and Six address issues related to scoring assessments, analyzing assessment results, and reporting student outcomes.

Chapter Seven introduces general steps that schools and districts can take to effectively support the overall development and implementation of a standards-based assessment system, including the following: strengthening organizational support for change, phasing in an assessment system over time, meeting the needs of all students, establishing community-wide support, and coordinating local and state assessment efforts.

Following Chapter Seven are four appendices. Appendix A provides an example of how student work can illustrate a performance standard. Appendix B illustrates several ways to combine scores from multiple assessments. Appendix C provides two examples of how teachers have collaborated when planning and implementing the C-TAP portfolio. Appendix D lists additional assessment-related resources that may be useful when developing, reforming, or refining a school or district assessment system.

As the name suggests, standards are the anchor for a standards-based assessment system. In the development of such a system, the identification of standards is the first step. Standards set forth expectations for what students should know and be able to do and/or expectations for levels of performance, using the identified knowledge, skills, and abilities. Two examples of standards are provided in the italics below.

As a result of activities in grades 9-12, all students should develop:

- *abilities of technological design; and*
- *understandings about science and technology.*

(National Science Education Standards, National Academy of Sciences, 1996)

Students will understand information processing concepts necessary to gather, create, and analyze data. They will perform multiple tasks required to process data effectively and produce usable information (Career Preparation Standards: Draft Interim Content and Performance Standards in Business Education — Computer Science and Information Systems, California Department of Education, 1995).

Establishing clear, rigorous standards that specify what students should know and be able to do is critical to transforming the way we educate students and assess their performance. Advocates for standards-based education reform call for "high standards for all students." They reason that setting high expectations for everyone is the first step to improving student achievement.

Standards can define shared achievement targets that can help guide curriculum development, instructional planning, and student assessment across schools. In the past, the targets focused on low-level skills and competencies that most students could easily meet. In contrast, emerging standards emphasize thinking, problemsolving, and application skills at levels beyond those achieved by most of today's students. These challenging

standards are considered essential building blocks for improved curriculum and assessments. They play a critical role in preparing our nation's workforce for successful competition in the global economy of the 21st century.

This chapter begins by defining what a standard is and then discussing two different types of standards, content and performance standards, providing several examples of each. It then identifies characteristics of effective standards and concludes with a discussion of how standards can be developed or adapted for local use. Later chapters in the document describe the role that standards play in assessment development (Chapters 3 and 4), scoring (Chapter 5), and the reporting of student achievement (Chapter 6).

What Is a Standard?

A standard is one or more statements or phrases that clearly define the knowledge and skills to be taught and/or the level of performance that is expected in a content or career area. A set of standards should represent consensus among stakeholders on what is most important for students to know and be able to do.

As such, a set of standards provides a common language for educators, students, parents, and other community members to discuss the performance of students, schools, and school districts. A standard sets a goal that can be used to guide the development of curriculum and instruction. A set of standards provides a common set of criteria that can be used to evaluate the success of individual students, schools, and school districts. Standards also provide the opportunity to forge strong links among the efforts of various stakeholders. Some of the unique benefits of standards to key constituencies are outlined in Table 1.1.

Table 1.1 **Benefits of Standards to Key Constituencies**

- Educators know the important content to be covered and can design high quality, focused programs and curricula aligned to meaningful assessments.
- Students have clear goals for their education and career preparation.
- Workers are apprised of underlying expectations for jobs and career development, enabling them to better meet employer criteria and increase their chances for mobility and advancement.
- Employers have criteria to recruit, screen, place, and evaluate potential employees more efficiently.

What Are Content and Performance Standards?

To realize the benefits outlined in Table 1.1, two different types of standards are needed: content standards and performance standards. Content standards identify *what* areas of knowledge, understanding, and skills students are expected to learn in key subject and career areas. Performance standards describe *how well* students are expected to have mastered these areas of knowledge, understanding, and skills. They define how good is good enough by identifying the levels of achievement that students must reach or exceed to meet the standard. Serving different purposes, both types of standards are essential for building an effective assessment system.

Content Standards

Content standards for a particular discipline or career area identify important knowledge, understanding, and skills to be covered in the curriculum and mastered by students. As a set, they convey a vision for learning.

The structure of content standards for career areas will be used to help illustrate different types of content standards and how the different types of standards can be interrelated. Content standards for career education students come in several different forms: *core academic standards, career preparation standards,* and *career-technical standards.* As their names suggest, each type of standard corresponds to a particular focus.

Core academic standards focus on traditional subject matter areas such as mathematics, language arts, and science, as well as other areas such as thinking skills or technology. Core academic standards identify a subject area's important concepts or thematic areas, specific skills (e.g., computation, writing), and sometimes methods of thinking and communication that characterize the subject area. The following example of a core academic standard, taken from the National Council of Teachers of Mathematics' *Curriculum and Evaluation Standards* (1989), addresses many of these aspects.

Sample Standards: Geometry from a Synthetic Perspective

In grades 9-12, the mathematics curriculum should include the continued study of the geometry of two and three dimensions so that all students can:

- interpret and draw three-dimensional objects;
- represent problem situations with geometric models and apply properties of figures;
- classify figures in terms of congruence and similarity and apply these relationships; and
- deduce properties of, and relationships between, figures from given assumptions;

And so that, in addition, college-intending students can:

- develop an understanding of an axiomatic system through investigating and comparing various geometries.

Core academic standards can serve as a framework to identify important knowledge and skills used in academic subject areas or career areas. For example, although C-TAP assessments do not specifically target academic standards, they address the academic skills required for success in a specific career-technical field. These skills vary among career areas, but include skills in writing, application of mathematical concepts, and application of biology and chemistry concepts and facts. While not a core academic standard, the following career-technical standard, taken from the *Draft Agriculture Performance Standards and Integrated Activities* (California Department of Education, 1993), addresses specific biological concepts, facts, and skills that students in Animal Science should know and be able to apply.

Sample Standards: Animal Science – Animal Physiology

Students will understand the structure, function, and maintenance of major organ systems of animals. Students will explain the interrelationships between the circulatory, respiratory, excretory, endocrine, digestive, reproductive, skeletal, and muscle systems.

Career preparation standards, or workplace readiness standards, cover generic skills and qualities that students and workers must have in order to learn and adapt to the demands of any job. They are the most general of several levels of specialized standards for workplace preparation and performance that respond to a growing awareness that along with academic preparation, students and workers need better preparation for the world of

work. Recent studies (e.g., U.S. Department of Labor Secretary's Commission on Achieving Necessary Skills (SCANS), 1991, Council of Chief State School Officers' Workplace Readiness Assessment Consortium, 1995) have identified general career preparation standards, such as those which focus on critical thinking, problemsolving, communication, or technology skills, as key to success in the 21st century workplace. In response to these studies, California has adopted a set of career preparation standards that apply across career areas as well as across specializations within career areas. Similar workplace readiness standards have also been developed at the national level. The following is an example of a career preparation standard developed by SCANS (1991).

Sample Standard: Interpersonal Skills — Works with Others

A. Participates as a Member of a Team — contributes to group effort
B. Teaches Others New Skills
C. Serves Clients/Customers — works to satisfy customer's expectations
D. Exercises Leadership — communicates ideas to justify position, persuades and convinces others, responsibly challenges existing procedures and policies
E. Negotiates — works toward agreements involving exchange of resources, resolves divergent interests
F. Works with Diversity — works well with men and women from diverse backgrounds

Career-technical standards help further prepare students for the workplace by addressing the knowledge and skills necessary for successful employment within specific occupations or industries. There are three different levels of career-technical standards:

(1) *Industry core standards* cover fundamental skills needed in nearly all the occupations within a particular industry. In the health industry, for example, core standards may cover such broad topics as infection control, working on a health care team, or fundamentals of physiology. Many career-technical programs at the high school level incorporate industry core standards in an introductory or survey course given at the ninth, 10th, or 11th grades (e.g., Introduction to Health Careers). The industry core standards ensure that the introductory or survey courses provide students with the foundation they need to decide whether to pursue additional preparation in the field.

(2) *Career cluster standards* specify the knowledge and skills needed to perform functions across a cluster or family of occupations either within a particular industry or across industries. For example, health workers, such as nurses and nursing assistants, respiratory technologists, and aides who provide direct therapeutic services to patients or clients, can be thought of as part of a cluster of therapeutic occupations within the larger health care industry. In some high schools, after a student has completed an introductory health careers course or program, he or she can specialize in a specific career cluster, such as the therapeutic cluster. A course focusing on therapeutic standards would prepare students with knowledge and skills common to a number of related occupations, rather than focusing on one specific occupation.

Given the ever-changing world of learning and work, it is important that career-related standards are of sufficient breadth to afford students some flexibility in future career and education choices. At the secondary level, instruction targets both industry core and career cluster standards to provide this broad training.

(3) *Occupation-specific standards* pertain to skills of a particular job or occupation, such as that of a medical assistant or a lab technician. Because these standards are most relevant to students who have narrowed their job interests, occupation-specific preparation is the focus of Regional Occupational Centers/Programs (ROC/Ps) or post-secondary training.

Career preparation, industry core, career cluster, and occupation-specific standards are all part of California's Model Curriculum Standards and are organized and categorized in the following way for each particular career area: 1) general workplace readiness standards, labeled Career Preparation Standards; 2) industry core standards, labeled by industry (e.g., Home Economics Related Occupations Standards); 3) career cluster standards, labeled Career Path Cluster Standards; and 4) occupation-specific standards, labeled Career Path Specialization Standards.* When development of the sets of standards is coordinated, as is the case with the California Model Curriculum Standards, each set of more specialized standards builds upon and incorporates the more generalized areas to which they relate. Figure 1.1 on the next page shows examples of industry core, career cluster, and occupation-specific standards (California Department of Education, 1996) pertaining to the area of Industrial and Technology Education. As one moves from the base to the tip of the pyramid, the standards move from more general to more specialized.

* These standards documents are currently being revised. The revised documents may include changes in the way the standards are organized and categorized.

Figure 1.1 **Relationships between Industry Core, Career Cluster, and Occupation-Specific Standards**

Occupation-Specific Standard — Carpenter
Carpentry Materials and Supplies: Students will know the names, properties, and appropriate use of materials and supplies (e.g., wood, plywood, gypsum board) used in carpentry. They will identify building materials and supplies, discuss their properties and appropriate use, and demonstrate ability to safely use the material in carpentry activities.

Career Cluster Standard — Construction
Commercial Construction: Students will understand the concepts of commercial construction (e.g., concrete forming, heating and cooling, steel framing) and how commercial structures are built. They will define terms used in commercial construction and sequence the steps involved in building a commercial structure. They will select and safely use tools and machines in a variety of commercial construction activities.

Industry Core Standard — Technology Core
Construction Technology: Students will understand planning and design (e.g., surveying and mapping, problem solving, ideation, drafting, construction plans), constructing and servicing structures (e.g., preparing for construction, foundation setting), and electro/mechanical systems and services (e.g., plumbing, electrical, HVAC) as they relate to construction activities. They will work individually or cooperatively (student teams) to demonstrate an understanding of these construction concepts through the construction of models and/or written analysis of actual construction examples.

As evidenced by the different examples of content standards in this section, there is a lack of consensus on formats for standards. Standards can take different forms and address different levels of specificity. For instance, the sample mathematics standard, *Geometry from a Synthetic Perspective,* addresses desired curricular emphases in terms of student performances. In contrast, the generic career preparation standard, *Interpersonal Skills*, provides a list of key aspects of communicating with others instead of indicating cognitive and performance expectations. The examples of career-related standards provided in Figure 1.1 are similar in format because all were developed for California's Industrial and Technology Education program.

The differences in formats among standards result from the typical pattern of development, where content standards are developed independently for separate subject or career areas. Similar formats across standards help in the identification of connections and common themes across standards and make the development of an assessment system based on multiple sets of standards a much easier task.

Performance Standards

While content standards tell us what individuals should know, *performance standards* indicate how well we expect individuals to perform. Performance standards define and illustrate levels of expected accomplishment with respect to one or more content standards. Performance standards are used for a variety of purposes, including exemplification of content standards, as well as accountability and certification (McLaughlin et al., 1995).

A performance standard that exemplifies one or more content standards adds more details to what is meant by the content standard(s) as it defines an acceptable level of performance. For example, the national New Standards Project has a content standard titled Problem Solving that identifies "Designing," "Planning and Organizing," and "Improving a System" as three key features of problem solving (National Center on Education and the Economy, 1995). A short, narrative definition of each feature is provided in the standard and is followed by a list of characteristics of each feature that need to be present in a satisfactory performance. The example on the next page shows the list of characteristics for the "Designing" feature of problem solving.

The New Standards Project document from which this example was taken also provides examples of tasks (e.g., designing, building, and racing an electric car) that can be completed by students to show mastery of the standard. In addition, pieces of satisfactory student work are included in the document to illustrate both the content and the performance standards for the broader Problem Solving standard of which "Designing a Product, Service, or System" is a part. There is growing consensus in the education community, among teachers in particular, that performance standards must include samples of satisfactory student work in order to be useful. Exemplars of student work clarify a performance standard and illustrate what high quality work might look like. In other words, they exemplify what the standard looks like in application. Appendix A contains excerpts of student work from the New Standards Project and describes connections between the work fragments and the standard.

Sample Performance Standard: Designing a Product, Service, or System

The student designs and creates a product, service, or system to meet an identified need; that is, the student:

- develops a design proposal that:
  - shows how the ideas for the design have been developed;
  - reflects awareness of similar work done by others and of relevant design standards and regulations;
  - justifies the choices made in finalizing the design with reference, for example, to functional, aesthetic, social, economic, and environmental considerations;
  - establishes criteria for evaluating the product, service, or system; and
  - uses appropriate conventions to represent the design;
- plans and implements the steps needed to create the product, service or system; and
- makes adjustments as needed to conform with specified standards or regulations regarding quality or safety;
- evaluates the product, service, or system in terms of the criteria established in the design proposal, and with reference to:
  - information gathered from sources such as impact studies, product testing, or market research; and
  - comparisons with similar work done by others.

Another purpose served by performance standards is *accountability* (e.g., the evaluation of a school or a program). For example, the scoring guide or rubric used to evaluate student responses to the C-TAP project was originally designed to serve accountability purposes by providing a summary of aggregate student performances in a school or program according to three levels of performance: Basic, Proficient, and Advanced. The number of students performing at each level can be compared to previously set expectations for adequately performing programs. Programs exceeding those expectations are potential models of curricular and instructional effectiveness. Programs failing to meet these expectations need planned improvements. Another example of performance standards used for accountability is provided by the National Assessment of Educational Progress (NAEP). This assessment system articulates rules for translating results from its assessments into student achievement categories of Basic, Proficient, and Advanced. These achievement levels address the question

"how good is good enough," with Proficient and Advanced categories indicating performance levels that are satisfactory and above. For accountability purposes, NAEP reports the percentage of students in each state at each achievement level.

A third purpose served by performance standards is *certification*. For purposes of certification, performance standards are used to determine whether an individual student has reached a certain level of achievement, such as high school graduation, based on mastery of specific standards, or whether a program has met agreed-upon criteria (e.g., in program accreditation). For both accountability and certification purposes, performance standards must be explicitly tied to an assessment system that is built from and reflects required content knowledge (i.e., content standards).

Regardless of the purpose(s) for which content or performance standards are used, their usefulness depends on the quality of the standards developed. The next section describes characteristics of effective standards.

Characteristics of Effective Standards

As the standards movement has gathered steam, the creation of standards has become a popular activity. A wide variety of standards in multiple areas has been produced, and these standards have been used to guide curriculum, instruction, professional development, and assessment development at both the local and state levels. These experiences have helped make it possible to specify some key characteristics of effective standards summarized in Table 1.2.

Table 1.2 **Key Characteristics of Effective Standards**

- Clear and easy to understand
- Focused
- Comprehensive yet manageable in number
- Inclusive of both knowledge and skills
- Linked to measurable student performances
- Reflective of high expectations for students

First and foremost, effective standards are *clear and easy to understand* so that different readers come to similar understandings as to what the standards mean. Clarity derives partly from the use of familiar terminology,

but not jargon. There exists, however, a tension in standards development between communicating to educators or practitioners within the content or career area and communicating to the general public. These two groups have quite different levels of knowledge regarding the subject matter of the standard. What may seem like jargon to outsiders may be a concise way of invoking common understandings to professionals. One way to resolve this tension is to use the New Standards Project approach of supplementing each standard with lists of relevant concepts and skills and/or examples of curricula, instruction, assessments, or student work that further illustrate the standard and communicate its meaning.

An effective standard also has an identifiable *focus*. If a standard is composed of multiple phrases or parts, all are clearly related to a single theme. There is some disagreement among educators, however, about the degree of specificity needed in the focus. The American Federation of Teachers (AFT), for example, advocates the use of very specific standards that set forth a core curriculum (see Gandal, 1996). Its representatives go so far as to argue that if the standards refer to a knowledge of war and its repercussions, that specific wars should be named to help align textbooks, teacher training, and staff development. Others, such as the National Council of Social Studies or the National Council of Teachers of Mathematics, are more open to describing general knowledge and skills and leaving the particular illustrations of that knowledge and skills up to local districts. This option enables districts to use instructional materials that are tailored to local student experiences and interests, but relies on teachers to create conceptual understandings that go beyond these localized examples.

Effective standards are *comprehensive yet manageable in number*. Developers of effective standards identify all of the most important families of concepts and skills that could serve as the focus of standards. Any knowledge or skill considered important to the content area should be related to one of the standards. When consensus on these families of concepts and skills is reached, the succinct set of standards provides an effective framework for understanding the content or career area. An effective set of standards, however, is not so large that the complete set is likely to confuse rather than clarify understanding of the content or career area. Lengthy or extremely detailed sets of standards can lead teachers to emphasize isolated facts and skills and ignore integrated applications in order to address all of the standards in the set. A lengthy set of standards, however, should not be reduced by combining dissimilar areas of content into a few standards through the use of multiple clauses or sentences. Combining dissimilar areas reduces the clarity and focus of a standard and inhibits understanding.

Effective standards are also *reflective of high expectations for students*, specifying what students need to know and be able to do in order to participate fully in society. As mentioned earlier, a consistent theme in the last decade has been increased expectations for student learning. Effective standards often suggest goals for improvement in student performances. These goals then can lead to concrete plans for achieving these improvements. If standards are set too high, however, the level of student performances relative to the standards is apt to be discouraging. In such a case, if a standard is considered attainable over an extended period of time, intermediate goals for improving student achievement can be set that eventually lead to reaching the standard. However, if upon reflection, a standard is considered to be either too ambitious or more appropriate for a later stage of student development, the expectations embodied in it should be scaled back.

Effective standards *include both knowledge and skills*. Ultimately, we are interested in what students are able to do as a result of their education. These performances draw not only upon knowledge of specific facts and concepts but also upon specific skills, such as calculating areas of geometric shapes, communicating clearly, or safely operating a particular piece of machinery that is essential in a job. Knowledge is powerless without the skills to translate it into action. Therefore, skills are as important as knowledge to include in standards.

In order to compare student achievement to standards, effective standards are *linked to measurable student performances*. All constituencies — whether they be students, parents, educators, employers, or other community members — should know how students are doing in relation to the standards and, in turn, how the educational system is performing in helping students to meet the expectations set by the standards.

Developing or Adapting Standards for Local Use

As standards are the key feature of a standards-based assessment system, this section provides a general description of the steps needed to develop or adapt standards for use at the local level. The prescribed steps, which are based on lessons learned from various national and statewide standards development efforts (including C-TAP), are summarized in Table 1.3.

Table 1.3 Steps for Developing or Adapting Standards for Local Use

- Conducting background research
- Producing draft standards using an inclusive process
- Reviewing and validating standards using multiple methods
- Refining standards through pilot testing

Each of these steps is described in general terms below. Although local efforts may be unable to fully follow all the steps indicated, awareness of the goals of each step will help inform districts of the risks involved in taking shortcuts.

Conducting Background Research

While the steps in Table 1.3 are to be followed in the development or adaptation of standards for local use, it is almost always a much easier process to adapt standards than to develop them. For this reason, the first step of any standards development process should be to research the standards and standards-related documents that have already been developed by national, state, and local sources. The goal of such research is to determine if there are existing standards pertaining to the desired content or career area, and, if there are, to consider whether these standards can be adapted for use in a locally developed, standards-based assessment system or whether new standards need to be developed.

Almost all states now have standards for specific content and career areas, and many districts do as well. If local standards are available in the desired subject or career area, they can be compared with the state standards to identify any differences in content and rigor. For example, local standards should be comparable to or greater in breadth, depth, and rigor than the state standards. All of the content in the state standards should be reflected in the local standards, although the standards may be labeled or organized differently. Local standards should have at least the same number of performance levels as the state standards, and these performance levels should also be comparable in rigor to those described by the state standards. Although local standards should cover the same content as the state standards, they can also cover additional content that reflects areas of emphasis that are important to the local community or include additional levels of student performance.

If local standards are not available, a number of other resources can be consulted as examples of different ways of thinking about the desired knowledge, understanding, and skills characterizing a subject or career area. One source is district documents. The specification of what students should know and be able to do is not a new idea to district educators. Many district documents already state expectations for students, albeit in a form different from standards. Examples include curriculum scope and sequence documents, areas of knowledge and skill reported by existing tests and assessments, as well as district mission statements. Examination of these documents is especially valuable for identifying those areas that are important to the local community but which are not included in the state standards.

Other sources of information relevant to standards development or adaptation are documents or research papers on student learning that exist in many disciplines and career areas. For example, many educational or professional agencies or associations have produced standards documents or research that identifies essential knowledge and skills to be learned in a particular subject or career area. Such documents and research can serve as sources of different ways of thinking about what is important for students to know, understand, and be able to do with respect to a particular subject or career area.

An analysis and comparison of the above documents can help inform the organization of a set of standards into content strands or career clusters. Although the specific strands or clusters will be influenced by many factors, including the structure of state standards, the requirements of postgraduate education and employment, and the dominant conceptual frameworks in the field, the research phase can help produce empirical support for one method of organizing standards over another.

Finally, upon completion of the research, the findings should be summarized by content or skill area in a format chosen for its usefulness in informing the different stakeholders who will be brought together to begin developing or adapting standards for local use.

Producing Draft Standards Using an Inclusive Process

After conducting background research, different stakeholders should be brought together to begin drafting standards. The experience from standards development efforts indicates that all key stakeholders should be involved from the beginning in defining and developing standards.

25A

Stakeholders include educators, parents, students, employers, consumers, workers, and labor representatives. Standards development efforts at the national, state, and local levels have been obstructed by disagreements over the inclusion or exclusion of specific emphases in the standards. This has led to opposition to specific sets of standards from some sections of the public. Broad representation of stakeholder groups in the development of standards, not just in their review, can surface disagreements early in the production process. Including influential representatives from key stakeholder groups in standards development also contributes to the building of public support for the resulting standards. Given the potential strength of community opposition, education reformers have come to appreciate that the public must assume ownership of the standards if the standards are to be successfully incorporated into educational and training programs.

This is not to say that coalition- and consensus-building are easy tasks. Educators and other stakeholder groups often lack a history and process for communicating among themselves (Ananda et al., 1995). In addition, standards, once developed, may serve different purposes for different groups. Educators, for example, increasingly want less prescriptive and less narrowly defined standards in order to allow the greatest flexibility for purposes of program and curriculum development. In contrast, some employers desire more specific standards that can be used as criteria for screening potential job applicants or promoting existing employees. Thus, sufficient up-front time for coalition-building and "translation" between stakeholders is often required for standards, if all constituencies are to be satisfied.

A separate committee should be convened for each set of standards that are to be developed (i.e., a separate committee for each academic or career area). Each development committee should both consist of individuals with expertise and interest in the targeted area and have representation from the different constituency groups. Within a committee, there may also be subcommittees with completely different tasks. For example, in the development of content standards, the committee as a whole might work on industry-level standards and subcommittees might work on the more specific career cluster or occupation-specific standards. After much facilitated discussion and review of the summaries of background research, each development committee produces a draft version of the desired standards, which is subject to review and validation in the next step of the standards development process.

Reviewing and Validating Standards Using Multiple Methods

To result in credible end products, the draft standards must undergo an extensive and iterative review process. One or more of the following forums for review might be used:

External Review Committee — An external review committee is a large committee that includes a cross-section of representatives from the various stakeholder groups listed previously. To maintain the independent nature of this review, committee members should not have been involved in the standards drafting process. Members discuss and individually rate the draft standards with respect to content appropriateness, clarity, and usefulness.

Survey — Surveys by mail represent a cost-efficient means of securing widespread feedback on the relevance and importance of each draft standard. This process also ensures that multiple perspectives are incorporated into the review. The efficacy of this method, however, depends on the survey return rate.

Focus Groups — Focus groups at different sites and times might be convened to solicit stakeholder input. The focus group approach is particularly useful for getting input from practitioners (e.g., teachers, workers). Relative to other validation methods, focus groups provide some unique benefits, including the rich, in-depth information that emerges when participants respond to and build upon each other's different perspectives and thinking.

Evaluations of standards and recommendations from all sources of review should be collected, summarized, and analyzed to inform revision of the draft standards.

Refining Standards After Use

Although reaching consensus on the appropriateness of a given standard is important, it is not the same as actually putting the standard to practical use. Therefore, it is critical that the standards development process not be concluded before the standards have a trial period of application, such as in curriculum redesign or assessment. For example, in developing C-TAP assessments, the creation of classroom activities aligned with the standards helped ensure the appropriateness of the standards. This trial period should include use at various schools and for different purposes. Although many

potential problems can be anticipated through careful review, other needs for revision of standards will only be identified once they are actually used. Use of standards might indicate that important knowledge and skills were omitted from the standards, that there are too many standards to achieve (suggesting a need to reexamine and prioritize the standards), or that there are better ways of stating or communicating a standard. At any rate, a process should be planned for collecting information on the "usability" of the standards and for refining them that draws upon the experience of a variety of users (e.g., teachers, program chairs, work supervisors) after a period of implementation.

Summary

Standards play a key role in a standards-based assessment system, informing the development of all other elements. Standards communicate important aspects of what students should know and be able to do. They also serve as goals in the development of curriculum and instruction, provide a common language for educators, students, parents and other community members to talk about these goals, and function as criteria by which student and school performances can be evaluated.

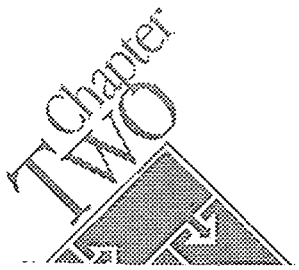
There are two general types of standards: content standards and performance standards. Content standards define the breadth and depth of knowledge and skills to be mastered by students by the time they complete an instructional program. Performance standards define and illustrate levels of expected accomplishment with respect to one or more content standards. They serve as the foundation of the scoring system used to evaluate student work.

Content standards in most career areas include three categories: core academic standards, career preparation standards, and career-technical standards. Core academic standards focus on traditional subject areas such as mathematics, language arts, or science. Career preparation standards cover generic skills and qualities that students and workers must have in order to learn and adapt to the demands of any job. Career-technical standards address the knowledge and skills that are necessary for successful employment within specific occupations or industries. Career technical standards may be further categorized into industry core standards, career cluster standards, and occupation-specific standards.

Based on the experiences of standards development efforts at national, state, and local levels, several characteristics of effective standards have been

identified. Effective standards are clear and easy to understand, focused, comprehensive yet manageable in number, inclusive of both knowledge and skills, linked to measurable student performances, and reflective of high expectations for students.

Based on the experiences of standards development efforts at both the state and local levels, four steps have been identified for developing or adapting standards for local use: conducting background research, producing draft standards using an inclusive process, reviewing and validating standards using multiple methods, and refining standards after use. While not all standards development efforts may be able to fully follow these four steps, each step is crucial to the success of the development process.



Understanding Key Characteristics of Effective Assessments and the Importance of a Multi-assessment System

Once a school or district has developed or adapted the standards that will form the basis of its assessment system, it can begin developing a set of individual assessments that, together, effectively measure student performance in relation to the selected standards. There are two general types of assessments that schools and districts can develop or use: on-demand assessments and cumulative assessments. *On-demand assessments* (e.g., multiple-choice and short written-response tests, performance tasks) are administered at a predetermined time under secure, uniform conditions. They demonstrate what students know at a particular point in time and, in the case of more complex tasks, their ability to integrate knowledge and skills in a single independent effort. *Cumulative assessments* (e.g., portfolios, projects) are typically completed over a time period ranging from days to months, and show the best work students can do when given opportunities to practice, reflect on, and revise their work in light of constructive feedback. A comprehensive standards-based assessment system will include both on-demand and cumulative assessments.

Before beginning the development of on-demand and cumulative assessments, schools and districts should first understand what makes individual assessments effective and why it is important to use multiple assessments to measure what students know and can do. This chapter describes key characteristics of effective assessments (on-demand and cumulative) and briefly explains the rationale for a multi-assessment system. The broad descriptions and points made here help lay a foundation for the more detailed discussions of on-demand assessments, cumulative assessments, and scoring-related issues in Chapters 3, 4, and 5 respectively.

Characteristics of Effective Assessments

An effective standards-based assessment system engenders learning and provides accurate and meaningful information about student achievement vis-à-vis standards. To achieve these results, such a system includes individual assessments that are themselves effective. Effective assessments share some key characteristics, a list of which is presented in Table 2.1.

Table 2.1 **Key Characteristics of Effective Assessments**

- Linked to standards
- Linked to curriculum and instruction and reflective of the most important content taught
- Cognitively complex, authentic, and integrated
- Supportive of self-evaluation and independence
- Meaningful and flexible
- Able to accommodate diversity in culture, language, cognitive/learning styles, and preferred modes of expression
- Legally defensible
- Efficient and cost effective

Not all effective assessments will exhibit all of the above characteristics. The nature of some assessments precludes some of the characteristics. For example, the structure of a multiple-choice test makes it unlikely that it will be "supportive of self-evaluation" or "flexible." The majority of characteristics presented, however, are shared by effective assessments. Thus, each characteristic is described in more detail below.

Linked to Standards

As indicated in Chapter 1, standards establish expectations for learning, outlining what students should know and be able to do (content standards) and how well they ought to perform (performance standards). If standards establish expectations for learning, then assessments should measure the degree to which students have met those expectations, or, in other words, the degree to which they have mastered the standards. To accomplish this, assessments must require students to demonstrate the specific knowledge, skills, and modes of thinking (e.g., problem solving) described by the standards. Without this direct link to the standards, it is extremely difficult

to make accurate and meaningful inferences about student achievement vis-à-vis standards.

Linked to Curriculum and Instruction

Assessments should be linked not only to agreed-upon content and performance standards, but also to the curriculum — that is, to what students are actually taught. Assessing students on content or skills that they have not been taught, or had an opportunity to learn, is unfair. Assessments must also address the *most important* concepts, skills, and thinking in the curriculum. Assessments that draw upon the heart of the curriculum are better able to provide evidence about how well students have mastered the targeted curricular domain(s).

Similarly, assessments should mirror instructional strategies that are regularly used with students. Current reform philosophy emphasizes the need to enable students to think at high levels, reason, reflect, self-evaluate, and problem solve. It is reasonable to expect students to demonstrate these skills during assessment activities only if they have had opportunities to use or develop them during classroom or work-based learning experiences. Recent studies suggest that when assessment becomes detached from instruction and learning opportunities, students are often forced to respond to tasks that are devoid of a meaningful context or familiar format (Herman et al., 1992; Koelsch et al., 1995). When this happens, students may be unable to fully demonstrate their knowledge and skills, making assessment results less meaningful and, therefore, less useful for evaluating student learning and planning further instruction.

Cognitively Complex, Authentic, and Integrated

The solution to many problems in daily life requires the integrated application of knowledge and complex thinking, reasoning, problem solving, and reflection skills. Assessments that model such real-world demands ask for more than simple recall of facts, concepts, principles, or procedures. They require students to actually apply their knowledge and skills in ways that parallel the use of knowledge and skills in real life. For example, they might include tasks that ask students to analyze, interpret, or explain cause-and-effect relationships; to identify or develop defensible hypotheses or valid conclusions; to justify ideas, methods, or procedures; to investigate and resolve realistic problems; to produce complex products or events; or to evaluate the self or others. Such assessments aim not only to measure what

students know and can do, but also to help prepare students in an authentic way for everyday living, including careers.

Some forms of assessment are inherently more cognitively complex, authentic, and integrated than others. For example, cumulative assessments (e.g., portfolios, projects) almost always require students to apply content knowledge and skills in an integrated manner to create a product or event (e.g., work samples, project). These assessments also require students to plan and organize activities, and effectively regulate their time — authentic tasks that mirror those required for success in school and the workplace.

On-demand assessments can also be cognitively complex, authentic, and integrated, although usually to a lesser degree than cumulative assessments. A written-response item, for example, can require students to evaluate a problem situation and to offer appropriate solutions (e.g., analyze an urban traffic problem and suggest appropriate solutions; analyze the symptoms of a sick cow, accurately diagnose the illness, and suggest appropriate remedies). Even multiple-choice items can extend far beyond simple recall of facts. If designed effectively, such items can reflect real-world tasks, as well as measure relatively deep levels of understanding and students' ability to think at high levels. For example, a multiple-choice item could ask students to look at several meal plans and then accurately identify the most balanced and healthy meal.

Supportive of Self-Evaluation and Independence

Most educators agree that it is important for students to become independent learners capable of continually monitoring and contributing to their own learning both inside and outside school. To do so, students must know how to reflect on and evaluate their work in order to identify their strengths and weaknesses (e.g., gaps in understanding). Assessments that require students to establish learning goals and scrutinize their progress over time help promote the development and use of the skills needed to help manage their own learning throughout their lives.

Cumulative assessments (e.g., portfolios, projects) are especially supportive of self-evaluation and independence. Such assessments typically require students to produce work (e.g., projects, work samples) independently, and then to evaluate how well their work demonstrates their strengths and abilities. To do so requires both self-reflection and self-evaluation.

Meaningful and Flexible

Whenever possible, the assessments in a system should not only measure student performance, but also help students further explore and refine important knowledge, skills, and thinking as they are applied in context. In other words, the assessments should be meaningful learning experiences in themselves. In addition, assessment tasks should be engaging, thought-provoking, and motivating. When possible, they should provide students with an opportunity to integrate their own interests and modes of learning into their assessment response, which is likely to increase students' motivation and desire to succeed.

Assessment tasks, especially cumulative activities, should also be flexible, allowing for a range of responses or performances that might demonstrate mastery of one or more standards. For example, two students may both want to complete projects that demonstrate their knowledge and skills in woodworking and construction. They may, however, choose very different topics (e.g., designing and building a guitar versus creating architectural plans and a model of a dream home) and use different processes for reaching their project goals. Both projects, if done well, can demonstrate the targeted standards equally well.

Able to Accommodate Diversity in Culture, Language, Cognitive/Learning Styles, and Preferred Modes of Expression

Differences in students' culture, language, cognitive/learning styles, and preferred modes of expression can and often do influence students' participation in the classroom and their performances on assessments. For example, many students have difficulty performing well on assessments that counter their cultural norms or require them to process information quickly in a language other than their first. Similarly, students who do not excel in verbal forms of expression are usually disadvantaged by assessments that require only language-based responses. It follows then, that assessments should accommodate differences among students, giving all students sufficient opportunities to effectively show their knowledge and skills.

While assessments must have some elements of standardization (i.e., basic requirements for completion) to ensure comparability of results, they should also provide enough diversity in task type, time allowed for completion, and opportunities for choice and support to accommodate differences in students' cultures, language, cognitive/learning styles, modes

of representation, aptitudes, and interests (Far West Laboratory, 1995; Stiggins, 1994). More specifically, assessments should allow for variation in language and cognitive and communicative styles, and, when possible, provide multiple avenues for demonstrating learning, including nonverbal forms of representation (e.g., creating "hands-on" projects; illustrating information or relationships through diagrams, graphs, or drawings).

In addition, all assessments should be written at the lowest possible reading level (i.e., using grade-appropriate vocabulary and simple, concise sentences) since the purpose of testing is to measure students' standards-related knowledge and skills, not their ability to read and interpret complex questions and instructions. Students who are still learning English should be assessed through their first language whenever possible, and time limitations for on-demand assessments should be dispensed with or modified when possible to give second-language learners adequate processing time.

Legally Defensible

An assessment must operate within governmental, legal, and professional measurement guidelines so that it can withstand legal challenges. The higher the stakes associated with assessment results, the greater the emphasis must be on legal defensibility. Basically, those developing and using an assessment must be able to show that the instrument provides information that can be used to make accurate and meaningful inferences about student achievement.

To be legally defensible it is particularly important that an assessment meet standards of validity and reliability. Validity relates to the degree to which an assessment measures what it is intended to measure. For example, an assessment that aims to assess student achievement vis-à-vis specific standards must elicit evidence of knowledge and skills directly related to those standards. An assessment is likely to demonstrate validity if both it and its scoring criteria are closely aligned with targeted standards. In addition, it is very important that a process exist for maintaining this alignment over time.

Reliability relates to the consistency of test results, or the degree to which students' assessment performances and the scores on those performances are replicable over time and across different circumstances. An assessment is likely to meet standards of reliability if:

- students who complete the assessment perform similarly when they complete the same assessment a second time or when they complete a similar, yet different, assessment (e.g., different questions/tasks measuring the same standards at the same level of difficulty); and

- the same performance on an assessment receives the same, or a sufficiently similar, score no matter who scores it or when it is scored.

Both validity and reliability are discussed in greater detail in Chapter 5.

Efficient and Cost Effective

At a time when resources for educational endeavors are shrinking and there are competing demands for existing funds, schools and districts must be concerned with the economic feasibility of the assessments they use. The implementation of any assessment by a school or district should not result in unbearable costs. The costs associated with an assessment depend in large part on the efficiency of the assessment itself, which is influenced by a number of factors, including:

- the amount of content (e.g., number and range of standards) that can be covered by the assessment;
- the time and effort needed to administer the assessment (including the level of support and feedback that must be provided to students throughout the assessment process); and
- the ease and speed with which student responses to the assessment can be scored.

Generally speaking, the efficiency, and therefore cost effectiveness, of an assessment increases the more content it covers, the less time and effort it takes to administer, and the easier and quicker it is to score. On-demand assessments such as multiple-choice tests tend to be the most efficient and cost-effective assessments.

The Importance of Using Multiple Assessments

The basic structure of every assessment affects the extent to which the assessment can incorporate the key characteristics described above. For example, all assessments (i.e., on-demand, cumulative) can be linked to standards, curriculum, and instruction, but different types of assessments vary in the degree to which they can be cognitively complex, support self-evaluation, accommodate diversity, or achieve efficiency and cost-effectiveness.

Consider the following examples that relate to an assessment's ability to achieve efficiency and reliability, measure complex achievement, engender learning, and accommodate diversity.

- **Achieving Efficiency (e.g., comprehensive coverage of standards):** Some assessment structures make it possible to quickly measure a greater number and wider range of standards than others, making them more efficient means of assessment. For example, assessments that include numerous questions that can be answered in a short period of time (e.g., multiple-choice tests) are likely to be more efficient than those with a limited number of complex and very focused tasks that take weeks or even months to complete (e.g., portfolio, project). The former type of assessment allows for more comprehensive coverage of standards, making it possible to quickly gauge the breadth of students' knowledge and skills.

- **Achieving Reliability:** While efforts must be made to ensure the reliability of all assessments, some assessment structures can achieve reliability easier than others. For example, assessments that ask students to select the "right" or "best" answer from several options (e.g., multiple-choice tests) are likely to be more reliable than assessments requiring students to construct their own answers or products. Assessments such as multiple-choice tests can be evaluated objectively and consistently since no human judgment is required to determine the quality of each answer (i.e., an answer is either right or wrong). In addition, machines can even be used for scoring, which also makes the evaluation process cost effective.

- **Measuring Complex Achievement:** While all assessments can be designed to measure some degree of complex achievement, some assessment structures are able to measure certain forms of complex achievement (e.g., ability to successfully apply knowledge and skills) more effectively and fully than others. For example, assessments that involve "hands-on" activities (e.g., performance tasks, projects, portfolio work samples) are likely to provide more accurate information about students' ability to successfully apply knowledge and skills in realistic contexts than are assessments that require students to simply choose the correct answers to questions (e.g., multiple-choice tests) or to write short responses to questions (e.g., written-response tests).

- **Engendering Learning:** Similarly, assessment methods that take place over substantial periods of time and provide opportunities for revision and improvement along the way (e.g., cumulative assessments such as

portfolios and projects) are more likely to promote self-reflection and self-evaluation and hence engender learning than are timed tests that require students to demonstrate what they know "on the spot" (e.g., on-demand assessments such as multiple-choice and written-response tests).

- **Accommodating Diversity:** Some assessments are better able to accommodate diversity than others because of their formats. For example, assessments such as portfolios or projects typically have a flexible format that allows students to decide how best to meet the assessments' basic requirements. In contrast, assessments such as multiple-choice tests have a format that strictly limits the ways in which students may respond. An assessment with a flexible format is more likely to accommodate diversity in cognitive/learning styles, preferred modes of expression, cultures, and linguistic backgrounds than an assessment with an inflexible format.

As these examples help illustrate, each assessment type has advantages and disadvantages that should be considered when selecting assessments for implementation. While one type of assessment may be very effective for measuring complex achievement and engendering learning (e.g., a project that requires the integration and application of knowledge and skills over time), it may cover fewer standards and be less reliable than other types of assessment. Conversely, the type of assessment that may be very efficient and reliable (e.g., a multiple-choice test that covers a broad range of standards and can be scored quickly and objectively) may limit the kinds and depth of knowledge and skills that can be measured and the variety of ways in which students can respond.

Because all assessments have some disadvantages, no one assessment alone can provide a comprehensive view (e.g., breadth and depth, recall of knowledge and application of knowledge) of what students know and can do. To develop such a view, and to ensure that students have some opportunities to refine knowledge and skills and deepen understanding through assessment experiences, it is necessary to use a multi-assessment approach. A multi-assessment approach uses a variety of different types of assessments at different points in time, using each type of assessment for the purposes to which it is best suited. For example, a school or district using the multi-assessment approach might at one point in time use a multiple-choice test to measure students' breadth of knowledge related to a variety of standards and, at another point in time, a project assessment for measuring depth of knowledge and hands-on application of knowledge and skills related to one or two standards. The multi-assessment approach allows districts to select

assessments in such a way that the strength of one assessment type helps compensate for the weaknesses of another type and vice versa.

Table 2.2 summarizes the key benefits of using multiple assessments to measure student achievement.

Table 2.2 Benefits of Using Multiple Assessments

- Allows for sufficient coverage of all targeted standards
- Makes it possible to assess both the breadth and depth of students' knowledge
- Makes it possible to measure students' knowledge and skills as well as their ability to apply that knowledge and those skills in realistic contexts
- Helps ensure that all students have sufficient opportunities to demonstrate what they know and can do (i.e., by accommodating diversity in cultures, linguistic backgrounds, cognitive/learning styles, modes of representation)
- Helps ensure that students have some opportunities to refine knowledge and skills and deepen understanding through assessment experiences

The ACE/C-TAP system is an example of a system that uses the multi-assessment approach to provide a comprehensive view of student achievement. The ACE/C-TAP system includes multiple-choice and written-response tests (ACE) and portfolio, project, and written scenario assessments (C-TAP). Each assessment type has been thoughtfully chosen to maximize the utility of assessment information for students, teachers, districts, parents, and potential employers or receiving schools. All of the assessments are complementary, and each focuses on different foundational skills. For example, the ACE multiple-choice tests are used to measure students' overall breadth of cluster-specific career-technical knowledge. The ACE short written-response questions and the C-TAP written scenarios are used to probe the depth of students' understanding in relation to targeted content, as well as students' ability to analyze information and pose written solutions to realistic problems. C-TAP projects are used to assess in-depth knowledge and skills related to one or two standards and students' ability to apply that knowledge and those skills and to plan, organize, and implement a project over time. C-TAP portfolios are used to assess students' skill in writing, reflection, and self-evaluation, as well as their standards-related knowledge. Together, the different assessments in the ACE/C-TAP system capture the breadth and depth of student learning in a range of ways and present a rich depiction of student achievement that no one method of assessment alone could do.

Summary

An important first step in developing the assessments that will comprise an effective standards-based assessment system is understanding what makes individual assessments effective and why it is important to use multiple assessments to measure student achievement. To help ensure that an assessment system engenders learning and provides accurate and meaningful information about student achievement vis-à-vis standards, efforts should be made to include assessments that are themselves effective. To the extent possible, each assessment in the system should be 1) linked to standards; 2) linked to curriculum and instruction and reflective of the most important content taught; 3) cognitively complex, authentic, and integrated; 4) supportive of self-evaluation and independence; 5) meaningful and flexible; 6) responsive to differences in culture, language, cognitive/learning styles, and preferred modes of expression; 7) legally defensible; and 8) efficient and cost effective.

Because of their different structures, assessments vary in the degree to which they are able to incorporate the key characteristics of effective assessments listed above. As a result, no one assessment alone can provide a comprehensive view (e.g., breadth and depth, recall of knowledge and application of knowledge) of what students know and can do. Instead a multi-assessment approach should be used to measure student achievement. A multi-assessment approach uses a variety of different types of assessments at different points in time, using each type of assessment for the purposes to which it is best suited (e.g., a multiple-choice test is used to assess breadth of specific knowledge; a project assessment is used to measure application of specific knowledge and skills). The use of multiple assessments helps ensure that all students have adequate opportunities to demonstrate both the breadth and depth of their knowledge and skills.

Developing Written On-Demand Assessments

Once schools or districts have developed or adapted standards for assessment, and have a general understanding of the characteristics of effective assessments and the rationale for a multi-assessment approach, they are ready to develop or select the assessments they will use to measure student achievement. The two types of assessments that schools and districts are likely to include in their assessment system are written on-demand assessments (e.g., multiple-choice and written-response tests) and cumulative assessments (e.g., portfolios, projects). This chapter provides information about written on-demand assessments, Chapter 4 discusses cumulative assessments, and Chapter 5 focuses on scoring systems for assessments.

While not all schools or districts actually develop their own written on-demand assessments (e.g., multiple-choice and written-response tests), most schools and districts choose to include such assessments in their assessment system, in part because they are readily available (especially in academic subject areas) and have been used for decades. Written on-demand assessments are also efficient and cost-effective means of measuring students' knowledge, factors that make them especially popular for wide-scale implementation purposes (e.g., at the national, state, and district levels).

This chapter begins with a discussion of the general features of written on-demand assessments, including a detailed description of the format (structure and uses) and advantages and disadvantages of two specific types of written on-demand assessments: multiple-choice and written-response tests. Following this information is a description of the key steps involved in developing written on-demand assessments. Even if a school or district chooses not to develop its own written on-demand assessments, this chapter can help educators better understand the usefulness of such assessments and the process used to develop them.

40

General Features of Written On-Demand Assessments

Written on-demand assessments share several features that distinguish them from other types of assessments (i.e., cumulative assessments). For example, all written on-demand assessments measure students' knowledge and skills at a particular point in time (e.g., at the end of a unit or course of study; prior to placement in a particular educational program). In addition, they are typically administered under uniform conditions and within a specified time period. Some examples of written on-demand assessments and their allotted time periods are a sixth-grade spelling test (10 minutes), a class mid-term (45-50 minutes), and the Scholastic Aptitude Test (3-4 hours).

The results of written on-demand assessments can be used both formatively and summatively. For example, an end-of-unit classroom test may give a teacher information about how well students have mastered content, as well as point to concepts that may need to be reviewed again or taught differently. Written on-demand assessments that are standardized and implemented on a wide-scale basis (e.g., a statewide achievement test) can be used to measure how well students have mastered content and to evaluate the effectiveness of educational programs and institutions.

While all written on-demand assessments share the features described above, each of the two types of written on-demand assessments discussed in this chapter has specific distinguishing features. Multiple-choice tests, for example, are able to assess a broad range or breadth of standards-based knowledge. Written-response tests are able to probe students' depth of knowledge in relation to a select number of targeted standards. For this reason, some assessment systems (e.g., California's Golden State Examinations, Assessments in Career Education) include both types of written on-demand assessments so that together the assessments can measure both the breadth and some depth of student achievement. Combining these types of assessments also makes it possible to measure relatively simple aspects of achievement (e.g., basic recall of factual information) and more complex forms of learning (e.g., ability to apply knowledge to solve problems or develop ideas). This will be discussed in more detail later in the chapter.

Table 3.1 summarizes the general features of written on-demand assessments.

Table 3.1 General Features of Written On-Demand Assessments

- Measure students' knowledge and skills at a particular point in time (e.g., at the end of a unit or course of study)
- Are administered under uniform conditions within a specified time period (e.g., minutes, hours)
- Produce results that can be used formatively and summatively
- Can be used to assess the breadth of students' standards-based knowledge (e.g., multiple-choice test) and/or to probe the depth of students' knowledge in relation to a select number of standards (e.g., written-response test)

Additional features specific to multiple-choice tests and written-response tests are discussed next, with a focus on the format (structure and uses) and advantages and disadvantages of the two different types of assessments.

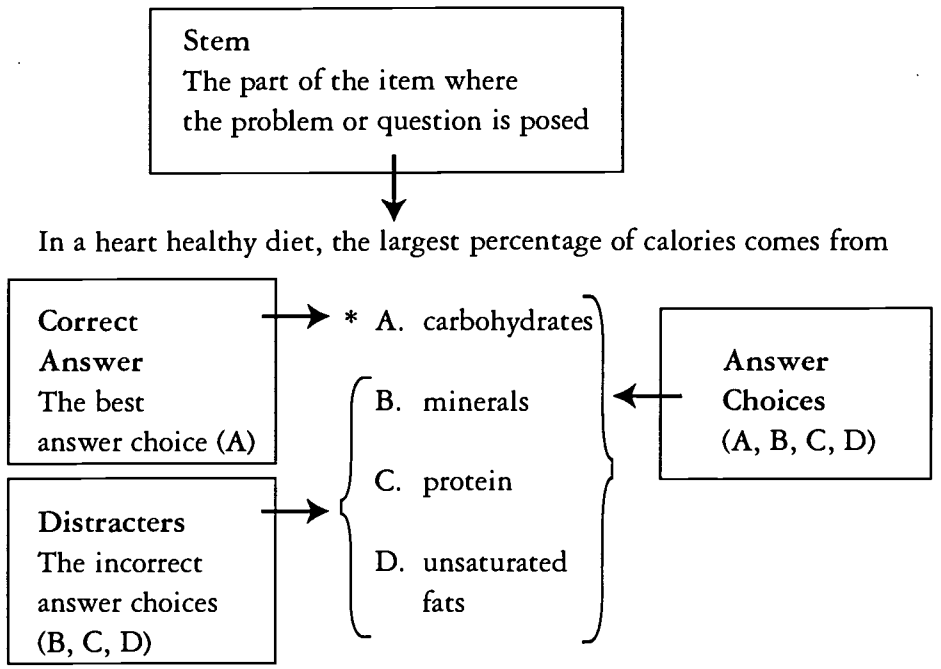
The Multiple-Choice Format: Structure and Uses

A multiple-choice assessment is a collection of selected-response items, each of which presents students with a highly-structured question and four or five possible answer choices. Students are asked to select the correct or best answer from the available choices.

Table 3.2 shows the "structure" of a typical multiple-choice item (taken from sample ACE items in the area of Health Care). The *stem* presents a question or problem that is solved by one of the *answer choices* listed below it. One of the choices is the *correct answer*. The other alternatives are called *distracters,* or incorrect answer choices, which are plausible yet unquestionably wrong or weak response options.

Multiple-choice items can be used to measure a range of basic achievement. They can measure basic recall of facts, concepts, principles, and procedures. They can also be designed to tap deeper levels of understanding, assessing how well students think, reason, and even problem solve. For example, multiple-choice items can assess whether students can identify correct applications of facts and principles, accurately analyze and interpret relationships, and select appropriate justifications for methods and procedures. Series of complex multiple-choice items that use written passages or graphics (e.g., charts, graphs, photos, drawings, figures, maps,

Table 3.2 Structure of a Typical Multiple-Choice Item



tables) can be used to gauge a host of evaluative capacities, among them the degree to which students can recognize and state inferences, recognize the relevance of information, develop and identify tenable hypotheses, formulate and recognize valid conclusions, recognize assumptions underlying conclusions, recognize the limitations of data, recognize and state significant problems, and design experimental procedures (Gronlund, 1985).

The sample multiple-choice items shown on the next page help illustrate how such items can be designed to measure 1) simple recall of memorized information; or 2) deeper levels of understanding, high-level thinking, and evaluative abilities. Both questions (taken from sample ACE items in the area of Food Service and Hospitality) aim to assess students' knowledge of recommendations outlined in the Food Guide Pyramid. The first sample item asks students to recall a specific recommendation, while the second requires students to use their knowledge of the Food Guide Pyramid to select the most balanced and healthy menu from several choices.

Sample Multiple Choice Item 1

According to the Food Guide Pyramid, how many servings from the "bread, cereal, rice and pasta" group should an individual eat daily?

A. 2-3 servings
B. 2-4 servings
C. 3-5 servings
* D. 6-11 servings

Sample Multiple Choice Item 2

Which menu is most healthy and includes foods from each major food group in the Food Guide Pyramid?

A. Hamburger with lettuce, tomato, and mustard
   Potato chips
   Carrot sticks
   Sliced watermelon
   Diet soda
B. Spaghetti with meatballs
   Garlic bread
   Fruit salad
   Ice cream
   Iced tea
* C. Chicken breast sandwich
   Mixed green salad
   Pretzels
   Frozen yogurt
   Orange juice
D. Stir-fry chicken with peanuts
   Rice
   Mixed vegetables
   Almond cookie
   Milk

Some Advantages of the Multiple-Choice Format

The multiple-choice format offers an efficient means of quickly assessing students' basic subject matter knowledge. Because multiple-choice items are usually brief and can be answered relatively quickly, many can appear on a single test form. Although each item samples only an isolated bit of student

learning, a set of such items can provide a reliable index of the breadth of a student's knowledge in a subject area. In addition, the highly structured format of the items (i.e., the student must select one answer from a limited set of possible choices) makes the scoring process relatively simple, fast, and accurate. Responses can be quickly judged as either "right" or "wrong" by a teacher or even a machine. This particular feature has made multiple-choice assessments one of the most efficient and cost-effective methods of testing.

Another advantage of the multiple-choice format is that of clarity. The multiple-choice format is often preferred over other selected-response formats (e.g., true/false and matching questions) and some short written-response formats (e.g., fill-in-the-blank questions) because of its potential for clarity in both the question and response. Consider the fill-in-the-blank item below:

*Clara Barton founded*_____. *(answer: the Red Cross)*

Even a seemingly simple question like this can be misinterpreted by students and answered in more than one way. For example, when reading the question, a student might think of a type of patient care associated with Barton (e.g., the practice of attending to wounded soldiers) and fill in the blank accordingly. His or her answer might be technically correct, but not the answer intended by the item developer(s). Similarly, an English language learner might be unfamiliar with the term "founded" and interpret it as meaning the past tense of "found." He or she might respond that "Clara Barton founded *the care of wounded soldiers poor*." Using the multiple-choice format, the same item can be framed and answered unambiguously:

Clara Barton founded the

A. Blue Cross.

* B. Red Cross.

C. White Cross.

D. Blue Shield.

Unlike some forms of selected-response items (e.g., true/false questions), multiple-choice items do not require absolutely definitive answers. Students can be asked to select the right or "best" answer depending on the nature of the question and possible answers. The "best answer" format may include answer choices that are all correct to some degree, with one answer being the best or most appropriate choice. In this way, the multiple-choice format has potential for tapping deeper levels of understanding because it requires students to make finer distinctions among answer choices.

Finally, the results of multiple-choice tests can be used to inform teaching. Wrong answers by students can be used as clues to their misunderstanding and need for additional instruction. For example, when many students choose the same incorrect answer choice for a particular item, a teacher may be able to identify an important idea or concept that needs to be reviewed or taught in a new way.

Some Disadvantages of the Multiple-Choice Format

Because multiple-choice items are highly structured and focus on isolated bits of information, they make measuring certain types of complex achievement difficult (e.g., the ability to generate, develop, integrate, and express ideas; the ability to plan and organize writing). In addition, while multiple-choice items can be used to measure students' ability to apply knowledge, they are somewhat limited in what they can do in this regard. For example, multiple-choice items can require students to recognize appropriate applications of various facts, principles, or procedures in relation to a particular problem situation, but the application is done hypothetically. Students' answers will indicate whether they know, in an abstract sense, what to do in the situation, but cannot show how well the students would perform if actually in such a situation. This does not mean multiple-choice items should not be used, only that they should not be the sole means of assessing students' ability to apply what they know.

Another limitation of multiple-choice items is their heavy reliance on reading skills. By design, multiple-choice items are meant to be brief and concise, including only information and words that are absolutely necessary for interpreting the question. Although succinct multiple-choice items may require less reading, some believe that such items may actually be more difficult for students with limited proficiency in English. Because the items include only the bare essentials, there are fewer clues available (e.g., extra words or phrases that students might recognize; explanatory information that places each question in some understandable context) to help these students make sense of what is being asked. Thus, a multiple-choice item may demand a surprisingly high level of skill with regard to both literacy and language. As an example of how a multiple-choice item may demand a high level of proficiency in English, consider the following multiple-choice items that focuses on math problem solving.

Sample Multiple-Choice Items

Molly is shopping at Superfoods and has a 50-cent double coupon for a six-pack of 12-ounce Double Bubble Cola priced at $1.49. She could also buy the same Double Bubble Cola in a 32-ounce container for $0.79. If she bought two 32-ounce containers, she could use the same coupon.

1. What is the cost of the six-pack when applying the double coupon?
   - A. $1.49
   - B. $0.99
   - * C. $0.49
   - D. Cannot tell from the information given

2. What is the cost of the two 32-ounce containers when applying the double coupon?
   - * A. $0.58
   - B. $0.79
   - C. $1.58
   - D. Cannot tell from the information given

From Haladyna, 1994, p.104.

Not only does the first sentence require knowledge of mathematics concepts (e.g., money, liquid volume), it also requires that students understand the concepts of six-packs and double coupons (i.e., what they are and how they work), concepts that may be unfamiliar to students from some cultures. In addition, the item's stem includes many ideas in very dense syntax. If we break the first sentence apart, for example, we find it includes several ideas and much information. A proficient reader of English can more quickly take the syntax apart and discount the unnecessary information (e.g., where Molly was shopping and what the product was) to solve the problem than can a person who is not yet fluent in English. In addition, time limitations penalize students who are processing in a second language. Some assessment specialists (e.g., Stiggins, 1994) suggest that time constraints for written on-demand assessments be reduced and that students be allowed ample time to complete such assessments. Such a move puts greater emphasis on mastery and less on speed.

Finally, effective multiple-choice items are not easy to write. In particular, generating enough plausible distracters (incorrect answer choices) can be a challenge. Distracters are intended to divert the attention of students unsure of the correct answer. When distracters are not plausible, students in doubt are more likely to guess correctly or easily eliminate one or more incorrect answer choices. This greatly reduces an item's effectiveness.

The Written-Response Format: Structure and Uses

In contrast to the multiple-choice format, the written-response format requires students to supply (or construct) their own answers to test questions. These answers may be as short as one word (e.g., fill-in-the-blank) or as long as an extended essay. This handbook focuses on written-response items that require answers of between several paragraphs (i.e., short written-response items) and several pages (i.e., long written-response items).

The written-response format is designed to measure students' depth more than breadth of knowledge, and their ability to manipulate such knowledge in relatively complex ways. Typically, a written-response item requires the application of knowledge, often asking students to pose written solutions to realistic problems. Students must not only recall knowledge, but be able to use the information to carry out a range of complex cognitive behaviors, such as organizing, summarizing, classifying, comparing, relating, analyzing, synthesizing, evaluating, generalizing, inferring, predicting, concluding, applying, solving, and/or creating.

As mentioned above, there are short and long written-response items. Short written-response items tend to be limited in scope, focusing only on one or two content standard(s), and using a simple, straightforward question format with few variables. They can usually be answered in two or three paragraphs within 10-15 minutes. Short written-response tasks are used in ACE assessments in conjunction with multiple-choice items. An example of the structure of a short written-response item is shown in Table 3.3.

Table 3.3 **Structure of a Short Written-Response Item**

Item Name	"Selecting a Preschool Program"
Prompt	Your neighbors are trying to decide how to educate their young child — at home or in a preschool. Because you work at a preschool, they have come to you for advice.
Instructions	A. Explain in detail two different reasons why attending preschool can be valuable for a child. B. Describe two different types of preschool programs the parents could consider for their child. Make sure your description includes information about the basic philosophy and key characteristics of each program.

In contrast to short written-response items, long written-response items tend to be broad in scope, covering several standards, and using a complex question format with multiple variables. These items usually take at least one full class period to answer and sometimes longer depending on the context. An example of a long written-response item is a C-TAP written scenario. A written scenario, like many long written-response items, presents a "real life" problem for students to solve "on demand." Students are required to read the scenario, think about possible solutions, organize their thoughts, and propose a solution in writing within a 45-minute time period. Long written-response items such as C-TAP written scenarios are not assessments of memorized knowledge and rote skills. Rather, they elicit students' ability to apply knowledge, interpret information, and explain ideas clearly. A sample C-TAP written scenario in the area of Agricultural Education is presented in Table 3.4. Long written-response items like the scenario shown in Table 3.4 are frequently used in assessment systems that include multiple-choice items, but they can also be used as stand-alone assessments.

As Table 3.3 and Table 3.4 illustrate, the formats of well-developed short and long written-response items contain some common features, such as an item name or title, a prompt, and instructions. The item name or title identifies the item with a word or short descriptive phrase related to the prompt. The prompt provides background information on the setting and context of the item, usually describing a problem or situation to be considered. This information is meant to set the stage for writing and to capture students' interest in the topic. As the name implies, the instructions tell students what to do. They convey the nature of thinking and writing required (e.g., evaluation, analysis) and clearly outline the specific "question(s)" to be answered and aspects of content to be considered when responding. Long written-response tasks often include evaluation criteria that clearly articulate what students must demonstrate to receive a satisfactory (e.g., Proficient) rating.

Table 3.4 Structure of a Long Written-Response Item
(based on the C-TAP written scenario model)

TITLE	**THE SICK CALF**
Prompt	You are asked by your neighbors to look at their sick calf. You arrive and observe that the calf looks unhealthy. It has a dull coat, watery eyes, little appetite, and is scouring. Its pen is muddy and has no dry areas. There is no shelter in the pen. Your neighbors don't have much experience with cattle. They do not want to call in a veterinarian because of the cost. As far as you know, the calf has been given no medication or vaccinations.
Instructions	Consider what you know about animal science. Prepare a list of recommendations for your neighbors to improve and maintain the health of the calf. Give reasons for what you suggest.
Evaluation Criteria	To receive a Proficient rating on this task, you must show all of the following: 1. Knowledge of ■ Animal health ■ Animal parasites and pests ■ Animal nutrition ■ Animal facilities, equipment, and handling 2. Ability to propose a solution to this scenario 3. Ability to communicate effectively in writing

Some Advantages of Written-Response Items

Written-response items are a good complement to multiple-choice items and assessments that are performance-based. They can measure forms of complex achievement that multiple-choice items cannot; for example, students' ability to generate, develop, integrate, and express ideas and their ability to plan and organize writing.

Written-response questions can also provide a valid and relatively cost-efficient means of measuring some of the prerequisite knowledge and cognitive behaviors needed to actually perform a variety of complex skills and activities. For example, it is possible to get a sense of students' readiness to build a chair that meets certain specifications by asking them to write about the various steps they would take to complete such a task and why

each step is important. Their responses would indicate their ability to recall relevant furniture making and construction knowledge and to discuss appropriate procedures for building chairs. Of course, this application is still somewhat hypothetical. A performance-based assessment requiring a student to construct a real chair would still be the only way to see whether students could successfully employ relevant content and procedural knowledge to actually build a chair.

Requiring students to supply, rather than select, answers to test questions can be advantageous for several reasons. First, written-response items reduce the risk that unprepared students will be able to guess correctly. While some educators (e.g., Haladyna, 1994) suggest that guessing is not as big a danger as some might think, it is still a factor to consider when measuring student achievement. When faced with a typical multiple-choice item, students without a firm grasp of necessary content knowledge and skills have a 1 in 4 (or 1 in 5) chance of guessing the correct answer. With written-response items, such guessing is impossible. Students must consider a problem situation and provide their own solutions based on what they know about the topic at hand.

In addition, students' answers to well-constructed written-response items can be instructionally informative, more so than answers to multiple-choice items. For example, items that require students to explain the rationale for their response can provide a window into their mental processes. Careful review and evaluation of their explanations can often reveal gaps in content knowledge, misconceptions, or weaknesses in reasoning and problem-solving skills. The information gathered from this review can be used for planning additional or future instruction. This is not always possible with multiple-choice items because reasoning and problem-solving processes are not usually revealed in student answers. With multiple-choice items, a teacher may know a student answered incorrectly, but may not know exactly why, which makes altering instruction more difficult.

Some Disadvantages of Written-Response Items

Written-response items are less efficient than multiple-choice items. Because most written-response items take more time to answer than multiple-choice items, fewer can be included on a test, thereby limiting the range of content knowledge that can be assessed at one time. For this reason, it is best to reserve the use of written-response items for measuring aspects of learning not well-tapped by multiple-choice items or for content that students should know in more depth.

Another disadvantage of written-response items is that they are not easy to score. Unlike answers to multiple-choice items, answers to written-response questions cannot be quickly judged as either right or wrong. Rather, evidence of what students know and can do must be gleaned from the content of their writing (e.g., in the ideas they express, the concepts they explain, the reasons they provide, the way they tie ideas together). Their responses must be read and judged by people, not machines, and judgments must be based on the overall quality and the level of standards-based mastery demonstrated. To make this possible, criteria for evaluating students' responses must be clearly articulated, and then scorers, whether they be teachers or other trained educational professionals, must be taught how to use these criteria to make fair and uniform judgments. Developing scoring criteria and training people to use them effectively takes time. Moreover, when scoring criteria are unclear, or are applied improperly or inconsistently, scoring results can be unreliable. The concept of reliability in scoring is discussed in more detail in Chapter 5.

Developing Written On-Demand Assessments

As just indicated, both multiple-choice and written-response items have advantages and disadvantages. When used together in an assessment system, however, the advantages of one help to compensate for the disadvantages of the other and vice versa. For this reason, the key development steps described in this section relate to the process of developing written on-demand assessments that contain both multiple-choice and written-response items.

Table 3.5 summarizes some of the key steps involved in developing written on-demand assessments. Each step is discussed in more detail following the table.

Table 3.5 Key Steps in Developing Written On-Demand Assessments

- Creating a test blueprint that clearly outlines key characteristics desired in the final assessment (e.g., the specific standards-related content to be addressed, the relative emphasis to be given to each area of content covered, the type of cognitive performances desired of students, the specific item/question format(s) to be used, the overall level of difficulty desired)
- Developing a variety of multiple-choice and written-response items, always keeping the test blueprint in mind (e.g., developing items for each specification outlined in the blueprint)
- Assembling items into a draft assessment that meets the requirements outlined in the test blueprint
- Reviewing the assessment prior to classroom tryouts to ensure that all items adhere to effective item development guidelines, are clearly linked to targeted standards, and are free of grammatical errors, inconsistencies, and bias
- "Trying out" the assessment with representative samples of students (i.e., through informal classroom tryouts and more formal, large-scale field tests) and statistically analyzing the test results (student response data) to identify items that do and do not work as intended
- Refining the test based on insights gleaned from classroom tryouts, field tests, and subsequent item analysis

Creating a Test Blueprint

As mentioned in Chapter 2, the most important characteristic of any standards-based assessment system is the inclusion of individual assessment items and tasks linked to standards that clearly define what students should know and be able to do. To ensure that written on-demand assessments are tied to standards and cover a range of content as intended, it is important to develop a test blueprint to guide assessment construction. A good test blueprint, sometimes called a table of test specifications, describes the features desired in an assessment. It defines the purpose of testing, outlines the specific standards-based knowledge and skills (i.e., content) to be measured, and specifies the relative emphasis (e.g., number of questions or percent of total questions) to be devoted to each standard or aspect of content covered. In the case of complex assessments that involve multiple tasks (e.g., a project, portfolio), a test blueprint may also specify how each part of the assessment relates to targeted standards.

In addition, a thorough test blueprint will stipulate the types of questions to be included on a test (e.g., multiple-choice, written-response), and designate the different types of cognitive performances that will be required of test takers (e.g., basic recall of facts, interpretation/evaluation of concepts and principles, application of knowledge), again with an indication of the emphasis or weight to be given to each type of cognitive performance. It will also outline the methods that will be used to score student responses to the assessment.

In some cases, a test blueprint may provide some indication of how difficult an assessment should be. The level of difficulty desired will depend in large part on the intended purpose(s) of the test and the targeted student population(s). For example, if the ultimate goal of a test is to identify very high-achieving students (e.g., for advanced placement or formal recognition), the test should include a large proportion of difficult items (i.e., those that only high-achieving students are likely to answer correctly).

In summary then, a useful test blueprint outlines the content that an assessment should address, the types of cognitive performances it should elicit, the item/question format(s) that should be used, and in some cases the level of difficulty sought. Creating such a blueprint prior to assessment development 1) facilitates the development or selection of appropriate items for inclusion in an assessment, and 2) helps ensure that the standards that are considered most important are sampled more heavily than those deemed less important, and that different levels of student understanding, not just basic recall, are tapped.

Table 3.6 provides an example of a partial test blueprint for a mock written on-demand exam related to workplace readiness (career preparation) standards.

Table 3.6 Partial Test Blueprint

Types of questions to be included on the test:

- Multiple-choice questions: 75 (total)
- Written-response questions: 2 (total)

Standards (content) to be covered by the test, including the relative emphasis to be given to each standard (area of content):

STANDARDS to be covered	Percent of questions that should focus on: KNOWLEDGE	Percent of questions that should focus on: APPLICATION OF KNOWLEDGE
Standard 1: Communication	10%	10%
Standard 2: Thinking and Problem Solving	10%	10%
Standard 3: Team Building and Leadership	9%	9%
Standard 4: Decision Making	7%	7%
Standard 5: Employment Literacy	7%	7%
Standard 6: Technology Literacy	7%	7%
	TOTAL 50%	TOTAL 50%

Developing Multiple-Choice Items

Once a test blueprint is complete, item development can begin. Before developing multiple-choice items for inclusion in a written on-demand assessment, schools or districts should first see if there is an existing multiple-choice assessment or individual multiple-choice items that can be used or adapted for use. Most textbooks and some packaged programs come with tests, and many test publishers maintain banks of test items in various subject areas. These tests and individual items have been created by skilled test developers who have the expertise and resources necessary to both pilot

test items with large numbers of students and make revisions as necessary to ensure that all items effectively measure targeted knowledge and skills. In addition, increasing numbers of teachers are now sharing test items electronically. Using or adapting available resources can greatly expedite the test development process.

If teachers, schools, or districts choose to use or adapt existing multiple-choice items (or other types of items), they should be sure to evaluate both the quality of the items and how well the items meet their own purposes and goals. They need to judge whether the items reflect the standards they want assessed and, if they don't, they should adapt them accordingly. In addition, they should review the assessments to ensure that extraneous skills (e.g., reading ability in a test of mathematics skills) are not being overemphasized.

Whether reviewing existing multiple-choice items or writing new ones, several general guidelines should be considered. These general guidelines are presented in Table 3.7. They apply to the development and review of both multiple-choice items and written-response items. Discussing these guidelines before beginning the development or review process can be helpful.

Table 3.7 General Item-Writing and Review Guidelines

1. Use the standards and related documents (e.g., test blueprint) to guide item-writing and review efforts. As mentioned previously, this will help ensure that all the items and tasks developed are linked to targeted standards and measure the types of cognitive performances (e.g., recall vs. application) desired.
2. As items are developed or reviewed, ensure that they meet the following criteria:
   - focus on high-level thinking, reasoning, and problem-solving skills as much as possible;
   - use simple, concise language to clearly articulate the tasks to be completed;
   - include only information that is relevant and necessary for answering the items or completing the tasks;
   - are within the appropriate range of difficulty for the intended student population;
   - use the lowest readability level possible (e.g., grade-appropriate vocabulary; simple, concise sentences) since the purpose of each item is to measure students' standards-related knowledge and skills, not their ability to read and translate the item or task;
   - use graphics (when applicable) that are clear and easy to understand;
   - do not use language or content that could be offensive or inappropriate for a population or subgroup; and
   - do not include or implicitly support negative stereotypes.
3. Develop two or three times the number of items actually needed for the final assessment. This will make it possible to drop ineffective items following analysis of test results from classroom tryouts and field tests.
4. Allow ample time for editing and proofreading of items. Check for clarity, as well as for errors in spelling, grammar, and punctuation.

In addition to the general item writing and review guidelines outlined in Table 3.7, there are additional guidelines specific to the development or review of multiple-choice items. These guidelines, listed in Table 3.8, are intended to make the development or review of the multiple-choice format easier and more successful for schools and districts. Two of the recommended resources listed in Appendix D (i.e., Gronlund, Haladyna) provide sample multiple-choice items that help illustrate the recommendations made in these guidelines.

Table 3.8 Specific Guidelines for Developing or Reviewing Multiple-Choice Items

- Present a clearly formulated, concise problem in the item's stem. The best stems focus on a single aspect of content (e.g., a principle) and one type of cognitive performance (e.g., application of knowledge).
- State the item stem in positive terms whenever possible. Students, especially those with limited English proficiency, often have difficulty understanding questions that are phrased in negative terms (e.g., "Which is not an example of..."). They often overlook the word "not," and, therefore, misinterpret the question. If it is necessary to phrase a question using negative terms (e.g., not, except), make sure to capitalize or bold-face the negative terms so that they stand out to students.
- Avoid the use of unnecessary details in the item stem and answer choices.
- Use answer choices that are brief and parallel (e.g., if one answer choice begins with a verb, make sure all answer choices start with verbs).
- Use answer choices that are grammatically consistent with the stem of the item. Grammatical inconsistencies can provide clues that help uninformed students correctly guess the appropriate answer.
- Include distracters that are plausible and attractive to uninformed students. For example:
  - Use common misconceptions or errors of students as distracters.
  - Make distracters similar to the correct answer in both length and complexity of wording.
  - Use scientific- and technical-sounding words to help make distracters enticing.
- Do not give clues that might enable students to guess the correct answer or to easily eliminate incorrect alternatives. For example:
  - Avoid using similar wording in the item stem and correct answer.
  - Avoid writing the correct answer in a style that is distinctly different from the distracters.
  - Avoid stating the correct answer in greater detail or length than the distracters.
  - Avoid including absolute terms (e.g., always, never, all, none, only) in distracters.
- Make sure each item has a correct answer that is unquestionably correct or clearly best.

By following the guidelines in Table 3.7 and Table 3.8, schools or districts can help ensure that the multiple-choice items they develop and use are effective.

Developing Written-Response Items

As already noted, the same general guidelines that apply to developing and reviewing multiple-choice items also apply to written-response items (see Table 3.7). Listed in Table 3.9 are several additional guidelines specific to the development or review of written-response items.

Table 3.9 **Specific Guidelines for Developing or Reviewing Written-Response Items**

For all written-response items:

- Present a clearly formulated problem or situation (in paragraph form) in the item's prompt. Make sure that the described problem or situation is novel but not entirely unfamiliar to students. The context or details in the prompt should not be beyond the ability of students to imagine.
- Provide specific instructions that tell students everything they need to do when responding to the prompt. Be sure, however, not to provide excessive information, or you may remove the challenge for students.
- Present the instructions in the form of statements rather than questions whenever possible (e.g., Explain three reasons...." rather than "What are three reasons....").
- Avoid unnecessary detail in both the prompt and instructions. Ask yourself, "Is this essential information?" If the answer is "no," eliminate it.

For long written-response items:

- Clearly state the evaluation criteria (i.e., what students must demonstrate to receive a satisfactory rating). Providing this information helps students understand what is expected. (See Table 3.4 for an example.)
- Make sure that the information presented in the prompt, instructions, and evaluation criteria is consistent. For example, concepts included in the evaluation criteria should reiterate or support information given in the instructions and the prompt.

Assembling a "Draft" Written On-Demand Assessment

Once a variety of multiple-choice and written-response items have been developed, they can be assembled into a "draft" on-demand assessment that meets the overall requirements outlined in the test blueprint. This should be a relatively straight forward process as long as development of the individual

items was completed with the blueprint in mind. As a reminder, a good strategy for developing a written on-demand assessment that measures both breadth and depth of knowledge is to use a combination of multiple-choice and written-response items, employing each type of item for the purpose(s) to which it is best suited. In addition, when assembling the draft version of the assessment, it is best to include more items per targeted content area than the table of test specifications actually calls for because not all items are likely to be found effective following classroom tryouts and analysis of student response data.

Reviewing Written On-Demand Assessments Prior to Classroom Tryouts

After a draft assessment of individual items has been assembled, the process of determining the effectiveness of the assessment begins. This process should include several types of preliminary review before the items are then tried out in classrooms. The overall purpose of these reviews is to begin to identify ways in which the assessment can be improved to more effectively and fairly measure targeted knowledge and skills. Suggested types of preliminary reviews are as follows:

- **Editorial review:** Someone well-versed in English grammar and composition should ensure that all test items are stated as clearly and concisely as possible and are free of grammatical errors that could distract or provide clues for test takers.
- **Item-writing guideline review:** One or more assessment experts should ensure that items meet the criteria outlined in the general and specific item-writing guidelines presented earlier. At this point, reviewers should also check across the set of items to make sure there are no items that provide clues that might help students correctly guess the answers to other items.
- **Content review:** A respected group of content experts (e.g., teachers, industry representatives) should ensure that the assessment items are linked to targeted standards and that the test as a whole samples the range of standards-related content outlined in the test blueprint.
- **Bias review:** A committee sensitive to bias issues should ensure that the test's content and design are free of bias that might unfairly disadvantage one or more groups of test takers.

"Trying Out" and Evaluating Written On-Demand Assessments

Following the preliminary reviews described above, the assessment should be "tried out" by several samples of students. It is common to start by trying the assessment out with students in a small number of classrooms (i.e., classroom tryouts) to get a preliminary feel for how the items on the exam are performing (e.g., Do students seem to understand the test items? Are the test items appropriately difficult? Do the test items make it possible to effectively differentiate between high and low performing students?). Informal classroom tryouts should be followed by one or more formal field tests in which the assessment is administered to large groups of students representative of the population that will take the final version of the test. Both informal classroom tryouts and more formal, large-scale field tests should be followed by careful statistical analysis of test results to identify weak or potentially problematic test items that should be revised and/or removed from the assessment entirely. Statistical analysis of test results will be discussed in more detail in Chapter 5.

Refining Written On-Demand Assessments based on Item Analysis following Classroom Tryouts and Field Tests

The information collected during classroom tryouts and subsequent item analysis can be used to revise and improve weak or potentially problematic items, and to select the most effective items for inclusion in a refined version of the assessment. This is a very important step in development since the overall quality (i.e., validity and reliability) of a test is closely linked to the quality of the individual items that make up the test.

This process of trying items out in classrooms, analyzing the results of classroom tryouts and field tests, and refining test items accordingly should be repeated until test developers are confident that the assessment effectively measures student achievement in relation to targeted standards.

Helping Students Succeed on Written On-Demand Assessments

While, in most cases, teachers cannot provide direct assistance to students as they take a written on-demand assessment, they can take steps to help prepare their students for such tests. The most important step that teachers can take is to provide instruction related to all the standards covered by an exam. Students are more likely to do well on a test if they have had

should teach directly to the test, but instead, that they should teach to the standards that underlie the exam. Designing curriculum that covers targeted standards improves students' chances of doing well on assessments that are linked to those standards.

In addition to providing instruction that is linked to standards, there are many activities that teachers can use to build students' capacity to successfully answer the types of questions they are likely to encounter on written on-demand assessments (e.g., multiple-choice and written-response questions). Examples of several activities are described below:

- Familiarize students with the basic structure of a multiple-choice question (see Table 3.2) and give students multiple opportunities to practice answering such questions.
- Familiarize students with the basic structure of short and long written-response questions (see Tables 3.3 and 3.4) and plan a variety of classroom activities that will help students learn to interpret, think through, and answer such questions successfully. For example:
  - Define and explain terms that are commonly used in written-response questions (e.g., list vs. describe vs. explain in detail). Knowing the meaning of terms like these will help students better understand the depth of response expected for a question.
  - Walk students through the wording of several practice written-response questions, helping them to identify and understand the key requirements of each question (i.e., what is being asked and what a student must do to answer the question completely).
  - Model for students different strategies for "thinking through" and outlining answers to written-response questions.
  - Explain to students the importance of including details in their answers to written-response questions. Then, model different strategies for identifying details to include in their answers. This is a very important step in helping students do well on written-response questions. It is quite common for students to provide very general answers to these questions. When they do this, they leave their readers wondering how well they really know the content covered by the question. Helping students learn to articulate what they know in writing and to provide the level of detail and specificity necessary to convince their readers that they have mastered targeted content is often critical to students' success on written-response questions. Table 3.10 provides one specific example of an exercise that teachers can use to help students identify important details to include in their answers to written-response questions.

Table 3.10 The "Keep Asking HOW/WHY" Technique: An exercise for helping students provide detailed answers to written-response questions

By continuing to ask themselves "how" or "why" as they attempt to answer a written-response question, students can identify important details to include in their answers, details that demonstrate their understanding of a topic. This technique is best demonstrated through an example. The example provided is based on a written-response question in the area of food service and hospitality that asks students to discuss how four different factors can affect the quality of deep-fried foods in a restaurant. One factor that students must address is the temperature of the oil in the fryer. The example focuses on how a student might respond when discussing oil that is too hot.

Question:

Several customers have complained about the quality of the deep-fried foods at your restaurant.

Explain in detail how each of the following factors could be causing the poor quality of the fried foods:

- the temperature of the oil in the fryer;
- the amount of food fried in the fryer at one time;
- the types of foods fried in the fryer; and
- the procedures used to clean the fryer.

Exercise for Developing Student Response:

If the oil is too hot, the outside of the food may cook faster than the inside....

Ask: HOW/WHY does this affect the quality of deep-fried foods?

The outside of the food will be done, but the inside of the food may still be raw....

Ask: HOW/WHY does this affect the quality of deep-fried foods?

The raw food might not taste good to customers.... **Ask: WHY**.... Its flavor or texture may be unappealing if it is a food that is not typically served raw.

The raw food could also pose a health risk to customers... **Ask: WHY**.... Because it may contain bacteria that was not killed during the cooking process.... **Ask: WHY**.... Bacteria can cause food-borne illnesses if consumed.

- Have students, as a class, brainstorm the answer to a written-response question. Then have each student write his or own response to the question, using ideas generated during the group discussion.
- Provide students with multiple opportunities to practice writing about what they know (e.g., through homework assignments, in-class projects, written-response assessments given in the classroom). The more opportunities that students have to answer practice written-response questions, the more likely they are to do well on such questions when they appear on formal written on-demand assessments.
- Before administering written-response questions, show and explain to students the general criteria that will be used to evaluate their answers. This can help students better understand what is required to achieve a top score.
- Have students evaluate their own answers to practice written-response questions, as well as the answers of their peers, using a scoring guide (rubric). Then, encourage students to discuss strategies for improving their own and others' work.
- Allow students to revise and improve their answers to practice written-response questions, using your feedback and/or feedback from their peers.
- After students have answered a practice written-response question, provide them with examples of student work that illustrate the different levels of performance outlined in the scoring guide (rubric) for the question. Explain why each piece of student work received the score it did. Again, this will help students better understand what is required to achieve a top score. This understanding may help students improve their performance on the next written-response question they answer.

■ Prior to administering written on-demand assessments, review effective test-taking strategies with students. Examples of some specific test-taking strategies are provided below:
- Remind students to read test directions carefully.
- Remind students to pace themselves by considering the number of questions on the test and the amount of time given to complete the exam.
- Encourage students to read each test question very carefully. For written-response questions, remind students to read each part of the question before responding. Suggest that they underline the key requirements of these questions to make sure they clearly understand all that they must do to provide complete responses.

- For multiple-choice questions, encourage students to generate their own idea of the most accurate answer to a question before reviewing and selecting from the answer choices provided.
- For written-response questions, encourage students to briefly outline their answers before actually writing their responses. This might involve thinking quickly of the main ideas that will serve as a framework for an answer and then organizing these ideas into a logical sequence. Also, remind students to include in their answers details that will demonstrate their knowledge of the topic covered.
- Remind students to check their work carefully when they have finished to make sure that they have answered all questions as completely and accurately as possible.

Summary

Written on-demand assessments such as multiple-choice tests and written-response tests can be an important part of a standards-based assessment system. All written on-demand assessments share some general features. For example, they measure students' knowledge and skills at a particular point in time (e.g., at the end of a unit or course of study) and are typically administered under uniform conditions and within a specific time period. In addition, written on-demand assessments can be used to assess the breadth of students' standards-based knowledge (e.g., multiple-choice test) and to probe the depth of students' knowledge in relation to a select number of standards (e.g., written-response test).

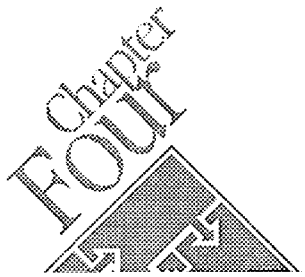
Features specific to multiple-choice and written-response items respectively are their formats and some of the advantages and disadvantages associated with their use. The format of a multiple-choice item includes *a stem* (the question or problem that is to be solved) and four to five *answer choices*. One of the choices is the *correct answer*; the other choices are called *distracters* (or incorrect answer choices). Students must select the correct answer from among the answer choices. The format of a written-response item varies, often depending on whether the item requires a short or long response from students. Short written-response items, such as those used by ACE, and long written-response items, such as C-TAP written scenarios, include an *item name*; a *prompt,* which usually describes a problem or situation to be considered by the student; and *instructions,* which tell students what to do. In addition, long written-response items usually include *evaluation criteria* that clearly articulate what students must demonstrate to receive a

satisfactory rating. For both short and long written-response items, students must respond in writing to a question or prompt, providing an answer that ranges in length from several paragraphs (short written response) to several pages (long written response).

Multiple-choice items and written-response items also have distinct advantages and disadvantages. Some advantages of multiple-choice items are that they can be used to cover a broad range and number of standards efficiently and they can be easily scored by people or machines. They cannot, however, be used to measure certain types of complex achievement (e.g., students' ability to develop or express ideas) and they are difficult for English learners because of their heavy reliance on reading skills. Advantages of written-response items are that they can be used to probe students' depth of knowledge and understanding in relation to a select number of standards, and they are particularly useful for measuring students' ability to pose appropriate solutions to realistic problems or to develop or express ideas. Some disadvantages of written-response items are that they cannot be quickly judged as either right or wrong. Instead, they must be evaluated by people, not machines, and their results are apt to be unreliable if scoring criteria are not clearly articulated, or are applied improperly or inconsistently.

When multiple-choice and written-response items are used together in an assessment, the advantages of one help to compensate for the disadvantages of the other and vice versa. For this reason, schools and districts should consider using one or more written on-demand assessments that include both types of items. The process for developing such an assessment involves a number of key steps including 1) creating a test blueprint, 2) writing individual multiple-choice and written-response items, 3) assembling a draft assessment, 4) reviewing the assessment to ensure links to standards, clarity, accuracy, and freedom from bias, 5) "trying out" the assessment and analyzing the test results to determine if the items work as intended, and 6) refining the assessment based on insights gleaned from preliminary reviews, classroom tryouts, and formal field tests.

Teachers can take a variety of steps to help prepare their students for written on-demand assessments. Among the steps that teachers can take are providing instruction that is directly linked to the standards covered by a test, designing classroom activities that will build students' capacity to answer multiple-choice and written-response questions successfully, and reviewing specific test-taking strategies prior to administering written on-demand assessments to students.



Developing Cumulative Assessments

Chapter 3 introduced written on-demand assessments such as multiple-choice and written-response tests. As indicated, these assessments are effective for measuring the breadth and some depth of student knowledge at particular points in time, but are somewhat limited in their capacity to assess hands-on application of knowledge and skills and students' ability to revise and improve their work over time. To achieve these additional purposes, schools and districts can consider implementing cumulative assessments as part of their assessment systems.

This chapter focuses on cumulative assessments, beginning with a description of the general features of cumulative assessments, including several advantages and disadvantages associated with their use. Next, the chapter addresses two specific types of cumulative assessments: projects and portfolios. Key features of these two assessments are presented along with a description of each assessment's structure. The chapter ends with a discussion of some challenges associated with implementing cumulative assessments, challenges that must be addressed in order to realize the advantages of cumulative assessments. The information shared throughout this chapter is in no way exhaustive, but should help schools and districts begin to identify key features and structural elements to incorporate into their own cumulative assessments.

General Features of Cumulative Assessments

Cumulative assessments are completed over time (e.g., usually weeks or months) and demonstrate the best students can do when given opportunities to practice and revise their work based on self-evaluation and constructive

feedback from others (e.g., teachers, peers). Portfolios may be the best known form of cumulative assessment, but "hands-on" student projects and written compositions (e.g., research papers) are also commonly used to assess student learning and achievement over time.

Cumulative assessments are designed to measure students' depth of standards-based knowledge and their ability to apply that knowledge and related skills in meaningful, often hands-on, ways. They tend to be more complex than written on-demand assessments, often requiring the integration of knowledge and skills across disciplines.

Cumulative assessments usually result in substantial work products. A completed portfolio, for example, is likely to include a piece of writing that required extended research and thought. It may also include the final product(s) of one or more hands-on projects, such as a set of historical maps and charts or a cabinet built by hand. Products like these can provide teachers and others (e.g., potential employers, college admissions officers) with concrete evidence of what students know and can do.

Cumulative assessments are, however, more than the final products of students' efforts. They are also multi-step processes (e.g., a series of thoughts, actions, judgments, and decisions), during which students purposely and systematically demonstrate and refine their knowledge and skills. During cumulative assessments, evidence of learning can be gleaned both from students' final products and from the various efforts leading up to and following them. For example, when completing a project assessment, students may develop project plans and document evidence of progress before actually creating their final products. After producing their final products, they may also write summaries of project results. Indices of achievement, or pieces of evidence, can be collected during all phases of this process. Together these pieces of evidence can present an integrated view of what students have accomplished.

Finally, cumulative assessments encourage high levels of student involvement in and responsibility for learning throughout the assessment process. For example, as students compile portfolios of their work and manage long-term projects, they help establish learning goals (e.g., by selecting topics and themes for various writing and work samples), reflect on their experiences, evaluate their efforts, and revise their work based on their own and others' feedback. More specific examples of student involvement and responsibility in cumulative assessments will be provided later in this chapter when projects and portfolios are discussed.

Table 4.1 summarizes the general features of cumulative assessments.

Table 4.1 General Features of Cumulative Assessments

- Take place over substantial periods of time (e.g., weeks, months)
- Represent the best students can do given constructive feedback and opportunities to revise their work
- Focus primarily on depth of knowledge and students' ability to apply knowledge and skills
- Usually result in substantial work products
- Value the process, as well as the products, of student learning
- Require students to actively participate in and take responsibility for their learning

Some Advantages of Cumulative Assessments

Because cumulative assessments emphasize the process(es), as well as the final products, of learning, they can be very informative for instructional planning. By reviewing work in progress and conferencing with students during the assessment process, teachers can identify students' instructional needs and plan accordingly.

Another advantage of cumulative assessments is that they tend to blur the line between instruction and assessment, engendering as well as revealing learning. Throughout the cumulative assessment process, students are provided with constructive feedback and encouraged to reflect on and evaluate their own work. They are expected to revise and improve their work based on their own insights and recommendations made by others. Through these activities, together with coaching from their teachers, students can actually deepen their understanding of standards-based content and refine a variety of skills during the assessment experience. As students become more aware of how they learn, what help they need, and what it takes to manage complex tasks effectively, they also strengthen their capacity to be independent, self-motivated learners (Rogers, 1996).

Finally, cumulative assessments are generally less dependent on instantaneous production of language and other traditional modes of representation than are written on-demand assessments. Cumulative assessments allow students to express what they know in a variety of ways (e.g., hands-on demonstration, visual representations) and provide more time for completion than do most written on-demand assessments (i.e., usually

weeks or months vs. minutes or hours). In addition, teacher-student conferences, a common component of cumulative assessments, can provide students with multiple opportunities to clarify task requirements and performance expectations. As a result, students may participate more readily and perform better. This is particularly true for English language learners and other students who tend to perform poorly on traditional written on-demand assessments. Many C-TAP teachers have reported their surprise at seeing their "average" students perform above average on cumulative assessments. They suggest that their students are more motivated when they have choices and can move beyond the limited modes of representation required by many written on-demand assessments.

Some Disadvantages of Cumulative Assessments

The substantial benefits afforded by cumulative assessments must be weighed against the disadvantages of these assessments, particularly for large-scale or high-stakes assessment purposes. These disadvantages are related primarily to technical adequacy, standards coverage, and cost.

Cumulative assessments face a number of challenges related to their technical adequacy. Of particular concern is reliability. Studies on cumulative assessments indicate that the levels of reliability for tasks such as portfolios are lower than those for traditional multiple-choice tests (Koretz et al., 1994). There is recent evidence, however, that reliability for cumulative assessments is increasing, particularly interrater reliability (i.e., the agreement on ratings by two or more raters on the same assessment task). Specifically, interrater reliability increases with increased prescriptiveness of task requirements, and when raters are thoroughly trained on well-defined scoring rubrics (see, for example, Resnick, 1996). The concept of reliability is discussed more fully in Chapter 5.

The usefulness of cumulative assessments is also limited by their inability to cover a large number of standards, given practical time constraints. For example, while a multiple-choice assessment given in a standard 45-minute period can cover a wide range and number of standards, a cumulative assessment that takes several days or months to complete may cover only a handful of standards at best.

Finally, the cost associated with developing and implementing cumulative assessments for large-scale or high-stakes assessment purposes is often prohibitive. For example, scoring procedures used to ensure high levels of reliability (i.e., the development of scoring rubrics and intensive training

for scorers) cost much more than scoring procedures used for multiple-choice assessments (i.e., electronic scanning).

Although the disadvantages of cumulative assessments must be seriously considered when developing any assessment system, so must their substantial benefits. When combined with written on-demand assessments in a standards-based assessment system, the benefits of cumulative assessments help compensate for the disadvantages of written on-demand assessments and vice versa.

Project Assessments

Two types of cumulative assessments that have become increasingly popular over recent years are project assessments and portfolio assessments. This section introduces key features and structural elements of effective project assessments.

In the most general sense, a project is an in-depth, hands-on exploration of a topic, theme, idea, or activity, resulting in a product, performance, or event for assessment (Katz & Chard, 1989). Project assessments can measure students' standards-based knowledge and skills and their ability to apply that knowledge and those skills in authentic situations. They can also assess how well students are able to evaluate their own work, solve problems, plan and carry out complex activities, and communicate findings to an audience.

Key Features of Project Assessments

While projects come in many shapes and sizes, most effective project assessments share several key features in addition to the general features outlined for cumulative assessments in Table 4.1. These key features are summarized in Table 4.2 and discussed in more detail below.

Table 4.2 **Key Features of Effective Projects**

- Involve hands-on application of knowledge and skills in a purposeful, authentic activity
- Encourage students to integrate knowledge and skills, often across several subject areas
- Focus on one or two content standards at most

Perhaps the most important feature of project assessments is that they involve hands-on application of knowledge and skills in purposeful, authentic activity. Consider the examples of C-TAP project ideas presented in Table 4.3. In each case, students must explore a complex and realistic question, problem, or activity over time. During the process, they must do more than "learn about" a topic. In fact, they must actually use their acquired knowledge and skills to create products, performances, or events that are related to the topic. Why is this important? Many educators suggest that students learn best when placed in situations requiring the actual use of the knowledge and skills to be learned.

Another key feature of project assessments is that they encourage students to integrate knowledge and skills, often across several subject areas. Rarely in life do individuals engage in activities that call for only one type of skill or for skills relating to only one discipline. Usually endeavors are more complex, requiring the integration of a variety of information and skills. Project assessments mirror this reality. When working on challenging projects, students will invariably need to use content knowledge from a variety of subject areas (e.g., career-technical, mathematics, English-language arts, science, social studies), as well as a range of high-level thinking and management skills. For example, analysis of any of the sample C-TAP project ideas described in Table 4.3 suggests that as students work to create targeted products, performances, or events, they are likely to use multiple kinds of knowledge and skills, such as planning, organizing, researching, experimenting, writing, computing, calculating, creating, making, collaborating, evaluating, and presenting.

It is also evident from Table 4.3 that the sample project ideas are relatively focused and complex in nature. This, too, is a key feature of effective project assessments. Project assessments usually address one or two standards at most, and take sustained effort, over time, to complete. As students explore and work with project topics, they are given ample time to develop and demonstrate in-depth understanding and mastery of relevant standards and to practice and sharpen other important skills. As with all cumulative assessments, evidence of student achievement can be gleaned from the final fruits of project work (i.e., a product, performance, or event) and from the various actions students take to complete their projects.

Table 4.3 Examples of C-TAP Project Ideas

- **Livestock Facility**: Develop and construct a livestock facility (or scale model) that provides for the safe movement and handling of livestock. *(Agricultural Core. Standard: Animal Science — Animal Facilities, Equipment and Handling)*
- **Computer Software Manual:** Investigate the typical computer software questions asked by beginning users and develop a software reference manual to help users find answers to their questions. *(Business Education. Computer Science and Information Systems Cluster. Standard: Information Processing)*
- **The Infomercial:** Investigate the current infomercial trend and create an infomercial using a videotape and product of choice. *(Business Education. Marketing Cluster. Standard: Promotion)*
- **Seniors' Home Health Care or Nursing Home Care:** Provide current information to senior citizens about contrasts in health care options, costs, insurance coverage, and services for home health care and residential nursing home care. *(Health Care Core. Standard: Socioeconomics.)*
- **Preparation of Teaching Materials:** Prepare and facilitate three age-appropriate activities for children ages 4 through 6 that promote the development of physical, intellectual, emotional, and social skills. *(Home Economics. Child Development and Education Cluster. Standards: Child Growth and Development; Developmentally Appropriate Activities)*
- **Child Nutrition:** Develop a one-month snack schedule that meets the state's standards and licensing regulations, and implement one week of the schedule. *(Home Economics. Child Development and Education Cluster. Standard: Nutrition, Health, and Safety Practices)*
- **Video Broadcast:** Produce a 15-minute video broadcast, including two commercials, two news stories, one weather report, three songs, and one station identification. *(Industrial Technology Core. Standard: Communications — Photography and Motion Pictures)*
- **Frameless Cabinet:** Develop a frameless style cabinet that includes a door, drawer, and solid surface top. *(Industrial Technology. Construction Technology Cluster. Standard: Wood Product Manufacturing — Cabinetmaking)*

Structure of Project Assessments

Most project assessments consist of four basic steps, or parts, each of which can result in student work for assessment. These four basic steps are as follows: 1) planning and organizing the project; 2) researching and developing the project; 3) producing a final product, performance, or event;

and 4) presenting the final project. These four steps are discussed in more detail throughout the remainder of this section.

Because the C-TAP project provides a good example of the basic project assessment structure, it will be referenced in this section to help explain typical project requirements. The C-TAP project, which includes the four basic steps described above, is summarized in Table 4.4.

Table 4.4 C-TAP Project Structure

1) **Planning and organizing the project**
   Project Plan: a document describing the goals of a project and how the project will be completed

2) **Researching and developing the project**
   Evidence of Progress: three pieces of evidence showing how a project was developed

3) **Producing a final product**
   Final Product: a physical product or documentation of a performance or event that is the result of project work

4) **Presenting the final product**
   Oral Presentation: a presentation describing the project, the knowledge and skills used to complete it, and what was learned during the process

Planning and Organizing the Project

A common first step in completing a project assessment is planning and organizing the project itself. During this step, students determine project goals and identify the activities and resources needed to complete their projects. When completing project plans, for example, C-TAP students are required to do all of the following: specify what they intend to accomplish and the standards-based knowledge and skills they will demonstrate; fully describe the final product, performance, or event they will create; outline the major steps necessary to complete their work; determine the materials and resources they will need to be successful; and consider how they can document evidence of progress during the project process.

Requiring project planning as part of the assessment process communicates the importance of thinking through a project in its entirety before beginning work. It also helps students develop and refine their project planning skills. As they brainstorm and finalize project goals, determine the

actions needed to achieve these goals, and consider the resources needed to support their efforts, students get first-hand experience in the type of planning and organizing that is typical of project work in the workplace and life in general.

Some project assessments require that students submit a formal written plan as part of the assessment. This can be useful to teachers and students in several ways. By reviewing written project plans, teachers can ensure that students' project topics are linked to standards, are appropriately difficult, and that the time and resources necessary to complete the project are available. They can also get a sense of students' ability to think through and plan complex, long-term activities and how well students use high-level thinking skills (e.g., brainstorming, analyzing, evaluating, strategizing, synthesizing, estimating).

For students, written project plans can serve as road maps, helping guide their work throughout the assessment process. Students can and should be encouraged to use their plans to stay organized and focused on their goals. By reviewing their project plans regularly, both alone and with their teacher(s), students can determine how well they are meeting their intended goals.

Table 4.5 presents suggested guidelines for what students should include in a project plan. These guidelines were adapted from the C-TAP project assessment.

Because project planning may be a new skill for students, teachers may need to provide support both by helping students brainstorm project topics and by giving students feedback on their written project plans. Ideally, teachers should meet with students to review and discuss their plans, keeping the following kinds of questions in mind: Is the project topic related to the program of study and corresponding standards? Is the topic related to the student's interests? Is the project challenging? What knowledge and skills does the student need to learn? Given the time and resources available, does the plan for completing the project seem feasible? How can progress be documented?

As students begin project work and monitor their progress, unanticipated problems may arise, requiring them to rethink their strategies for reaching their goals. Or, once into the project process, students may think of ways to expand on or improve their goals and strategies. In either case, students should be encouraged to adjust their plans and pursue improvements rather than be required to adhere strictly to their original plan.

Table 4.5 **Suggested Guidelines for Project Plans**
(based on the C-TAP project assessment)

When preparing project plans, students should include most, if not all, of the following:

- **Project idea/topic**, including a brief overview of the focus of the project
- **Project purpose(s) and goal(s)**, including the final product or event the student will create, and the standards-related knowledge and skills that will be learned and demonstrated
  (Note: Students should be encouraged to state purpose(s) and goal(s) in very specific and measurable terms so both they and their teachers are clear about what will be produced.)
- **General process for completing the project**, including an outline or description of the major steps that will be taken
  (Note: Students should be encouraged to consider how the steps are connected and in what order they must occur to successfully achieve project goals.)
- **Resources needed to complete the project**, including strategies for obtaining the resources
- **Evidence of progress that the student will collect** to help monitor the project's development and demonstrate standards-based achievement
- **Timeline for completion**, including the due date for the final product, and the target dates for the major steps toward completion

While students should be encouraged to revise their project plans during the project as needed, they should be discouraged from writing their plans "after the fact." Experience with C-TAP has shown that if students believe that their final product and the process leading to it must exactly match that outlined in their original project plan, they are tempted to create their plans after their projects are completed to guarantee a "perfect fit." When they do this, students miss out on the learning involved in project planning. Both teachers and students should keep in mind that the project plan serves several important purposes, such as giving students a chance to practice and refine their capacity to plan, to organize complex tasks, and to facilitate the flow of work during the project process.

Researching and Developing the Project

The second step in most project assessments is researching and developing the project. This process takes place over a period of time, ranging from several days to most of a school year. During this time, students are usually

given class time to do project work (e.g., reading relevant literature, sharing information with classmates, developing a physical product). Students may also conduct research and development activities outside the classroom (e.g., interviewing or observing professionals at their workplaces).

As students research and develop their projects, they are often required to collect "evidence" of their progress. Evidence of progress can take many forms (e.g., journal entries, research notes, interview questions, sketches, photographs, rough drafts). Such evidence is meant to document the major steps taken and the milestones reached during the process of completing a project.

Evidence of progress can be used in a variety of ways during this step of a project assessment. For example, teachers can informally check the evidence at various points in time both to ensure that students are progressing at a satisfactory rate and to provide feedback that will help students improve their work. In addition, students can periodically review their own evidence as one way to monitor and regulate their progress toward completing their projects. By doing so, they may also identify ways to improve their work.

Teachers can also require students to collect evidence of progress for formal evaluation, as is the case with the C-TAP project. C-TAP teachers use this evidence to assess students' standards-based knowledge and skills and their ability to apply that knowledge and those skills as they develop a specific product, performance, or event. In cases where evidence of progress is formally evaluated, students must concentrate on collecting and submitting evidence that effectively demonstrates their mastery of the standard(s) targeted by their projects.

Producing a Final Product

The third step of most project assessments is the production and submission of a final product for assessment. The final product is the culmination of students' research and development efforts. It typically demonstrates students' ability to apply knowledge and skills related to the standard(s) targeted by their project topics.

If the final product is a physical product (e.g., a brochure, cabinet, sculpture, written document), then students may submit the product itself for assessment. If the final product is a performance or event (e.g., a poetry reading at a local nursing home, implementation of an age-appropriate learning activity in a preschool), then students can submit documentation of the performance or event for assessment (e.g., a series of photographs or a

videotape showing the event and the reaction of some of the participants). Often, as is the case with the C-TAP project, the final product of a project can be presented, either alone or as part of a larger collection of work, to potential employers and others as evidence of mastery of valued knowledge and skills.

Presenting the Final Product

The final step in most project assessments requires students to present their projects to the teacher and sometimes others (e.g., classmates, parents, community members, a panel of professionals). The presentation usually includes a showing of the final product itself (i.e., physical product or documentation for a performance or event) and an explanatory or reflective piece that may be oral or written in form, or both.

Presenting the final product can serve several important purposes. Because of the in-depth, long-term nature of project work, students are required to invest substantial time and energy to reach their goals. Both written and oral presentations provide an opportunity for students to show off the fruits of their hard work. During presentations, students can describe the process of completing their projects and explain how their efforts reflect what they know and can do. They can also share knowledge and insights gained during the project process. Table 4.6 presents suggested guidelines for what students should include in their project presentations. These guidelines have been adapted from the C-TAP project assessment.

As students prepare project presentations, they have opportunities to practice and demonstrate important knowledge and skills. For example, as students prepare to describe and explain their work and accomplishments to others, they engage in reflection and self-evaluation. By practicing these skills, students can become more aware of their strengths and weaknesses, as well as gain insight into how to work more effectively.

In writing or delivering project presentations (including responding to questions from the audience), students continue to demonstrate important knowledge and skills, particularly written or verbal communication skills, depth of understanding of the project topic, and an awareness of how their work relates to important standards.

Project presentations, whether written or oral, are not without their challenges. Because such presentations normally occur at the end of the project process, some students believe that they have completed all the "really tough work," so they do not take adequate time to think through or

prepare their presentations. When this happens, students rarely demonstrate their actual ability to communicate or show mastery of important knowledge and skills. They may also miss out on the benefits gained from reflecting on and evaluating their finished work and the project process. In addition, those on the receiving end of presentations (i.e., readers, a live audience) miss out on the informative and educational value that project presentations can have.

Table 4.6 **Suggested Guidelines for Project Presentations**
(based on the C-TAP project assessment)

When preparing project presentations, students should describe and explain most, if not all, of the following:

- **Project topic/idea**, including why the topic/idea was selected and how it relates to relevant content standards
- **Final product, performance, or event** created by the student, including what it is and its key features
- **Major steps** taken to complete the project
- **Knowledge and skills applied** during the project process
- **Challenges encountered as the project was completed** and how these challenges were resolved
- **What was learned and accomplished** as a result of project work
- **What the student would do differently** if given a chance to do the project again
- **How the student can use what was learned** to inform future endeavors

To help ensure that both students and teachers get the most from project presentations, teachers can keep several guidelines in mind. First, they should schedule oral presentations well in advance and encourage students to practice their presentations ahead of time, helping them when possible. In addition, they should invite "outsiders" (e.g., parents, community members, professionals from a field related to the class) to read or view the presentations. Students tend to take the presentation task more seriously when they know their readers or audience will include professionals from the real world and not just their classmates and teacher. Teachers should also encourage the readers or audience to ask questions. Informing students that they are responsible for answering questions about their projects encourages them to be prepared.

Portfolio Assessments

A portfolio assessment involves the structured collection of student work that documents students' knowledge and their ability to apply both knowledge and skills in a variety of authentic contexts. Whereas projects typically require students to produce one product or event that is related to one topic/theme and one or two standards, portfolios generally require a variety of student work related to multiple standards and sometimes to multiple topics/themes. As a result, portfolio assessments can provide a more comprehensive view of students' standards-based knowledge and skills than projects alone can.

Key Features of Portfolio Assessments

While portfolio assessments vary, most effective portfolios share several key features in addition to the general features outlined for cumulative assessments in Table 4.1. These key features are summarized in Table 4.7 and discussed in more detail below.

Table 4.7 Key Features of Effective Portfolios

- Include a variety of student work reflecting multiple standards
- Grow out of the regular classroom curriculum and students' work-based experiences
- Can be used to document performance improvement over time and/or to showcase overall achievement

Portfolio assessments, which are completed over a period of weeks or months, include a variety of student work samples, each sample reflecting knowledge and skills related to at least one standard. As a collection, the portfolio work demonstrates student progress and/or achievement vis-à-vis multiple standards. Some types of student work that can be included in portfolios are the following:

- **work products/samples** (e.g., physical products resulting from hands-on project work; photographs or videotapes of student performances)
- **writing samples** (e.g., research papers, fictional narratives)
- **career-related materials** (e.g., a resume, letters of recommendation, performance evaluations by supervisors)

Effective portfolio assessments identify a range of standards that students must demonstrate (i.e., required standards) and outline the different types of work that students must submit to show their mastery of these standards. When using portfolio assessments, teachers can examine their existing curriculum to identify work assignments that could naturally generate portfolio entries. Similarly, teachers can ask students with jobs or internships to review their work responsibilities and identify work products that may be appropriate for their portfolios. In these ways, the student work that goes into a portfolio can grow out of the regular classroom curriculum and students' work-based experiences, rather than be created exclusively for the purpose of assessment.

Another key feature of most effective portfolio assessments is that they can be used to document student learning and improvement over time (i.e., a formative "working portfolio"), to showcase overall achievement following a program or course of study (i.e., a summative presentation portfolio), or both. For a formative working portfolio, students keep drafts (or documentation) of work at various stages of completion so that progress can be reviewed periodically by both the teacher and student. When examining working portfolios, teachers can determine whether a student's work has improved from one draft to the next, and identify areas needing further improvement. Based on their periodic reviews of students' work, teachers can provide the students with constructive feedback about their portfolios and revise or expand their own plans for instruction. Students, too, can evaluate the work in their portfolios, making decisions about how to improve their work and proceed. As students revise their work, based on their own reflections and teacher feedback, they can deepen and refine the knowledge and skills they are attempting to demonstrate.

Many formative working portfolios can be transformed into summative presentation portfolios (i.e., final portfolios) that showcase evidence of overall standards-based achievement. This transformation takes place as a program or course of study draws to a close. At that time, students can finalize each piece of work in their working portfolios and decide which pieces reflect their best work and meet the various requirements of the portfolio assessment (i.e., the specific types and quantity of work required, the particular standards that must be demonstrated). The students include these pieces of "best" work in a final presentation portfolio that showcases their overall standards-based achievement. These summative presentation portfolios can be given to teachers and others (e.g., potential employers, parents, college admission officers) as evidence of what students know and can do.

Structure of Portfolio Assessments

Most portfolio assessments provide a basic structure, or framework, within which student work can be completed and submitted. This structure usually specifies the number and range of standards to be demonstrated and the specific quantity (e.g., number of entries) and types of work (e.g., hands-on work samples, writing samples) required. It may also outline how work will be evaluated and whether it will be judged as a whole or in pieces. The structure of portfolio assessments can vary greatly from one site (e.g., classroom, school, district) to another depending on the specific purposes of assessment and the nature of the standards to be demonstrated.

Perhaps the best way to illustrate how portfolio assessments can be structured and the types of work they can include is to provide a concrete example: the C-TAP portfolio. The C-TAP portfolio is designed to help students in career-technical programs prepare for postsecondary training and work by: 1) requiring demonstration of knowledge and skills needed in the workplace, 2) showcasing students' best work to potential employers, colleges, and training programs, and 3) improving students' ability to plan work, document progress, identify strengths and weaknesses in their work, and refine and improve their work over time.

The structure of the C-TAP portfolio requires students to submit entries (i.e., student work) within four or five major sections: *Portfolio Presentation (Introduction), Career Development Package, Work Samples, Writing Sample,* and *Supervised Practical Experience Evaluation* (optional). Through this work, students must demonstrate achievement related both to career preparation standards (i.e., workplace readiness knowledge and skills) and to a range of required industry-specific standards (usually five or six standards). Table 4.8 summarizes the structure of the C-TAP portfolio, showing the entries required within each section.

Each section of the C-TAP portfolio is described more fully below. Before continuing, it is important to note that the C-TAP portfolio presents just one of the many ways to structure portfolio assessments. Even the C-TAP portfolio itself can vary from site to site depending on the career-technical program, the specific career paths of students, and the relative mix of career cluster, career preparation, and academic standards that the teacher, school, or district decides to emphasize.

Table 4.8 C-TAP Portfolio Structure

Portfolio Presentation (Introduction)

1. Table of Contents
2. Letter of Introduction

Career Development Package

1. Resume
2. Employment or College Application
3. Letter of Recommendation

Work Samples
Four examples and descriptions of work, demonstrating mastery of important career-technical standards

Writing Sample
A sample of writing, demonstrating investigative, analytical, and writing abilities

Supervised Practical Experience Evaluation (optional)
Documentation of a student's practical or work experience, demonstrating workplace readiness

Portfolio Presentation (Introduction)

This section of the C-TAP portfolio includes a table of contents and a letter of introduction. The table of contents outlines the different sections of the portfolio and the materials included in each section. It helps reviewers locate each section of the portfolio and also helps students keep track of their different portfolio entries. The letter of introduction introduces the portfolio to an outside reviewer. In the letter, students must describe their goals (both educational and career) and personal strengths, as well as the best work sample in their portfolio. They must also discuss how their work has improved over time and how it reflects their career goals and the knowledge, skills, and qualities important to employers.

A completed Portfolio Presentation demonstrates career preparation skills, an awareness of career goals, the ability to reflect on and evaluate work, and the ability to compose a well-written original letter. It also illustrates students' organizational and presentation skills (e.g., neatness, accuracy).

Although the Portfolio Presentation is the first section of the portfolio, it is usually the last step in the portfolio process since it requires a great deal of

reflection and self-evaluation by students. They must review their portfolio in relation to the required standards and what is valued by employers, and then make decisions about which aspects of themselves and their work to highlight in their letter of introduction. They must also reflect on how their work has improved over time and identify, from their work, their personal strengths. This process can be educational, as well as motivating, as students see tangible evidence of their learning or achievement beyond a set of numbers or grades on assignments and tests.

Career Development Package

This component of the C-TAP portfolio consists of a resume, an employment or college application, and a letter of recommendation. Its primary purpose is to prepare students to search for a job, seek advanced training, or apply to college. It should demonstrate evidence of career planning and preparation and an awareness of the knowledge, skills, and qualities important to potential employers.

Industry representatives and teachers working with C-TAP agree that a career development package is an essential part of a career-technical portfolio. Because U.S. workers typically change jobs several times during their careers, the skills needed to complete this portfolio entry are likely to be used throughout students' future work lives. These skills include formulating and articulating career goals; identifying and describing personal strengths; documenting and explaining work experiences; organizing information; writing effectively; and completing work that is clear, neat, and accurate.

Work Samples

Work samples are concrete, hands-on examples of what students have learned in the classroom, on the job, or in a volunteer placement. They may be actual products (e.g., a spreadsheet, a plan for a construction project) or visual documentation of a product, event, or performance accompanied by descriptions of knowledge and skills learned (e.g., a photograph of a student replacing a carburetor accompanied by text explaining the knowledge and skills used).

Work samples can be used to show both the depth and some breadth of student learning. For example, the C-TAP portfolio requires four work samples. Each work sample must show strong evidence of at least one required standard. Collectively, the four samples should illustrate proficiency in relation to the range of required standards. Table 4.9 shows several examples of C-TAP work samples and the standards they address.

C-TAP teachers believe that work products like these are the most important part of the portfolio because they provide direct and powerful evidence of students' career-technical knowledge and skills and their ability to apply that knowledge and those skills in a variety of ways.

Table 4.9 Examples of C-TAP Work Samples

- Computer-generated reports in different formats on the demographics of the 50 states, with annotations of optimal uses for each format *(Business Education: Computer Science and Information Systems Cluster. Standards: Computer Applications, Document Processing, Computer Systems, File Management)*
- Storyboard for a preschool lesson based on "The Hare and the Tortoise" *(Home Economics Careers and Technology: Child Development and Education Cluster. Standards: Child Growth and Development, Positive Interaction, Guidance and Discipline, Developmentally Appropriate Activities)*
- Scale drawings and production cost estimates for a grandfather clock *(Industrial and Technology Education: Construction Technology Cluster. Standards: Planning and Layout)*
- Brochure comparing autologous and allogenic bone marrow transplants in terms of procedure, length of patient hospital stay, costs *(Health Careers Core. Standards: Human Growth and Development; Socioeconomics)*

For each work sample, students must submit a written summary that briefly describes the work sample, identifies the specific knowledge and skills demonstrated, and explains what was learned as the work sample was completed. Writing work sample summaries requires students to reflect on and evaluate their work in order to identify evidence of standards-based achievement. Completed summaries can be used to help teachers or others interpret and evaluate the students' work samples.

Writing Sample

The C-TAP writing sample requires students to investigate and write about a standards-based topic of their choice. By completing this task, students demonstrate important skills, including their abilities to obtain and evaluate information and data; analyze, evaluate, and organize information; and communicate effectively in writing (e.g., using correct grammar and spelling, attending to audience and purpose). At the same time, they learn about, and demonstrate understanding of, their selected topic. Some examples of C-TAP writing sample topics and the standards to which they relate can be found in Table 4.10.

Table 4.10 Examples of C-TAP Writing Sample Topics

- **Livestock and Internal Parasites:** The types of parasites that affect various livestock; the symptoms associated with the parasites; their impact on the health of the animal; and, finally, the methods of control and elimination currently available and their costs. *(Agricultural Education: Animal Science Cluster. Standard: Animal Parasites and Pests)*
- **The Changing Technology of Computer Hardware:** A brief historical summary, including a picture of what is current and a forecast for future development or innovations. *(Business Education: Computer Science and Information Systems Cluster. Standard: Changing Technology)*
- **Marketing to Multicultural Buyers:** A description of how marketing strategies are changing to meet the needs of multicultural consumers, with an assessment of the impact of these new marketing strategies on profits. *(Business Education: Marketing Cluster. Standards: Marketing Principles, Promotion)*
- **Documenting a Site Visit:** A summary of conclusions about safety issues, construction procedures, organization, working conditions, and worker morale drawn from two visits to a residential construction site during the construction of a new foundation. The summary is based on observations as well as interviews with two workers. *(Industrial and Technology Education: Construction Technology Cluster. Standards: Assembling Processes, Finishing Processes, Residential Construction, Work Site Safety, Site Preparation)*

Initially, the C-TAP portfolio required a formal research paper that heavily emphasized the collection and synthesis of information, as well as the demonstration of proficiency in academic writing. When implementing C-TAP, however, some teachers reported that this formal approach was not necessarily the most appropriate vehicle for measuring students' ability to express career-technical knowledge in writing, since the writing required in most career-technical fields is different than that used for research papers.

The current writing sample approach used by C-TAP is more flexible, stressing student interest and expertise as a starting point. In one health careers class, for example, a student wrote about the identification and treatment of breast cancer. She had a personal interest in this topic because her mother had been diagnosed with the disease. Her writing sample was also clearly linked to a specific and important standard (i.e., Health Maintenance).

As with project topics, teachers may need to help students choose appropriate topics for their writing samples. When doing so, they should provide advice (rather than prescriptions) that help students link their work

to standards. Some C-TAP teachers allow students to include in their portfolio a paper written for one of their academic classes such as English or history. In such cases, students and teachers need to clarify, in advance, which standards are addressed by the paper and how the paper relates to a particular career path. For example, a paper addressing the causes of the Civil War would not be well-matched to any career-technical standard. However, a paper focusing on hygiene in Civil War camps or construction techniques used for pre-Civil War era forts might relate to career-technical standards and the student's career path.

Supervised Practical Experience Evaluation

The Supervised Practical Experience (SPE) Evaluation form, which is completed by a student's work supervisor, gives students feedback on their performance in the workplace. This portfolio entry provides important documentation of students' work-related experience and can be used to evaluate not only job-specific but general workplace readiness skills (e.g., time management, organization, communication). Because it is typically completed by an outside supervisor or reviewer and not a classroom teacher, it is a direct link to the world of work, providing real feedback about workplace skills. If a student has not had a formal work placement, he or she can be evaluated via the SPE form by a teacher or other appropriate adult on the basis of community volunteer work, school club work, or other activities that require workplace readiness skills. This is an optional component for the C-TAP portfolio.

Challenges Associated with Developing and Implementing Cumulative Assessments

The general process for developing and refining a cumulative assessment like the C-TAP project or portfolio is similar to that used for developing written on-demand assessments (see Chapter 3). Although a test blueprint is not expressly created, all other steps in the development process remain the same. To recap, the steps are 1) define the purposes of the assessment; 2) determine the number and range of standards that students must demonstrate; 3) develop the assessment itself (i.e., basic structure and specific requirements for completion); 4) review the assessment prior to use to ensure links to standards, clarity, and freedom from bias; 5) "try out" the assessment with teachers and students in classrooms; 6) analyze student responses to identify places where instructions or expectations are unclear or

where tasks fail to elicit the type or depth of performance desired; and 7) revise and improve the assessment based on insights gleaned from the entire development process.

What makes this development process different from that used for written on-demand assessments is the level of ongoing effort required by teachers and students as the cumulative assessments are tried out. To realize the advantages of cumulative assessments, both teachers and students must respond to challenges different from those required by written on-demand assessments. Among these challenges are the following:

- adjusting to shifting roles and relationships among teachers and students;
- planning and managing complex, long-term assessment work;
- providing student choice while ensuring links to standards;
- promoting effective reflection and self-evaluation by students;
- documenting progress and products effectively; and
- anticipating needed resources.

Each of these challenges is discussed in more detail below. An additional challenge associated with implementing cumulative assessments, reliably scoring student work, is discussed in Chapter 5.

Adjusting to Shifting Roles and Relationships among Teachers and Students

Implementing cumulative assessments requires both teachers and students to redefine their traditional roles and how they relate to one another. Teachers must relinquish some control, allowing students to assume considerable responsibility for their learning and to work more independently. They must also create instructional environments that promote student responsibility. As students begin to make more choices and decisions about their learning, teachers must learn to work effectively as coaches, carefully orchestrating a delicate balance of demands and support. In other words, teachers must encourage student independence while at the same time providing the support and guidance that students need to succeed on cumulative assessments. In addition, as part of the assessment process, teachers must monitor students' work for information that can help shape future learning goals and inform instruction (e.g., students' misconceptions

or gaps in understanding and skills). Both teachers and students will need time to become comfortable and effective in their new roles.

Planning and Managing Complex, Long-Term Assessment Work

Both projects and portfolios require considerable time to complete, even when effectively integrated into the existing curriculum. Teachers using C-TAP assessments have learned from experience that portfolio work samples and project products that effectively demonstrate proficiency in one or more standards cannot be churned out in a single day. They require several days or sometimes months and adequate time for revision and improvement. Even portfolio components that seem relatively straightforward, such as the letter of recommendation required by the C-TAP portfolio, can take more time to complete than anticipated. Students who wait until the last minute to request such letters may be surprised when employers or advisors are unable to write them overnight.

For these reasons, time management is a critical issue for both teachers and students when using cumulative assessments. Teachers need time to plan how the portfolio or project assessment will unfold in their classrooms. Such planning should begin well before the school year starts. Once the school year has begun, teachers will need further time to do the following:

- familiarize students with standards and the specific requirements of the cumulative assessment(s) they will be completing;
- provide relevant instruction; and
- provide ongoing feedback and support that will help students revise and improve their work over time.

In addition, for best results, teachers should introduce standards early in the school year so that students understand (at least in general) what they must demonstrate before beginning assessment work. One way to introduce and continually reinforce standards is to reference them frequently during instruction and teacher-student conferences. Many teachers also find it helpful to develop a large poster for each standard and to hang the posters in the classroom as a reference. When planning instruction, teachers should allow adequate time to teach students the knowledge and skills needed to demonstrate proficiency on any standards-based cumulative assessment being implemented. Building in time for feedback and revision provides students with opportunities to strengthen their work based on new insights and deepened understanding.

Experienced C-TAP teachers recommend that all teachers devise strategies to ensure that both they and their students stay on track and on time when administering and completing cumulative assessments. Table 4.11 provides an example of a form that can be used to plan and monitor project or portfolio progress.

Table 4.11 **Project (or Portfolio) Timetable**

Student: _____

Class: _____

Date: _____

Teacher: _____

Activity:	Time Required	Expected Completion Date	Teacher Sign-off (with date)	Actual Completion Date

Finally, an obvious, but significant, challenge is the planning required to manage and store student work associated with cumulative assessments. As students work on projects and portfolios, they will generate and collect various materials to meet the assessment requirements and to show evidence of their progress and learning over time. Students will need to access these materials regularly and should be encouraged to manage the materials they collect and produce. This requires careful thought by the teacher on how to organize assessment materials within limited space and how to structure students' access to the materials. (Some suggestions for how to manage the day-to-day flow of assessment materials can be found in the *Career-technical Assessment Program (C-TAP) Teacher Guidebook.*)

Providing Student Choice while Ensuring Links to Standards

A key feature of cumulative assessments is that they engage students' interests and encourage them to assume greater responsibility for their own learning. Giving students freedom to design and create projects and portfolios around personal interests and strengths is not, however, the same as giving them free reign. Though their work should be personally meaningful, it must also relate to standards. To make this possible, students may need assistance in forging links between their own interests and work, and the standard(s) they must demonstrate. Teachers can help by meeting with individual students or groups of students to brainstorm interesting standards-based topics (e.g., for projects and portfolio writing samples). In addition, they can regularly review students' work-in-progress to ensure that the work is appropriately linked to designated standards.

Facilitating Student Reflection and Self-Evaluation

During cumulative assessments, students are asked to reflect on and evaluate their work at numerous points in time. It is likely that students not practiced in reflection or self-evaluation may need help. At first, their reflections and insights may be somewhat irrelevant or shallow and, therefore, limited in usefulness. For example, students may reflect on aspects of their work that are unrelated to standards (e.g., their feelings about doing a project, their working relationship with a teacher or supervisor). Or, they may offer only global judgments such as, "I think I did a good job," or, "I know I need to work on this piece more." This is adequate as a beginning, but students must learn to think more analytically about their work if cumulative assessments are to be truly meaningful learning and assessment experiences. They must be taught to recognize the connections between their work and the standards, to identify strengths and weaknesses in their efforts, to brainstorm ideas for improvements, and to acknowledge their progress over time.

There are many strategies that teachers can use to help students develop their reflective and evaluative capacities. Some examples are provided below.

- Regularly ask students probing questions that require them to reflect on or evaluate their own or others' work (e.g., How does this piece of work show mastery of a standard? What do you see as the strengths of this piece of work? What would you do to improve this piece of work? What did you learn as a result of completing this work?). This can be done during whole class discussions or one-on-one teacher-student conferences.

- Have students work individually or in pairs to compare two or more pieces of student work (i.e., how they are alike and different) and identify both the strengths and weaknesses of each.
- Design and implement instructional activities that will help students practice and refine specific skills used in cumulative assessments (e.g., writing exercises that require students to write a rough draft and to then improve the draft based on their own and others' feedback).
- Provide students with support materials that help remind them of issues to think about when completing particular aspects of cumulative assessments (e.g., formatted worksheets that include questions students should consider when planning a project or summarizing a portfolio work sample).
- Nurture a culture of inquiry in the classroom, where mistakes are viewed as vehicles for learning, self-assessment, and problem solving.

Documenting Progress and Products Effectively

Documenting student work (both work-in-progress and final products) in ways that effectively illustrate what students know and can do is a challenging task that requires careful consideration by students and teachers. Students, with the help of their teachers, must decide not only what knowledge and skills to document, but also what form that documentation should take to be meaningful and informative to others.

Projects or portfolio work samples that focus on the creation of something physical (e.g., a piece of furniture) are fairly easy to document. Students can show the product itself, both finished and at various stages of completion, as evidence of their knowledge and skills. Some projects and work samples, however, are centered around performing technical procedures, organizing and conducting events, or other activities that do not result in physical end products (e.g., teaching a lesson at a child-care center). Documenting the progress and results of these kinds of projects or portfolio work can be more difficult. Students will need to use photographs, videotapes, audiotapes, interview transcripts, or other creative measures to document their knowledge and skills. This is as true for documenting evidence of progress as it is for showing the final product.

When planning how to document completed work and work-in-progress, students need to first consider the various methods available to them (e.g., not all students have access to videotape equipment) and then determine which

method(s) will best convey the essence of an activity and the knowledge and skills applied. For example, suppose that a student plans to interview health care professionals as part of a project assessment. The student wants to document this activity to meet three goals: to show evidence of progress, to show his or her ability to research and gather topic-related information, and to show his or her ability to make connections with community resources. There are several methods the student could use to document this activity, including typing a list of the interview questions, taping the interviews and transcribing the tape(s), or videotaping the interviews. When deciding which method of documentation to use, the student must carefully consider which method would best allow him or her to meet the goals above, as well as which is most feasible given the resources available.

Though students may be required to formally submit only a limited amount of evidence of progress, they may find it helpful to make documentation an ongoing activity during any cumulative assessment. A record of what has happened up to any point in a project or portfolio assessment, including what is working and what is not, is invaluable as students review their progress, evaluate their efforts, and make appropriate adjustments. Keeping a journal, saving draft copies of work and research notes, photographing key activities, and documenting informative conversations with the teacher, supervisor, and others are several ways that students can regularly document their progress.

Anticipating Needed Resources

Completing and documenting cumulative assessment work may require a variety of resources not always immediately available in classrooms (e.g., computers for word processing and creating charts or graphics, cameras for photographing work in progress). Districts and teachers may need to help make such resources available during the assessment process. When completing C-TAP assessments, for example, students usually need access to professionals and workplaces in career-technical fields in order to gather information for projects and writing samples or to obtain work experience. Schools or districts should spend time building new, and nurturing existing, relationships with businesses and other organizations that provide work-related experiences or staff willing to meet with students. These relationships are not only valuable resources, but can help create good will for the district and keep the community aware of what schools are doing.

Completion of projects and portfolios does not necessarily require the use of technically sophisticated equipment. However, the types of technological

resources available to students can affect the "presentation" quality of portfolios and some projects. Schools and districts may need to take measures to ensure that differences in available resources from one classroom or school to the next do not unfairly advantage or disadvantage students with regard to the quality of "presentation" they are able to produce.

Summary

Cumulative assessments are completed over time and demonstrate the best students can do when given opportunities to practice and revise their work based on self-evaluation and constructive feedback from others. They typically measure the depth of students' knowledge and their ability to apply knowledge and skills in meaningful, often hands-on, ways. Unlike written on-demand assessments, cumulative assessments encourage high levels of student involvement in and responsibility for learning. Throughout the assessment process, students are actively engaged in establishing learning goals, evaluating their own efforts, and revising and improving their work over time. Evidence of students' improvement or achievement, or both, can be gleaned from their final assessment products and from the various efforts (i.e., process) leading up to them.

There are several advantages and disadvantages associated with cumulative assessments. Among the advantages are their capacity to inform instructional planning, engender as well as reveal learning, and provide multiple ways for students to express what they know. With regard to disadvantages, cumulative assessments tend to be less reliable and efficient than written on-demand assessments. They cover fewer standards due to practical time constraints and can be prohibitively expensive to develop, implement, and score.

Two types of cumulative assessments that have become increasingly popular over recent years are project assessments and portfolio assessments. A project assessment involves in-depth, hands-on exploration of a topic, theme, idea, or activity, resulting in a product, performance, or event for assessment. Project assessments typically focus on one or two content standards and measure students' ability to apply knowledge and skills in authentic situations. They can also assess how well students are able to evaluate their own work, solve problems, plan and carry out complex activities, and communicate findings orally and/or in writing. Most project assessments involve four basic steps: 1) planning the project; 2) researching and developing the project; 3) producing a final product, performance, or event; and 4) presenting the final product. As with all cumulative assessments,

evidence of student achievement can be collected during each major step of the process.

A portfolio assessment involves the structured collection of student work that documents students' knowledge and skills. Most portfolio assessments require students to complete different types of work (e.g., writing samples, work products associated with hands-on projects, various career-related materials) that grow out of classroom curriculum and work-based experiences. Each piece of work typically reflects knowledge and skills related to at least one standard. As a collection, the portfolio work demonstrates student progress and/or achievement vis-à-vis multiple standards. Most portfolio assessments aim to measure the depth and some breadth of students' knowledge, and their ability to apply both knowledge and skills in a variety of authentic contexts. They can be used to document learning and improvement over time, showcase overall achievement following a program or course of study, or both.

The structure of portfolio assessments is likely to vary from one site (e.g., classroom, school, district) to the next depending on the purposes of assessment and the nature of the standards being assessed. Most portfolio assessments do, however, provide some basic structure, or framework, within which student work can be completed and submitted. Typically this structure outlines the specific standards and range of standards to be demonstrated and the quantity and types of work to be submitted.

The process for developing cumulative assessments is similar to that used for written on-demand assessments. Assessment tasks must be drafted, reviewed, tried out, revised, and tried out again until they effectively elicit evidence of performance in relation to targeted standards. To realize the advantages of cumulative assessments, both teachers and students must respond to challenges different from those required by written on-demand assessments. Among these demands are 1) adjusting to shifting roles and relationships among teachers and students; 2) planning and managing complex, long-term assessment work; 3) providing student choice while ensuring links to standards; 4) promoting effective reflection and self-evaluation; 5) documenting progress and products effectively; and 6) anticipating needed resources.

Educators interested in implementing the C-TAP project and portfolio should consult the C-TAP Teacher and Student Guidebooks. (See Appendix D for more information.)

Developing an Effective Scoring System

This chapter addresses the complex issues related to scoring student assessments. As stated previously, the development of an assessment's scoring system actually takes place at the same time as the development of the assessment itself. However, for most assessments, developing an effective scoring system is a complex and challenging process, one that often gets insufficient attention during the assessment development process. This chapter focuses solely on scoring issues in order to help schools and districts better understand both the importance of an assessment's scoring system and how to develop a scoring system that is effective.

Scoring systems are means of interpreting the relationship between standards (e.g., core academic, career-technical) and student achievement. Depending on the nature of the assessment being scored, the scoring system can be very simple and straightforward or very complex. Scoring multiple-choice tests is at the easy end of the spectrum. For tests such as these, which require students to select the correct response from a limited set of options, the scorer simply has to determine if a student has indeed selected the predetermined correct response. So simple and straightforward is the scoring process that machines usually do the scoring.

At the other end of the scoring spectrum, however, is the scoring of cumulative assessments such as portfolios and projects. These assessments require a student to independently produce a response, with a wide range of possible options available, and these options may differ in a variety of ways. In other words, there is no one right answer but sometimes a multitude of right answers. For these assessments, the scorer must interpret a student's response to determine its adequacy in relation to the appropriate standard(s). The process of interpreting a student's response to a cumulative assessment is not a simple or straightforward task, but rather a very complex task because of all

the variables involved. For this reason, the majority of this chapter applies to the scoring of cumulative assessments and to those written on-demand assessments that require a student to independently produce an answer.

The chapter begins with an overview of the activities involved in developing an effective scoring system and then elaborates on how each activity can be accomplished. While the information presented applies equally to the scoring of classroom assessments and districtwide assessments, developers should heed the following caveat: for high stakes assessments (i.e., those with significant consequences such as whether a student graduates), the degree of validity and reliability required demands complex technical procedures that are beyond the scope of this document. However, the scoring principles laid out here can help provide a knowledge base for collaborating with district staff or outside testing and measurement experts to design a high stakes assessment system that is consistent with local needs.

Developing an Effective Scoring System: An Overview

As mentioned in Chapters 3 and 4, the development of assessments and their accompanying scoring systems is an iterative process: Items or tasks are drafted, tried out, statistically analyzed based on student responses, revised, and tried out again until the standards, the assessment instrument, and the scoring system achieve a satisfactory match. Within this process, there are specific types of activities involved in developing an effective scoring system. These activities are summarized in Table 5.1.

Table 5.1 Activities in Developing a Scoring System

- Developing a scoring plan
- Drafting scoring scales for performance assessments
- Checking for validity
- Checking for reliability
- Choosing a cut score to reflect the performance standard

For the most part, these activities are accomplished as the assessment itself is developed. As was the case for developing the assessment, the development activities listed in Table 5.1 often take place more than once. For example, checking for validity will be done at several points during the development process, and the scoring plan and scoring scales may be revised

several times following tryouts based on patterns seen in student responses. However, selecting a score to represent the desired performance standard occurs after the scoring system is completely developed.

While teachers are not likely to try out and revise assessment items to the same extent as developers of districtwide assessments, it is still important for teachers to understand the process for developing an effective scoring system, and to engage in activities, when appropriate, to ensure that all assessment items or tasks accurately reflect student abilities as well as standards.

Each activity involved in developing an effective scoring system is described more fully below.

Developing a Scoring Plan

The development of a scoring system actually begins with the identification of the standards to be addressed by the assessment (see Chapter 1, "Identifying Standards"). Without the identification of these standards, an effective scoring system cannot be developed.

Once standards have been identified, the first step in developing an effective scoring system is to create a scoring plan. A scoring plan contains information that is part of the assessment blueprint described in Chapter 3. A scoring plan is that part of the blueprint that describes the specific standards-based knowledge and skills to be measured by the assessment and the method(s) of scoring to be used (e.g., holistic, analytic). For written on-demand assessments, the scoring plan also includes the assessment blueprint information that identifies the number or percentage of items addressing each standard. For cumulative assessments, such as projects and portfolios, the scoring plan also includes the blueprint information that identifies how the targeted standards are reflected in the project steps or portfolio entries that comprise the assessment. (Note: For both projects and portfolios, the targeted standards may identify specific core academic or career-technical standards that students must demonstrate, but often do not identify additional standards (e.g., occupation-specific standards that students may also demonstrate should they choose to do so.)

The next two sections provide additional information about the type of information included in a scoring plan. Specifically, the first section addresses how a scoring plan might describe the relationship between assessment items/tasks, standards, and scores. The second section addresses a scoring plan's description of the scoring method(s) used, specifically the selection of such scoring method(s).

Relating Assessment Items and Tasks to Standards and Scores

In order for scores to accurately reflect student performance on one or more standards, the assessment must be designed so that the assessment items or tasks are directly related to the targeted standards. For multiple-choice tests or mixed-method written on-demand assessments such as the ACE exams (i.e., a combination of multiple-choice and short written-response items), the scoring plan should specify the actual number of items that will measure each standard. For more complex assessments such as projects or portfolios, the scoring plan should specify how each assessment step or task comprising the assessment relates to the targeted standards.

A single score may represent proficiency related to a single standard or to multiple standards. For example, the C-TAP portfolio assessment yields several scores in different areas called scoring dimensions. Two of these scoring dimensions are "Content" and "Career Preparation." In the C-TAP portfolio assessment, the scores for the Content and Career Preparation dimensions reflect not one standard but multiple standards. When assessments can be broken down into multiple tasks or parts, it is useful to include in the scoring plan a "map" that indicates which parts of the assessment are likely to provide information for each standard or score. Table 5.2 shows an example of such a map for the C-TAP portfolio assessment. The assessment dimensions are listed on the left side of the map, and the headings at the top refer to the different parts (i.e., entries) of the C-TAP portfolio assessment. Each X indicates a portfolio entry that is designed to produce student responses related to a specific scoring dimension. Each (X) indicates an entry that is not designed to produce information related to a specific scoring dimension, but which may produce student responses that independently address the dimension. Blank areas indicate that the entry was not designed to, and is unlikely to, produce information for that dimension.

Table 5.2 Map of Dimensions to Portfolio Entries

	APPLICATION	LETTER OF RECOM-MENDATION	RESUME	WORK SAMPLES	WRITING SAMPLE	SUPERVISED PRACTICAL EXPERIENCE EVALUATION	TABLE OF CONTENTS	LETTER OF INTRODUCTION
CONTENT ■ Knowledge of major ideas and concepts in career-technical standards ■ Knowledge of how skills in career-technical areas are applied				X	X	(X)		
CAREER PREPARATION ■ Career planning ■ Personal qualities needed for employment	X	X	X	(X)		X		X
ANALYSIS ■ Evaluation of own skills and work ■ Investigation and information gathering			X	X	X	(X)		X
COMMUNICATION ■ Attention to audience ■ Using own ideas ■ Organization and clarity ■ Accuracy, neatness, and completeness ■ Language mechanics and sentence vocabulary	X		X	X	X		X	X

By preparing a map such as that shown in Table 5.2, developers can more easily identify when it is likely that insufficient information is being gathered on a particular dimension or when so much information is being collected that some tasks might be dropped without affecting the score. Such a map can also help scorers identify likely sources of evidence in student responses for each dimension, as well as identify where lack of evidence should be interpreted as a sign of a lack of understanding, since the entry was clearly designed to produce evidence related to the specified dimension.

Selecting Scoring Method(s) and Approach(es)

A scoring plan should also indicate which scoring method(s) and approach(es) will be used to score the assessment. As mentioned earlier, the method used for scoring multiple-choice tests and similar assessments is relatively straightforward, while the method for scoring cumulative assessments is more complex. For multiple-choice and similar tests, the scorer (often a machine) simply determines whether the student has correctly identified the right answer. If, in addition to a total score (i.e., the total number of items identified correctly), a set of subscores is desired, then the number of correct answers is counted for the relevant subsets of items.

The method for scoring cumulative assessments and other complex assessments (e.g., written on-demand assessments such as written scenarios) is more problematic, since the range of possible responses is large. Moreover, different scoring methods and approaches may be used to score the possible responses. The scoring plan should specify exactly which method(s) and approach(es) will be used. Two scoring methods that are commonly selected for use are the holistic scoring method and the analytic scoring method. Two types of scoring approaches commonly used are task-based scoring and dimensional scoring. These scoring methods and approaches are described below.

Holistic versus Analytic Scoring Methods

A scorer using the *holistic* scoring method views a student's response to an assessment as a whole or, in other words, as an integrated performance. Using this method, a scorer considers information about specific aspects of the performance only in relation to their contribution to the overall impression left by the entire performance. Holistic scoring is often used for complex assessments such as portfolios, projects, and written scenarios that

have several different features or parts that, when combined together, form an overall integrated performance. By definition, the holistic scoring method results in a single score or narrative description that summarizes the performance as a whole. This method is not suitable for assessments that consist of many multiple-choice, matching, or short written-response items because such assessments are not intended to reflect an integrated performance but rather a diverse sample of the student's knowledge, skills, and abilities in a particular career or academic area.

A holistic score is usually assigned on the basis of a scale (e.g., a rubric) that describes characteristics of performances at different score points. However, the application of a scoring scale is not self-evident. To be most meaningful, a scale must be accompanied not only by descriptors for each point on the scale, but also by sample performances, sometimes called benchmarks, that help scorers, teachers, and students understand how the scale is reflected in actual student performances. Using benchmarks in scorer training sessions helps scorers develop the knowledge and skills needed to score performances effectively. Using benchmarks in scoring practice is especially helpful when student performances exhibit qualities corresponding to more than one score point. A clear understanding of the overall nature of the performance that is represented by each score point is necessary to apply the scoring scale to uneven performances.

Table 5.3 shows the holistic scoring scale used to assess the C-TAP portfolio as a whole. The four points in the scoring scale include two levels of satisfactory performance: "Advanced," where the student has met the standards with distinction, and "Proficient," where the student has met the standards but not excelled. The other two score points represent levels of unsatisfactory performance: "Basic," where the student does not meet the standards at the present time, but shows promise of meeting the standards with some additional focused work, and "Limited," where the student does not come close to meeting the standards and may need substantial remediation. These performance levels are ordered such that each level of the scoring scale represents a point on a continuum ranging from weaker to stronger performances.

When determining the overall (holistic) level of performance for the C-TAP portfolio, several aspects or dimensions are considered: content (i.e., knowledge and application of knowledge and skills related to career-technical standards), career preparation (i.e., the ability to plan and prepare for a career), analysis (i.e., self-evaluation and investigative skills), and communication (i.e., the ability to communicate in writing effectively).

Table 5.3 C-TAP Portfolio Holistic Scoring Scale

Limited: Shows little or no content knowledge and application of content knowledge and skills related to the career-technical standard(s); shows little or no ability to prepare for a career; self-evaluation skills are weak; fails to present work effectively.

Basic: Shows gaps in content knowledge and/or application of content knowledge and skills related to the career-technical standard(s); shows some ability to prepare for a career, but major weakness(es) may be evident; demonstrates vague or sketchy self-evaluation skills; overall presentation makes some of the work difficult to understand.

Proficient: Shows adequate* content knowledge and application of knowledge and skills related to the career-technical standard(s); shows adequate ability to prepare for a career; demonstrates adequate self-evaluation skills; overall presentation is organized, making most of the work easy to understand.

Advanced: Shows superior content knowledge and application of knowledge and skills related to the career-technical standard(s); shows superior ability to prepare for a career; demonstrates superior self-evaluation skills; overall presentation is well-organized and effective, making all of the work easy to understand.

(*adequate = satisfies requirements)

In contrast to the holistic scoring method, a scorer using the *analytic* scoring method views a student's response to an assessment in parts. Using this method, a scorer rates different aspects of a performance separately and usually, but not always, combines these separate ratings into an overall score. Sometimes, some aspects of performance are deemed to be more important than others and therefore their scores are given more weight when calculating the overall score. This is analogous to scoring plans for multiple-choice tests where the different percentage of items devoted to specific topics or skills reflect their perceived importance in the subject or career area being assessed. Analytic scoring systems not only provide an overall score of the student performance but also have the potential to provide additional scores for specific standards or sets of standards. Table 5.4 illustrates an analytic scoring guide for different aspects of a written report on a specific science experiment (i.e., Exploring the Maple Copter) that is designed, conducted, and documented by groups of students. The experiment focuses on what makes maple seeds twirl as they fall to the ground. Each aspect included in the scoring guide uses a different scale, ranging from 0 to ≥ 6 or from 0 to 4 score points.

Table 5.4 Sample Analytic Scoring Guide for a Science Assessment

EXPLORING THE MAPLE COPTER

Part II: Group Experimentation

Directions: For each criterion below, circle the letters of the standards for which students have provided sufficient evidence in their written work. Then add the numbers of standards met and circle the corresponding total from 0 to ≥ 6 in the column to the right.

Criteria and Standards	Performance Levels				
	Excell.	Good	Fair	Poor	No Evid.
II.1 Identification of relevant factors	≥ 6	4-5	2-3	1	0
a. Total mass					
b. Distribution of mass					
c. Surface area, length and wing					
d. Shape and curvature					
e. Air (e.g., currents, pressure)					
f. Materials (e.g., seed's moisture, vein's structure)					
g. Dropping position					
h. Physical forces					
II.2 Experimental design	4	3	2	1	0
a. Matches the factor to be studied					
b. Defines independent and dependent variables					
c. Controls variables, when possible					
d. Includes description of model used					
II.3 Data collection and presentation	4	3	2	1	0
a. Sufficient repetitions of measurements					
b. Mathematical treatment of data (e.g., averages)					
c. Appropriate presentations (labeled charts, appropriate graphs)					
d. Adequate description of procedures					
II.4 Conclusions	4	3	2	1	0
a. Related to studied problem					
b. Supported by experimental findings					
c. Appropriate generalization					
d. Include discussion of effect of errors					

From Baron, 1996, p. 186.

Both the holistic and the analytic scoring methods can satisfy the demands of students, parents, teachers, policymakers, and others for a score that represents overall achievement. Holistic scoring is most appropriate for complex performances where the overall impact is of most interest, especially where extreme performances on one or more aspects can outweigh performance on other aspects. It also can accommodate two very different performances that have a similar effect. Although holistic scoring is often considered to take less time than analytic scoring, this may not hold true for lengthy or extremely complex performances where a scorer must identify and weigh many different, often contradictory, pieces of information. Analytic scoring is most appropriate when information is desired about both the overall performance and different aspects of the performance. The subscores can provide valuable diagnostic information about specific strengths and weaknesses of individual students, classroom instruction, and programs in different career areas.

Task-based versus Dimensional Scoring

Once a school or district has decided on whether to use the holistic or analytic scoring method (i.e., to assess a performance as a whole or in parts), still another decision must be made. In addition to including a description of the scoring method, a scoring plan should also include a description of the scoring approach to be used with the scoring method. The two scoring approaches used most often are the task-based scoring approach and the dimensional scoring approach. Both approaches may be used with both scoring methods (i.e., holistic and analytic), but they are most commonly used with the analytic scoring method.

A scorer using the *task-based* scoring approach evaluates an assessment by focusing on the assessment's different pieces or tasks. A scorer using this approach with the analytic scoring method would typically consider the quality of each task, give each task a separate rating, and then combine these separate ratings into an overall score for the assessment. The task-based scoring approach is often used when an assessment's tasks are relatively independent. For example, the different tasks or entries of the C-TAP portfolio assessment (i.e., Portfolio Presentation, Career Development Package, Work Samples, Writing Sample, and Supervised Practical Experience Evaluation) could be scored separately because each task is independent of the others (although many or all of the tasks are often connected by content). The Career Development Package entry could be given a score that reflects a student's mastery of selected career preparation standards, and the Writing Sample entry could be given a score that reflects

a student's mastery of core academic (i.e., writing) standards. (The C-TAP portfolio assessment is not scored using this approach because the assessment developers chose to view the portfolio as an integrated display of student knowledge, skills, and abilities in a particular career-technical field, not as a collection of separate tasks.)

A scorer using the *dimensional* scoring approach evaluates the assessment by focusing specifically on the assessment's different dimensions (i.e., the important elements that the assessment is designed to measure). A scorer using this approach with the analytic scoring method would typically assign a score to each dimension and then combine those scores to arrive at an overall score. Some examples of dimensions of different assessments are language mechanics for a writing assessment, problem solving for a mathematics assessment, and career preparation for a career-technical assessment. In multi-tasked assessments, the dimensions almost always cut across the different tasks. Looking back at Table 5.2, which shows the dimensions of the C-TAP portfolio assessment and the different tasks comprising the assessment, it is clear that not all dimensions are reflected in every portfolio entry, but that all dimensions are represented in multiple entries. This helps ensure that enough evidence is collected to assign each dimension a score.

The task-based scoring approach is best used when the tasks themselves are considered important independent features of a student's performance. For example, an assessment in the career area of Computer Science and Information Systems may include a task that requires a student to show skills in producing documents using appropriate software. If mastery of this skill is considered necessary to meet industry standards, then this assessment should probably be scored using the task-based scoring approach. The dimensional scoring approach is best used when the features or dimensions to be evaluated occur across, or are integrated within, different assessment tasks. The knowledge and skills measured by each dimension, however, must be relatively independent so that scorers can be trained to reliably assign specific evidence in a student response to one dimension or another. If dimensions are too similar, then it is difficult to achieve consistency in the assignment of evidence and, therefore, in scoring. When dimensions are strongly interrelated (i.e., similar), then they need to be reconceptualized, perhaps by combining some dimensions or eliminating others and creating new, more independent, dimensions. If this is not possible, then another scoring approach or scoring method (i.e., holistic) may be indicated.

As mentioned earlier, both the task-based scoring approach and the dimensional scoring approach can also be used with the holistic scoring method. A scorer using either approach with the holistic scoring method would use the approach simply as an aid to judge the performance holistically. For example, a scorer evaluating a C-TAP portfolio uses the holistic scoring method to give the portfolio an overall score. In order to arrive at that holistic score, however, the scorer is guided by looking at the different dimensions of the assessment. Although a holistic score reflects an overall performance, there are always aspects of that performance that a scorer reflects upon in order to make an holistic judgment. Either formally or informally, a scorer may use the task-based or dimensional scoring approach as a guide in holistic scoring.

Drafting Scoring Scales for Assessments

Another important step in developing an effective scoring system is to draft a scoring scale for each assessment used. A scoring scale is a system of classifying assessment performances in a progressive graduated series of points, grades, levels, or degrees, with one end of a scale always indicating higher level performances than the other end of the scale. For assessments consisting of items where the student responses are unambiguously right or wrong (e.g., multiple-choice or matching items), the scale is always based on the number of correct responses. For example, a multiple-choice test of 100 items could have a scale of 0-100 points, with a score of 100 indicating a higher level performance than a score of 0. Scoring scales can also be created for these assessments to reflect scoring levels of achievement (e.g., Limited, Basic, Proficient, Advanced) by assigning a range of test scores to each level. Thus, for a multiple-choice test of 100 items, a scoring scale could be developed in which the range of 85 to 100 reflects a performance level designated as Advanced. Other scales for such assessments (e.g., those used for the SAT) may be developed to reflect the comparability of scores between test administrations or the relationship of one student's score to the distribution of scores from all students who took the test. Developers who desire these types of scales, however, need to consult with district or external experts in testing and measurement, as the process for developing these scales is very complex and beyond the scope of this introductory document.

For assessments consisting of student performances that require human judgment to interpret (e.g., portfolios, projects, assessments with short and long written-response items), scoring scales are based on the different levels

of student performance. Drafting scoring scales for these assessments (hereafter referred to as performance assessments) is a complex task because, while consensus is often easily reached on identifying "very good" and "very bad" performances, consensus on what constitutes an "adequate" performance is likely to require substantial discussion. For this reason, careful thought must be given to developing a scoring scale that clearly identifies the elements that distinguish higher-level performances from lower-level performances. The remainder of this section describes the process for developing scoring scales for performance assessments.

Scoring scales, which are basically descriptors of different levels of student performance, may be developed for use with the holistic or analytic scoring method. If assessment developers are familiar enough with student work, they can often draft the scoring scales in the initial assessment development phase. If not, then the scoring scales should be developed during the tryout and analysis phases. During the tryout phase, examples of student responses should be collected that clearly represent strong, average, and weak performances. Several examples should be collected to represent each performance level. These examples should then be analyzed against the relevant content standards to identify the specific characteristics that distinguish the different performance levels. If the analysis is done by a committee, the characteristics are usually discussed until a consensus is reached as to the language to be used in the performance level descriptors. If classroom teachers are developing their own scoring scales, it helps to discuss the characteristics with one or more colleagues.

Scoring scale descriptors should focus on characteristics that are present in a student response, not those that are absent. They should also focus on the elements that are being measured by the assessment, and omit reference to any element that is not being measured. For example, Table 5.5 shows how the descriptor for the "Limited" performance level of the C-TAP portfolio scoring scale was initially written and how it was revised, after much discussion, to focus on characteristics present in the student response and those measured by the assessment. In the initial version, the descriptor began with "Completes few entries," but in the revised version all reference to the completed number of entries has been omitted. This change was made to reflect that the assessment was not designed to measure how many entries a student completed. The initial version also stated, "...fails to identify and evaluate own skills and work," a statement that basically focuses on a characteristic that is absent. The revised version reads, "self-evaluation skills are weak," which suggests that the characteristic, although weak, is still present in the student response.

Table 5.5 Revision of a Performance Level Descriptor
(from the C-TAP portfolio scoring scale)

LIMITED (INITIAL DESCRIPTOR)	LIMITED (REVISED DESCRIPTOR)
Completes few entries; shows little coverage of career-technical standards through understanding and application; fails to identify and evaluate own skills and work; writes poorly; presentation fails to enhance work	Shows little or no content knowledge and application of content knowledge and skills related to the career-technical standard(s); shows little or no ability to prepare for a career; self-evaluation skills are weak; fails to present work effectively

Like the development of assessment items and tasks, the development of scoring scales is an iterative process. It is also a process closely interwoven with assessment development. For example, during the development process, whenever an assessment is revised, the scoring scale may need to be revised. This is because a change to the assessment often results in a change in the student responses upon which the scoring scale is based. Therefore, if the student responses change, the scoring scale will likely need to be changed.

Scoring scales may also be revised during the assessment development process as a result of reviewing scored student responses. During the development of the C-TAP portfolio, for example, the initial scoring scale included three levels: Advanced, Proficient, and Basic. Teachers reviewing the scored portfolios, however, felt that the scoring scale should be changed because the range of student responses receiving a rating of "Basic" was too broad. In response to their comments, the assessment developers revised the scoring scale to include a fourth scoring category, "Limited." The new scoring level differentiated those responses that showed an almost complete lack of understanding (now designated as "Limited") from those responses that indicated the probability of achieving a "Proficient" rating if given minimal targeted support (designated as "Basic").

Scoring scale descriptors are usually not sufficient to communicate their meaning. To help communicate their meaning (and avoid having to write exceedingly lengthy scoring descriptors), scoring scales for performance assessments are usually accompanied by examples of student work, called benchmarks, that illustrate each level of performance. The benchmarks provide concrete examples of the scoring scale descriptors and offer an opportunity for others (e.g., teachers, students, parents, community

members, industry representatives) to understand more fully the meaning of each score or score level.

Developers should be aware that the first time a performance assessment is given, scores are typically low. In many instances, there are no performances that merit the highest rating of the scoring scale. Over time, however, as students and teachers become familiar with the assessment format, both instruction and student responses typically improve. For this reason, it is important to leave some room at the highest levels of the scoring scale for improvement. That is, if none of the responses are at the highest scoring level, than none of the responses should receive the highest score; or, if the highest score is assigned to some responses, it should be done so with the realization that the highest-scored responses in the early years of implementation will most likely look very different from the highest-scored responses in later years of implementation.

Checking for Validity

Scoring provides accurate information about students only if the assessment instruments and the scoring scales used to score the instruments accurately reflect the particular standard(s) being assessed. When such alignment is present, the scoring system is said to demonstrate validity. The alignment of both written on-demand and cumulative assessments to content standards was discussed in Chapters 1, 3, and 4. Checking for alignment, or validity, during the development process, helps ensure that the assessment items or tasks will elicit performances that provide scorable evidence relevant to specific standards.

There are two methods of checking for validity: 1) a review of the assessments and their scoring scales by content experts, and 2) a statistical review of student response data. Each is discussed below.

Review by Content Experts

One method of checking for validity is to ask content experts, such as classroom teachers, industry representatives, or higher education faculty, to review the assessments, their scoring scales, and benchmark performances (if available), and to then make a judgment as to whether they accurately reflect the standards targeted for assessment. At the district level, this review is usually conducted by a committee. At the classroom level, the review may

be conducted by one or more colleagues teaching in the same career or academic area.

Including a content review in the development process helps ensure the validity of an assessment. It is important, however, that the content review takes place more than once and at several points in the development process: just before the assessment items or tasks are tried out, when student responses are available after the tryout, and when the items are assembled into test forms for multiple-choice or mixed-format exams. Usually, it takes several cycles of trying out and revising an assessment to closely align the assessment and its scoring scales with the targeted standards. An assessment should not be used to make important decisions (e.g., graduation decisions) until this alignment is achieved.

A content review may include several activities. One is a mapping of the tasks or pieces of an assessment to identify where evidence can be found that reflects the targeted standards. Table 5.2 (shown previously) provides an example of such a map for the C-TAP portfolio entries and scoring dimensions. If a review of student work indicates a lack of scorable evidence related to the targeted standards, then a revision of the assessment task(s) may be needed. It is also possible that the task itself does not sufficiently reflect the standard, in which case a decision must be made to either drop the task or, if the standard is one of several being assessed by the task, to look to other tasks for evidence that reflects the particular standard.

In the development of performance assessments (e.g., portfolio, projects, short or long written-response items), content experts should review all assessment materials or prompts given to students, the assessment scoring scales, and scored student work to ensure that all the materials are consistent with the relevant content standards. For multiple-choice tests, the content experts should review the test forms to ensure that they are aligned with the relevant standards and the assessment blueprints. All content reviews should be performed by people with acknowledged expertise in the specific content or career-technical area. During the final phases of assessment development, an independent evaluation of validity should be conducted by reviewers who have not participated in the development of the assessment items, tasks, or scoring scales. Reviewers at this phase of the process might include career technical instructors, business and industry representatives, and, if specific disciplinary skills such as writing are assessed, teachers of academic content areas.

Another content review activity is that of ensuring that all students have equal opportunity to display their standards-related knowledge, skills, and abilities. The distributions of scores for items or tasks should be disaggregated

by such variables as gender, race/ethnicity, and level of English proficiency to identify particular differences in student performance. Items or tasks for which there are large differences in student responses should be reviewed by content experts to see if the content of the item or task is biased against a particular group. Such differences can help identify groups of students for whom assistance in specified areas (e.g., language mechanics) might be needed to enable the students to meet the standards.

Statistical Review of Student Response Data

Analyzing student response data for the purpose of determining the validity of a scoring system is a complex process, the full details of which are beyond the scope of this introductory document. There are, however, several key factors to consider when reviewing student response data for validity. Three of these factors — item discrimination for multiple-choice tests, item or task difficulty for all assessments, and distracter effectiveness for multiple-choice tests — are discussed below. For more detail and guidance and for other ways to analyze test results, teachers and district staff involved in test development should refer to the *Recommended Resources* in Appendix D and consult with experts in the field of assessment.

Considering Item Discrimination in Multiple-Choice Tests

One factor to consider when evaluating the validity of a multiple-choice test is *item discrimination*, or the "degree to which individual items *correctly* differentiate among test takers in the behavior the test is designed to measure" (Anastasi, 1982). When administering a multiple-choice test, it is expected that students' performances will differ from one another because their levels of mastery of targeted content are likely to vary. Some students have developed deep understanding of relevant concepts and principles while others are still attempting to grasp the meaning of the same ideas. For a multiple-choice test to effectively differentiate (or discriminate) among students, each question should reflect the tendency for test takers to differ, such that high-achieving students are more likely than low-achieving students to answer any particular item correctly. (Note: High-achieving students are those scoring well on the overall test or a similar measure of achievement. Low-achieving students are those scoring poorly on the overall test or a similar measure of achievement.)

A relatively simple way to determine whether an item is effectively differentiating among students is to calculate the difference between the

proportion of high-achieving students who answer an item correctly and the proportion of low-achieving students who get the same item right. The result of this calculation, known as an item discrimination index, will range between -1.0 and +1.0. If the index is positive, the item is discriminating well, meaning that, as expected, more high-achieving than low-achieving students answered the item correctly. A negative item discrimination index indicates that more low-achieving than high-achieving students got the right answer, a result which counters expectations. Items with negative item discrimination indices detract from the overall effectiveness of a test and should be examined carefully to determine the source of the problem (e.g., vague or imprecise language that leads to misinterpretation by students; lack of a clearly correct or best answer; lack of sufficient instruction or learning opportunities preceding testing). Any such items should be revised or replaced.

Considering the Difficulty of Items and Tasks

Another factor to consider when analyzing student response data is *item/task* difficulty, or the proportion (i.e., percentage) of all test takers who answer a particular item correctly or the percentage of students at each level of an assessment's scoring scale. For multiple-choice items, this proportion, sometimes referred to in research literature as the p-value, can range from 0.0 to 1.0 or 0% to 100%. Generally speaking, the higher the p-value, the easier the item. For example, an item with a p-value of .8 was answered correctly by 80 percent of the students who took the test. This item is considered easier than an item with a p-value of .2, which was answered correctly by only 20 percent of test takers.

Analyzing item difficulty is helpful in determining whether the level of difficulty of a particular assessment item is in line with the overall difficulty desired in the assessment. The level of difficulty desired in a test, and therefore in the items that make up a test, is closely related to the intended purpose of testing. For example, a test intended to identify the highest-achieving students should primarily contain difficult questions, or, in other words, questions with relatively low p-values (i.e., below 0.5 or 50%). An item found to have a p-value of .90 would probably be inappropriate for the test because such a question would not help distinguish the highest-achieving students from the others. It should therefore be replaced by a more difficult question or rewritten to be more difficult itself. Similarly, tests intended to identify low-achieving students (e.g., for remediation purposes) or to determine whether students have attained some minimum level of competency should contain mostly easy items with relatively high p-values

(e.g., above 0.7 or 70%). An item with a p-value of .20 may be too difficult for such a test and would likely detract from the test's ability to achieve its purpose.

Multiple-choice tests that aim to provide maximum information about the differences among test takers at all levels of mastery should include items with a range of p-values that average approximately 0.5-0.6. This will ensure that the test discriminates between a variety of performance levels and that it shows the full range of student achievement in the group. Items on such tests that are found to have p-values too far outside this range (e.g., .2 or .9) may be inappropriate and should be eliminated or revised.

Multiple-choice items with p-values very close to 0.0 or 1.0 are, in general, too difficult or too easy, respectively, for any test. Their ability to discriminate among students is very weak, making them minimally useful for most testing purposes.

It is also important to analyze or review the distribution of scores on the overall test or on the different levels of a scoring scale when considering whether an assessment is appropriately difficult. As with the item distributions, the general score distribution should reflect the purpose of the assessment. The distribution of scores for assessments measuring minimum competency should reflect most students scoring at the higher levels. Conversely, assessments meant to identify distinguished performances should have score distributions thatreflect most students scoring at the lower levels. Assessments aiming to distinguish achievement across levels should have score distributions that reflect the traditional bell-shaped curve, with most students scoring in the middle range. These "rules" for examining score distributions apply only to assessments taken by large numbers of students. At the classroom level, it may be more valuable to look at the scores against knowledge of the students and their past performance on similar work to get a sense of the appropriateness of the item or task difficulty.

Considering the Effectiveness of Distracters in Multiple-Choice tests

In addition to considering the proportion of students who answer a multiple-choice item correctly, it is also useful to examine the percentage of test takers who select each incorrect answer choice, or distracter. Distracters that are chosen by few or no students are likely to have been seen as implausible to test takers. Generally speaking, such distracters make items easier than they need to be because they reduce the number of plausible choices from which students

must choose. (Thus, instead of having a one in four chance of getting an answer right, for example, the student may have a one in three chance of correctly answering the item.) Ineffective distracters can weaken an item's ability to effectively discriminate among students and such distracters should, whenever possible, be replaced with more plausible incorrect answer choices.

Distracters selected by too many students may also be a problem. When substantially more test takers, especially high-achieving students, choose a particular distracter rather than the correct answer, it is possible that students are interpreting the item differently than intended or that there is more than one potentially correct answer to the item. If this is the case, the distracter should be clarified or replaced, this time with an answer choice that is more clearly incorrect yet still plausible. A distracter chosen by a high number of students could also be, however, an indication that a majority of students do not yet fully understand the aspect of content addressed by the item, pointing to a need for further instruction in that area.

Analysis of the validity of an assessment and its scoring system, whether by expert review or by statistical analysis, does not cease with the completion of the assessment development. Each assessment administration should be followed by an analysis of student response data and a subsequent refinement of the assessment items, tasks, or scoring scale(s) to help ensure that the assessment continues to be valid for each group or population of test takers.

Checking for Reliability

In addition to validity, assessment developers should check for the reliability of an assessment and its scoring system. Reliability is the degree of confidence that both scores and student performances are replicable over time and across different circumstances. Replicability of scores means that the same student response will receive the same score(s) no matter who scores it or when it is scored. For example, different scorers or the same scorer at a different time should assign the same score(s) to a given piece of student work. Replicability of performances also means that students will perform similarly on different tasks or items designed to measure the same standard at the same level of difficulty. If conventional standards for the replicability of both scores and performances are not met, then scores have little meaning.

For multiple-choice tests, statistical packages are available that calculate reliability measures for items on a particular test and for the test as a whole. Simply put, these indices compare different distributions of student scores

against each other to determine the degree of similarity. Developers should consult with district or external measurement experts to conduct and interpret these analyses.

Achieving reliability for performance assessments is a challenge because scoring such assessments requires interpretation and professional judgment. Rigorous methods of ensuring reliability when scoring performance assessments have been developed; the levels of reliability achieved for these assessments, though acceptable, are generally not as high as those achieved by multiple-choice tests. For this reason, high stakes assessments usually contain a mixture of multiple-choice items and performance tasks. In addition, when performance tasks such as writing samples are used in high stakes assessments, failing performances usually undergo an additional independent review to guard against inaccurate application of the scoring scale(s).

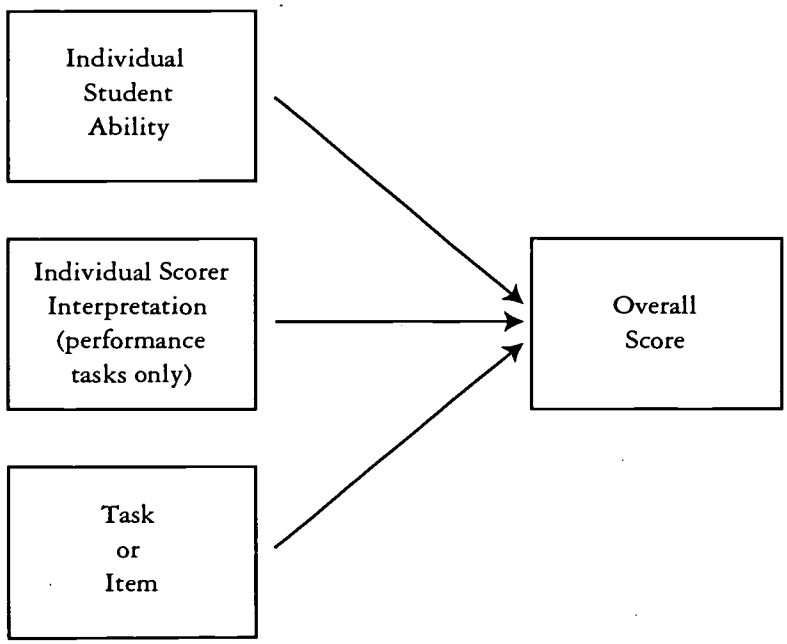
Once scoring scales are developed, their meaning must be accurately communicated to the scorers, so that scorers can both understand and internalize them. In large-scale, high stakes performance tasks, training sessions provide scorers with structured opportunities to become familiar with the scoring scales and their application to the particular tasks being scored. Understanding of the scoring scale is usually further facilitated by the systematic review of sample student responses designated as "benchmarks," which are accompanied by written and/or oral explanations of how each sample response reflects the relevant level on the scoring scale. As mentioned previously, these benchmark responses represent different points on the scoring scale. In training, scorers are taught how to apply the scoring scale to the benchmark responses. They then practice applying the scale on their own, receiving feedback from the trainer. Often, before actual scoring begins, the scorers in training participate in a process known as "calibration." In this process, the scorers are asked to score a previously scored set of student responses. If a scorer's judgment of a student response does not conform closely to the previous judgment, the scorer must receive additional training and participate in a second calibration. Scorers must achieve calibration before they are allowed to score independently.

During the scoring of high stakes performance assessments, reliability checks are built in by having at least a sample of responses scored by more than one scorer (i.e., doubled scored). The degree of agreement among scorers must be high enough to meet standards of reliability generally accepted in the field; if these standards are not met, the scorers must be retrained and scoring begun anew. In addition, individual scorers are often checked for "drift" from a correct understanding of the scoring scale. This is done by including unidentified previously scored student responses in their

workload. Once again, the individual scorer's judgements are checked against the previously assigned scores to ensure that they are applying the scoring scale consistently. If necessary, the individual scorer participates in more training and is "recalibrated" before being allowed to continue scoring. The scores that he or she previously gave to performances are also rechecked.

Large-scale assessments (whether multiple-choice tests, performance assessments, or a mixture) also check for generalizability of student scores, or how replicable student performance is across similar tasks. Generalizability studies statistically compare the effects of different factors (e.g., the student, the scorer, the task) on scores. This relationship between relevant factors and scores is portrayed in Figure 5.1. Ideally, an individual student's abilities should affect the scores more than an item or task, or a scorer. If individual scorers of performance assessments have a large impact upon scores, then it means that some scorers are consistently scoring more strictly than others. If the item or task has a large impact, then the performance on that item or task is unique and does not reliably predict performances on other items and tasks designed to measure the same set of knowledge and skills. Performance tasks tend to have relatively high task effects (i.e., student performance varies considerably as a function of the specific task.) This is why these tasks are not often used in large-scale assessments, or are used in combination with multiple-choice items, such as in the ACE tests.

Figure 5.1 Generalizability Model



Choosing a Cut Score to Reflect the Performance Standard

The final activity in developing an effective scoring system is the choice of a "cut score" to reflect the performance standard. As Chapter 1 discusses, performance standards describe how well students are expected to demonstrate knowledge, skills, and abilities with respect to content standards. For multiple-choice tests, the performance standard is represented by a chosen score, where all students scoring at or above the score are judged to meet the standard, while those scoring below the chosen score are not. For performance assessments, a single point or level of the scoring scale is chosen to reflect satisfactory mastery of the relevant standard(s). For C-TAP assessments, the cut score is set at the "Proficient" level. Students scoring at the "Proficient" and "Advanced" levels on the scoring scale are demonstrating satisfactory mastery of the content standards, while students scoring at the "Limited" or "Basic" levels are not.

The choice of the score or scoring scale level exemplifying the performance standard is based on the judgment of experts in the career or academic area. Classroom teachers developing their own assessments may confer with one or more colleagues; districts may form a committee composed of career-technical teachers, academic teachers, higher education faculty, and/or industry representatives. The task is to reach a consensus on the specific score or scoring scale level that best represents satisfactory student performance relative to one or more standards. There are two general approaches for accomplishing this task: 1) *test-centered approaches,* which involve reviewing the items or tasks that comprise an assessment and then deciding what level of performance on the items/tasks would be considered satisfactory, and 2) *examinee-centered approaches,* which involve using actual student responses or student performance data to determine the level of performance required to "pass" an assessment (Kane, 1994). Appendix D lists two resources that provide additional information on setting performance standards (see Kane, 1994 and Berk. 1986).

If, after administering an assessment, most students are meeting the performance standard, some consideration might be given to whether student learning can be improved by raising the performance standard. If many student performances are far from meeting the desired performance standard, both students and teachers will find it difficult to close the gap in a short time. In this case, consideration may be given to setting interim targets for student achievement that are raised over time to gradually

approach the desired performance standard as student performance improves. Of course, the setting of interim targets must be accompanied by a plan for raising student achievement. As mentioned previously, the first time an assessment, especially a performance assessment, is given, scores are typically low. Thus, developers should be careful about drawing premature conclusions based on assessment scores alone without a parallel analysis of the curriculum, instruction, and the familiarity of the assessment format to students and teachers.

Summary

Scoring systems are means of interpreting the relationship between standards (e.g., core academic, career preparation, career-technical) and student achievement. Depending on the nature of the assessment being scored, the scoring system can be relatively simple and straightforward or very complex. Scoring written on-demand assessments such as multiple-choice tests is relatively simple because the scorer has only to determine whether the student has or has not selected the correct response from a limited set of options. The scoring of performance assessments (e.g., projects, portfolios, short or long written-response items), however, is more complex because these assessments require a student to independently produce a response, with a wide range of answers possible. For these assessments, the scorer must interpret a student's response to determine its adequacy in relation to the appropriate standards.

All scoring systems should be developed at the same time as the assessments themselves. The process of developing a scoring system begins with the development of a scoring plan that explicitly states how the score(s) will be derived from the assessment (i.e., the scoring methodology to be used) and identifies which standards are to be reflected in particular scores.

For multiple-choice tests, the scoring methodology is built into the assessment format, although subscores can be generated from related sets of items. For performance assessments, there are two scoring methods commonly used (i.e., holistic and analytic). The holistic scoring method requires the scorer to evaluate the student response as a whole. It is best used when the overall impact of a student performance is of interest. The analytic scoring method requires scorers to identify important independent aspects of a student response and to score each aspect separately. It is best used when information regarding specific aspects of a student performance is desired.

Scorers using either the holistic scoring method or the analytic scoring method can also use one of two scoring approaches: the task-based scoring approach or the dimensional scoring approach. A scorer using the task-based scoring approach looks at the student's performance as a set of completed tasks, while a scorer using the dimensional scoring approach looks at the student's performance as reflecting a set of different dimensions or aspects that the assessment was designed to measure. If multiple scores are desired, the analytic scoring method is used. A scorer using the task-based scoring approach would give each task a score. A scorer using the dimensional scoring approach would give each dimension a score. The task-based scoring approach is best used when the assessment tasks are strongly and independently related to one standard or a set of related standards. The dimensional scoring approach is best used when the standards-related knowledge, skills, and abilities measured by the assessment are integrated within an assessment or occur across tasks.

After developing a scoring plan, the task of drafting the assessment scoring scales is usually undertaken. For performance assessments, scoring scales are typically used to communicate student achievement. These scales consist of descriptors of different levels of student performance. Scoring scales are most effective when accompanied by examples of student responses at each performance level, called benchmark responses.

The third and fourth steps in developing an effective scoring system are checks for validity and reliability. Validity is the degree to which the assessment, the student responses to the assessment, and the assessment's scoring scales are related to relevant standards. Checks for validity should be made at many points during the development of the scoring system and also at assessment administrations. Validity can be checked through multiple reviews by content experts and through statistical analyses of scores.

Reliability is the degree to which both scores and student performances are replicable over time and across different circumstances. Checks for reliability are typically done through statistical analyses of assessment scores. For performance assessments, reliability is also checked by comparing the scores of one scorer with the scores previously assigned by other scorers. Achieving reliability in scoring requires extensive training of scorers and monitoring of their scoring. Checking for reliability and validity of scoring systems is essential to ensuring that a score accurately reflects the extent of a student's standards-related knowledge and skills (and not other factors such as a scorer's training or understanding of the scoring scale or poorly designed assessment materials.)

The final step in developing an effective scoring system is choosing a cut score that reflects the performance standard for student mastery of designated content standards. The cut score identifies a dividing line where students scoring at or above the line are considered to have adequately mastered the content standards while those scoring below the line have not. If most students taking an assessment meet the performance standard, then the standard might be raised. If, however, most students taking an assessment do not meet the performance standard, consideration should be given, not to lowering the standard, but to setting interim targets that are gradually raised over time to approach the desired performance standard as student performance improves.

Reporting Student Achievement

After an assessment of student performance is scored, the next major step is to communicate the results to the appropriate audience(s). This chapter describes some of the purposes for communicating results of student achievement and some of the formats for reporting that achievement. The chapter also describes the characteristics of an effective reporting system, and provides some options for reporting information based on a combination of assessment measures.

Purposes for Reporting Student Achievement

Purposes for reporting student achievement are often directly linked to the type of information provided by the student assessment. On a basic level, information provided by student assessments is either primarily formative or primarily summative. Assessment information that is primarily formative is diagnostic in nature — that is, it identifies specific student strengths and weaknesses that have implications for instruction and for a student's work in progress. Formative information is usually used by the teacher for several purposes: 1) to identify gaps in both individual and whole group understandings that need to be addressed through instruction; 2) to identify elements in students' work-in-progress that need improvement; or 3) to help specify revisions in a student's work that would move the student's performance or product to the next level of achievement.

Assessment information that is primarily *summative* summarizes a student's achievement at a specific point in time. Summative information is usually used for several purposes: 1) to compare student performances to established standards; 2) to compare student performances against those of peers; 3) to

assess a student's performance in relation to his or her personal goals; and 4) to help decide what happens next in a student's school career. For cohorts of students or different student populations, summative patterns of achievement across students can be used to suggest revisions in the curriculum, to identify group differences in achievement that suggest different support strategies, and to evaluate program strengths and weaknesses.

Both formative and summative information about student achievement can be reported to individual students and their parents and to the larger public. Such information can be used by students and their parents to understand how well the student is progressing and what areas might need improvement. The larger public (e.g., administrators, community members) is more likely to use summative information to help make policy and administrative changes, instructional program changes, changes in staff development, and improvements in local assessment systems.

Different Reporting Formats

Both formative and summative assessment information can be reported in a variety of ways. Teachers can communicate such information to students by having individual conversations with students, providing written comments on student work, assigning points or letter grades, or using a combination of these methods. There are fewer options, however, for communicating such information to parents, administrators, and the wider community. Teachers may have individual conferences with parents, but other than the use of report cards and average scores on standardized tests, there is often no formal means for teachers or the district to communicate student achievement to those outside the classroom.

When a district adopts a new form of assessment, it is the perfect time to consider establishing a meaningful system to report assessment results. Such a system can include a variety of formats, including letter grades, numeric scores (such as percentages, percentiles, or scaled scores), developmental continua (rubrics or checklists), narratives, portfolios, and/or student-led and three-way (i.e., student, parent, teacher) conferences. Each of these various reporting formats is discussed below.

Letter grades are what most parents are accustomed to seeing on homework and report cards. Letter grades indicate roughly how well students are doing relative to teacher expectations. Rarely, however, do the grades provide specific information about how students are doing or the teacher's expectations for

student performance. Teachers may specify in advance the components that constitute a grade (e.g., homework counts for W % of the grade, quizzes for X %, major tests for Y %, and the final exam for Z %), but the actual knowledge, skills, and abilities being assessed are usually unspecified. Moreover, teachers sometimes incorporate into their grading schemes general workplace readiness or personal responsibility skills (e.g., handing in work on time, punctuality, or participating in class discussions) which further complicates the interpretation of a letter grade or percentage score.

Another format for reporting student achievement, one which is commonly employed by most standardized multiple-choice tests, is the use of *numeric scores* that summarize student achievement. Types of numeric scores frequently used to report student achievement include raw scores, percentage scores, percentiles, or scaled scores. A *raw score* is how many test items a student answered correctly. It is only informative, however, if it is reported along with the total number of items in the assessment. A *percentage score* indicates the percentage of items that a student answered correctly. It can be more informative than a raw score, especially if the total number of items is not included as part of the reporting. For example, reporting a student score of 5% conveys more information than reporting a raw score of 5. The value of percentage scores, however, is also dependent on the total number of items in an assessment. For example, if there are only three items in an assessment, then the possible percentage scores are 0, 33, 67, and 100. Reporting a score of 67% for this assessment is less informative than reporting a score of 67% on an assessment containing 100 items.

Percentiles, another type of reporting format, report a student's performance relative to other test-takers. For example, a student scoring at the 90th percentile did as well as or better than 90% of the students who participated in norming the test. Similar to percentiles are scaled scores, such as those used by the SAT. Scaled scores take into account the mean and standard distribution of the student population norming the test. In addition, the range of predetermined scales can vary widely. The SAT, for example, uses a scale ranging from 200 to 800.

Developmental continua, in the form of checklists or rubrics (i.e., scoring scales), can offer more specific information about student achievement than letter grades or numeric scores. A developmental continuum is a sequenced list of skills that represent increasing progress toward mastery of a content area. Some developmental continua, such as the *checklist* in Table 6.1, describe specific knowledge, skills, and abilities to be mastered by the student. At the elementary school level, for example, checklists often include specific skills such as "Counts to 10," or "Knows all letters of the alphabet."

Checklists can also be more complex or contain general categories rather than specific skills. For example, a checklist might include the general category, "Problem Solving," and instead of listing skills, it could include a scale that describes progress at several levels, such as "Not Yet Progressing," "Working on Progress," and "Progressing Satisfactorily."

Table 6.1 **Example of a Developmental Continuum (Checklist) for Work Ethic Standards**

	Exceeds Standards	Meets Standards	Below Standards
Makes decisions quickly after due time is given to fact-finding and consideration of the alternatives.			
When necessary, disagrees and debates with others in a professional, respectful manner and always uses positive methods of persuasion.			

From Academy High School Internship Preparation Program, cited in Bailey and McTighe, 1996, p. 123.

Rubrics (also referred to as scoring scales), like some checklists, include levels of progress, but each level is accompanied by a detailed description and, sometimes, by samples of student performances (see Table 6.2 for an example). The addition of descriptions and samples of student performances gives context and meaning to the different levels of progress, each of which is sometimes also assigned a numeric rating. The ACE/C-TAP assessment system uses rubrics to assess the written scenario, project, and portfolio assessments, and to report student achievement on the assessments. Both rubrics and checklists offer a more complex picture of student performance than do letter grades and numeric scores, and, if the number of performance levels is limited, are relatively easy to understand.

Narratives as a format for reporting student achievement are distinctly different from letter grades, number scores, and developmental continua. Narratives, or narrative reports, provide descriptions of individual students' achievement. Table 6.3 shows an example of a narrative report related to one student's achievement. Narratives have the potential to offer customized information about an individual student's strengths and weaknesses relative

Table 6.2 Example of Developmental Continuum (Rubric) for Economics Assessment

This rubric describes student achievement in economics at the high school level. The rubric is based on an assessment composed of multiple-choice items plus a written essay.

6 Student work at this level shows excellent understanding of economic concepts and principles, which are applied and extended; the work:
- goes well beyond assigned tasks
- includes sophisticated and insightful analysis
- demonstrates excellent communication and writing skills

5 Student work at this level shows a solid understanding of economic concepts and principles, and reflects awareness of alternatives; the work:
- completes all major assigned tasks
- includes well-developed analysis
- demonstrates strong communication and writing skills

4 Student work at this level shows a good understanding of economic concepts and principles; the work:
- covers all major tasks
- argues convincingly
- demonstrates good communication and writing skills

3 Student work at this level shows a basic understanding of economic concepts and principles; the work:
- focuses on several aspects of assigned tasks
- displays adequate communication and writing skills

2 Student work at this level shows a limited understanding of economic concepts and principles; the work:
- addresses a portion of assigned tasks
- demonstrates basic communication and writing skills

1 Student work at this level shows little or no understanding of economic concepts and principles; the work:
- only briefly mentions economics
- provides limited expression of ideas without much focus

From Golden State Examination in Economics School Report Form, 1996.

to desired elements of achievement. Informative narratives can be difficult to write, however, because a good narrative depends partly on the quality of a teacher's writing and analytical skills. The teacher must decide exactly what

to write for each student. It is often difficult to strike a balance between providing too much information (which makes the narrative time-consuming to write and read) and too little information (which greatly reduces the value of the narrative format). If, when using a narrative format, a teacher finds himself or herself repeating phrases to describe similar achievements of different individual students, then the rubric format may be equally or more effective and require less work. Individual narratives are also very difficult and time-consuming to aggregate, making their usefulness for larger groups of students extremely limited.

Table 6.3 **Narrative Report of a Student's Final Project, Work, and Study Habits**

Jennifer is an elaborate thinker who eagerly accepts cognitively challenging activities. She thinks in novel ways and approaches problems with clever, unusual solutions. She has the ability to embellish and expand ordinary ideas into unique ones. Her scientific investigation about bone calcification was just one example of her ability to follow proper testing procedures, collect and record data, analyze information, and most of all, draw logical conclusions. Her identification and assessment of variables which may have affected her experiment's results required high-level thinking skills. What a quality project! Jennifer is a positive, effective leader among her peers. She is a goal-setter who determines priorities and meets deadlines. She demonstrates good time management and organizational skills. Jennifer has the courage to take risks, defend her opinions, and dream....

From Peckron, 1996, p. 58.

The *portfolio* as a reporting format can provide a rich picture of student achievement. Chapter 4 discusses the use of portfolios as formative assessment tools that make almost seamless connections between instruction and assessment. As mentioned in Chapter 4, portfolios can also be used for summative assessment and for reporting the results of that assessment. Depending on its contents, a portfolio may have the ability to reveal student progress over time and/or represent different achievements by the student. (C-TAP portfolios focus on the latter.) To be effective, however, portfolios must be well-organized around a moderately prescriptive structure or framework. If they are only a hodgepodge of collected materials, they can be almost impossible to interpret (even by skilled readers) and hence there will be little information to report. Furthermore, to assess the achievement of groups of students, the performances captured by a portfolio assessment must be summarized and reported using such formats as narratives, letter

grades, or rubrics. The C-TAP portfolios, for example, use a rubric to summarize and report a student's performance. The rubric not only gives the student a global picture of his or her achievement, but also can be used to aggregate performances across students to evaluate instruction and the program. In addition, rubrics allow for the examination of aggregate achievement by special populations of students to determine the suitability of the curriculum for different types of students.

Finally, in the *conferencing* format for reporting student achievement, the teacher meets in person with one or more people (e.g., student, parents, guardians) and orally conveys information about student achievement. During such conferences, the teacher almost always uses one or more of the above-described reporting formats as a basis for the conference. Depending on the format used, the conference enables the teacher to interpret and expand upon the format, and to answer any questions the listener(s) may have about the format and/or student achievement.

Characteristics of Effective Reporting of Student Achievement

Once a district determines the format or formats to be used for reporting student achievement, it is important that those receiving the information (e.g., students, parents, teachers, school and district administrators, school board members, admissions staff for post-secondary institutions, employers, the general public) understand the assessment results and be able to interpret them accurately in order to take or support appropriate actions. Based on Stiggins' five prerequisites for the effective reporting of student achievement (1994), Table 6.4 summarizes the five characteristics of effective reporting of student achievement. The characteristics are discussed in more detail below.

The first two characteristics of effective reporting of student achievement were discussed in Chapters 1 through 4. These chapters point out that linking assessments to particular standards helps make clear the aspects of student achievement being addressed. What also needs to be clear, however, is that different assessments may focus on the same or similar standards, but produce different results with regard to the breadth or depth of the knowledge and skills being measured. On-demand assessments typically are better at assessing breadth of knowledge, while cumulative assessments can better assess depth of knowledge. In an Animal Science program, for example, an on-demand assessment, such as a multiple-choice test, might be used to assess the breadth of a student's knowledge about various species of

Table 6.4 Five Characteristics of Effective Reporting of Student Achievement

- The aspects of reported student achievement are clearly and appropriately defined.
- The information being reported about student achievement is of high quality.
- The reason(s) for reporting student achievement (i.e., how the information is to be used, by whom, for what purpose) is clearly understood by both those reporting the information and those receiving the information.
- The meanings of the words, pictures, scores, or other symbols used in reporting are clearly understood by all participants.
- All reports of student achievement include a provision of a time, place, and set of circumstances when the reporter of student achievement and the receiver(s) of the information can attend to the information being shared.

animals, while a cumulative assessment, such as a project, might be used to assess a student's depth of knowledge about one species in particular. Stakeholders need to understand the different aspects of student performance that on-demand and cumulative assessments can measure. One assessment rewards quick thinking and reasoning, while the other rewards planning, persistence, and revision over time. Basic knowledge of assessment capabilities and differences is especially important when interpreting the results of multiple-measures assessment systems (e.g., ACE/C-TAP) which provide a more complete picture of student performance. Multiple-measures assessment systems demand that stakeholders understand the complexities of interpreting the results across multiple types of assessments.

The third characteristic of effective reporting of student achievement implies that different reasons for reporting student achievement might make some reporting formats more appropriate for some stakeholder groups than others. Examples of stakeholder groups that have different needs and uses for information about student achievement are parents, the general public, and post-secondary institutions. Parents need information about their individual children and are the group that is most likely to be willing to invest the time and effort required to examine and understand more lengthy reporting formats such as portfolios and narrative reports. In contrast, boards of education and the general public will most likely be interested in summaries of group achievement which are easily interpreted (e.g., percentiles or rubric levels.) Post-secondary educational institutions require information for admission

purposes and continue to request standardized test scores and students' grade point averages as key admission criteria (which is one reason that most high schools have not embraced alternative reporting formats.) To ensure that reports of student achievement are useful to those receiving them, districts must identify the various purposes of assessment information, together with ways of gathering and reporting information that fit these purposes. A district using a multiple-measures assessment system increases the likelihood that each of its stakeholder groups will find one or more assessment measures that provide information directly relevant to the group's concerns.

The reports of student achievement not only have to meet the varied needs of specific stakeholder groups, but the fourth characteristic of effective reporting states that the reports must be readily understandable. Any information needed to interpret a word, score, or symbol should be provided as part of the report. There should be no need to consult another document to interpret a student achievement report.

Some types of reporting formats require more interpretation than others. Two examples are norm-referenced scores (e.g., percentiles) and criterion-referenced scores (e.g., performance levels in rubrics). *Norm-referenced scores* portray how well a student did in comparison to other students. *Criterion-referenced scores* tell how well a student did in comparison to a set of performance standards. For example, a high norm-referenced score means that the student is doing well compared to other students, but it does not necessarily mean that the student is doing well with respect to the performance standards; indeed, all students may be performing inadequately. In contrast, a high criterion-referenced score means that the student is doing well with respect to the performance standards.

It is important to note that when an assessment system is initially implemented to measure standards-related skills and abilities that have not been explicitly taught in the past, it is likely that smaller percentages of students than usual may attain the highest levels of scores. In this case, the system should be designed to allow for improvement in student work as teachers and students better understand the standards and how to improve work in relation to these standards. This shift in instructional emphasis needs to be communicated when scores and score distributions are released or else the efforts of students, schools, and districts will be judged too harshly.

Quantitative scores can be subject to misinterpretation. The various types of quantitative scores differ greatly in the information that they convey. For example, many scaled scores, such as those reported for high-

stakes multiple-choice tests such as the SAT, are calculated so that scores are not only ordered, but are also proportional. This type of scoring scale is called an *equal-interval scale.* For example, if one student's equal-interval scaled score is twice as high as another's, then the first student did twice as well. Unless well-informed, some people make the same assumption for all other quantitative scores. However, most quantitative scores are reported based on an *ordinal scale.* These scales only indicate the relative order of performances on a scale portraying better performances at one end and worse performances at the other. A rubric is an example of an ordinal scale. Although a "4" on a rubric is a higher level performance than a "2," it is not twice as good. Another common misinterpretation is to convert rubric levels to percentages. A "2" on a four-point rubric scale is not equivalent to a percentage score of 50. In fact, the difference between two performances receiving a high "3" and a low "4" respectively may be less than between two performances receiving a high "3" and a low "3." Each subsequent rubric level represents a threshold that performance must cross in order to receive the higher rating, but there can be — and often is — a lot of variance within a particular level. To reduce the possibility of confusion, rubric levels are often represented with labels such as "Proficient" rather than with numbers. To reduce confusion, districts must educate stakeholder groups as to valid and invalid conclusions to be drawn from reports of student achievement.

The fifth characteristic of effective reporting of student achievement is the provision of opportunities to discuss the reports in order to explain and clarify their meaning. Schools have traditionally done this through parent-teacher conferences as well as through schoolwide and school board meetings. When a new method of assessing or reporting student achievement is begun, these provisions for discussion are particularly important to reduce the types of confusion described previously. Assessment results can only affect instruction, curriculum, policies, and practices if they are examined, understood, and analyzed for any implications for subsequent action.

Since schools are publicly funded institutions, the opinions of stakeholders matter a great deal. By providing early opportunities to ask questions and express concerns about new assessment reports and instruments, teachers, schools, and districts can solicit and address concerns before they turn into controversies.

Combining Multiple Assessment Measures for District Reporting

Since more and more districts are using a variety of assessments (e.g., classroom tests, district tests, standardized assessments) to gauge student achievement and progress, it has become a challenge for districts to combine these multiple measures into one general evaluation judgment for reporting purposes. This challenge is especially acute for those districts that are implementing standards and using multiple assessments to determine whether students meet those standards.

Some districts meet this challenge by developing new report cards that reflect adopted standards and the multiple measures of assessment. One such district was profiled in *Educational Leadership* (Kenney and Perry, 1994). This Colorado district developed its own standards and a generic rubric for evaluating classroom work. When it came time to report student progress, teachers realized that their usual grading scheme did not match with their new standards-based instruction. They subsequently obtained approval from their School Board to convene a group of teachers and parents to work on designing a new report card.

The report card they developed addresses 38 core standards across several content areas (e.g., science literacy, arts and humanities, mathematics) and performance dimensions (e.g., works collaboratively, produces quality work). Every standard is assessed at least one time each year, but the reporting of results differs in frequency. Some standards are assessed and results reported every quarter; other standards are assessed throughout the year, but the results are only reported once at the end of the year. Instead of receiving a single grade in each subject area, students get several scores, an approach that has been recommended by many assessment reformers as an element of improved reporting of student performance and learning. The Colorado district uses a variety of measures to assess student performance, including project assessments which are scored by a teacher and two people from outside the classroom (e.g., an administrator, a parent trained in scoring, an industry representative).

To help inform parents about the new system for reporting student achievement, the district prepared a brochure explaining the differences between the old and new report cards and sent it home at the first reporting period. With the brochure, they enclosed a survey that asked for parents' opinions and questions about the new system. Parents also received information about the new reporting system from teachers during parent conferences. These efforts helped smooth the transition to the new reporting system.

While the Colorado school district reported several scores for a single content area, some stakeholders (e.g., school board members) around the country have expressed interest in seeing information from multiple measures of student achievement combined (e.g., into one score or description of performance) to summarize achievement in a content area. A look at suggested local assessment and accountability practices in California provides some ideas for ways to combine the results of multiple measures to determine whether students meet desired standards of achievement. In California, schools and districts are encouraged (but not required) to collect and organize local accountability data with multiple measures that are aligned with grade-level performance standards. The California Department of Education has developed several models for combining results from multiple measures that schools and districts can use if desired. The models represent approaches for combining multiple measures and for setting grade-level performance standards when results from one, two, or three measures of achievement are available for students at a grade level in a specific subject (California Department of Education, 1998).

In general, the approaches developed by the California Department of Education involve the creation of two- and three-dimensional matrices that shows different levels of performance within different measures of achievement. Table 6.5 below provides an example of such a matrix, in this case, a two-dimensional matrix that could be used to combine performance data for two measures of achievement (i.e., a norm-referenced test and a class grade), each with multiple levels of performance (i.e., six levels of performance for the norm-referenced test and five levels of performance for the class grade).

Table 6.5 **Example of a Two-Dimensional Matrix for Combining Multiple Measures to Establish Grade-Level Standards**

Score on
NORM-REFERENCED TEST*

Class Grade	1 – 29	30 – 39	40 – 49	50 – 59	60 – 69	70+
A						
B						
C						
D						
F						

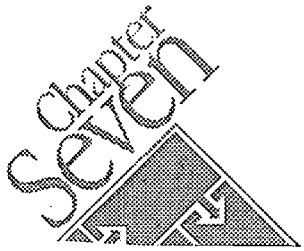
* All norm-referenced test scores are stated in terms of national percentile ranks.

Using a matrix like the one presented in Table 6.5, key stakeholders in a community (e.g., teachers, administrators, school board members, parents) can "look simultaneously at the different levels of performance on each of the two measures of achievement, and decide whether each possible combination of results (represented by each cell in the matrix) meets or does not meet desired grade-level standards. Deciding which combinations of results meet grade-level standards can be accomplished using a consensus process and by looking at students' work and/or assessment items for each combination. As stakeholders establish grade-level standards, they will need to consider the meaning of each standard, or in other words, what students who meet the standard are actually able to do and how well they can do it. They can then translate that understanding into a 'map,' or a line, separating the combinations of scores that meet the grade-level standards from those that do not meet the standards." (California Department of Education, 1998)

Appendix B briefly outlines California's models for combining multiple measures of achievement, presenting one or more examples of each. (NOTE: Schools and districts interested in combining results from multiple measures of student achievement will need to do further research into the different methods available to them.)

Summary

Reporting student achievement is a critical component of any assessment system. The assessment reports and the process that produces them should be made meaningful to important stakeholders such as parents, teachers, administrators, school board members, and the general public. A meaningful reporting system can be established by including a variety of reporting formats that meet the variety of purposes an assessment system will serve; by ensuring that all participants understand the achievement being reported and the methods of assessment and reporting being used; and by allowing for districts to report student achievement on multiple measures under one evaluation umbrella. Most stakeholders may not take or have the time to achieve the depth of understanding needed to fully interpret assessment reports. However, if a reporting system is meaningful and sound, the pieces will be in place to satisfy those with questions or concerns and to ensure that the information is accurate and able to contribute to appropriate decisions.



Supporting a Standards-Based Assessment System

Each of the previous chapters focused on a specific step in the development of a multiple-measures, standards-based assessment system: 1) identifying standards, 2) understanding key characteristics of an effective assessment system, 3) developing written on-demand assessments, 4) developing cumulative assessments, 5) developing a scoring system, and 6) reporting student achievement. This chapter introduces several steps that schools and districts can take to support the overall development and implementation of such a system. Among the steps discussed are the following:

- strengthening organizational support for change;
- developing and/or implementing a new assessment system in phases;
- meeting the needs of all students;
- establishing community-wide support; and
- coordinating local and state assessment efforts.

Many of the ideas presented in this chapter are based on the development and implementation experiences of ACE/C-TAP. Though no district has yet implemented the ACE/C-TAP system district-wide, the experiences of individual teachers and schools using ACE/C-TAP assessments provide important insights about the support that school and district staff will need as they develop and implement a multiple-measures, standards-based assessment system.

Strengthening Organizational Support for Change

A district or school will be able to develop and implement a multiple-measures, standards-based assessment system more quickly and smoothly if organizational conditions supportive of change are in place. Table 7.1 summarizes some key steps that schools and districts can take to strengthen organizational support for change. Each step is then discussed in more detail. Ideally, a school or district will consider these steps early in the development process.

Table 7.1 Steps for Strengthening Organizational Support for Change

- Committing time and resources to system development
- Establishing a strong leadership team
- Encouraging collaboration among staff
- Providing relevant professional development opportunities
- Facilitating change without jeopardizing student learning

Committing Time and Resources to System Development

The process of planning, trying out, and refining the various components and practices that make up a multiple-measures, standards-based assessment system is complex and lengthy. Expecting a new or reorganized system to be fully institutionalized in one or two years is unrealistic; five to seven years is more likely. Therefore, a school or district, and its wider community, must be willing to commit both time and resources to development and change over a period of years. This commitment begins with the development of standards, assessments, and scoring procedures, and extends through efforts to improve curriculum, instruction, and assessment over time so that all students receive the support needed to achieve targeted standards.

Establishing a Strong Leadership Team

As mentioned in previous chapters, new assessments and assessment systems must be tried out and modified repeatedly until they reliably measure student performance in relation to targeted knowledge and skills.

Complex and innovative change efforts such as these require strong leaders willing and able to shepherd changes through numerous stages of modification and refinement (Fullan with Stiegelbauer, 1991). Without such leaders, development efforts can stall before their full benefits are achieved.

When initiating assessment development efforts, a school or district should establish and support a leadership team responsible for keeping the development process moving forward. While a variety of individuals can participate on this team, it is especially critical that both teachers and administrators assume active leadership roles. For their part, teachers can continually gauge the impact of assessment efforts at the classroom level. As they observe students working on assessments and review the students' finished work, teachers can help determine whether assessments and related changes in instruction are positively influencing student learning. They can also discern gaps in students' knowledge and skills that indicate the need for additional support and guidance. Based on their observations and previous assessment experiences, teachers can help identify ways to continually improve assessments and instruction over time. In addition, they can help others (e.g., students, parents, teachers not yet using the new assessments and instructional practices) better understand the purposes and processes of the new assessment system.

The inclusion of administrators on the leadership team lends legitimacy to the development effort. Administrators can help keep the wider community informed about changes in assessment and instructional practices, and facilitate the commitment of resources to support development efforts. As the assessment system and its effects on student learning are monitored, committed administrators can often provide a more global perspective than can classroom teachers.

All members of the leadership team should be familiar with at least some of the national, state, and local standards and assessment development efforts taking place in related fields. The experiences of, and lessons learned by, others can serve as invaluable resources to the team as it devises strategies for guiding development efforts toward completion.

Encouraging Collaboration Among Staff

A primary purpose for developing or reorganizing a standards-based assessment system within a district is to ensure that school programs and departments teach toward the same set of standards, and assess student learning vis-à-vis those standards. Achieving this purpose requires an

organizational environment that encourages collaboration among staff in an effort to embrace change and meet shared goals.

There are several general types of collaboration that districts should encourage. First, in cases where standards have not yet been established or where existing standards are being revised, a district will want to provide adequate time for staff and the wider community to discuss the standards and reach consensus about the wording and meaning of new expectations for learning. An environment that supports collaboration is essential during such discussions.

Second, once standards are in place, teachers within schools should be encouraged to work together to ensure that the knowledge and skills covered in their individual classes contribute to combined coverage of all standards across the curriculum. Teachers must also explore the various ways in which their different courses complement each other and how they can work together to help students integrate and apply knowledge as they learn and demonstrate targeted knowledge and skills. The vignettes below provide several examples of specific ways in which teachers have collaborated when implementing the C-TAP portfolio and project assessments.

- Collaboration on portfolios: One agriculture teacher asked an English teacher to help students "polish" their work sample summaries, writing samples, resumes, and letters of introduction for their portfolios. Over time, three additional English teachers became involved in the collaborative effort. (Appendix E provides additional examples of how teachers have collaborated when planning and implementing the C-TAP portfolio assessment.)

- Collaboration on projects: One technology core teacher met with a mathematics teacher to discuss how both math and technology core knowledge and skills could be demonstrated in a project assessment on redesigning the Titanic to be seaworthy. The same teachers also discussed how knowledge of math concepts such as volume and surface area could be applied as a technology core student constructed a paper model of a robot using a process similar to that used to lay out sheet metal for construction of an actual robot.

Third, within a district, staff from secondary schools should be encouraged to collaborate with feeder schools to ensure that students are guided through an instructionally sound sequence of standards-based content that will adequately prepare them for secondary school learning experiences. Feeder school staff should also be encouraged to introduce students to a

variety of assessment experiences, including cumulative assessments like projects and portfolios. These early experiences will help students begin building the skills needed (e.g., identifying topics and key requirements in written-response questions, collecting and refining work over time, evaluating one's own work against known standards, showcasing work in ways that effectively demonstrate achievement) to complete such assessments in secondary school.

Finally, teachers and administrators should be encouraged to work together to strengthen a newly developed assessment system over time. The ACE/C-TAP experience indicates that development and implementation can be facilitated by ongoing collaborative conversations in which teachers compare experiences and share ideas as they try out new assessments and instructional strategies. These opportunities help teachers to deepen their understanding of key assessment processes, to determine the impact of new assessments and teaching strategies, and to identify ways to further improve instruction and assessments.

To engage in the types of collaboration described above, school and district staff need time to think and plan together, both within and across traditional departmental lines. This kind of time cannot easily be squeezed into the regular workload of teachers or administrators. Therefore, schools and districts must be willing to set aside specific time dedicated to assessment-related collaborative activities. For example, some schools using C-TAP provide such time in their staff and departmental meetings; others use time set aside for school-wide professional development activities. In this way schools and districts help institutionalize the practice of collaboration.

In addition, districts lacking a tradition of collaborative working relationships may need to engage in team-building efforts and professional development aimed at strengthening teachers' collaboration skills. The same will be true for districts or schools with a history of adversarial relationships between teachers and administrators or between different departments.

Providing Relevant Professional Development Opportunities

Any school or district developing and/or implementing an assessment system must be ready to help teachers master the knowledge and skills required to create and use the system effectively. As mentioned above, teachers may need support in learning to collaborate during development and implementation efforts. C-TAP experience suggests that schools or

districts may also need to provide professional development opportunities focusing on the following areas:

- understanding the standards-based knowledge and skills expected of students; and
- expanding assessment literacy.

Each of these professional development needs is discussed in more detail below, along with a description of selected resources which may be used to support professional development efforts. Large districts can usually support such professional development efforts alone. Smaller districts may need to collaborate with other districts or an institution of higher education that is interested in developing parallel skills in its own staff.

Understanding the Standards-Based Knowledge and Skills Expected of Students

When a new set of standards is developed or adopted within a district, school, or program, teachers must be given structured opportunities to become familiar with the standards and to develop an understanding of the specific knowledge and skills that students are expected to master. Developing such an understanding is essential if teachers are to effectively articulate the expectations to students, parents, and others. It is also critical if teachers are to help ensure that both curriculum and instruction provide students with opportunities to learn the specific knowledge and skills needed to demonstrate mastery of standards.

Often new or refined standards may emphasize some knowledge and skills not well covered by existing instructional practices or curriculum. When teachers notice such gaps between standards, curriculum, and instruction, they can make changes in their programs accordingly. Doing so will likely require professional development aimed at helping teachers develop new instructional techniques and new ways of thinking about student achievement.

Expanding Assessment Literacy

The type of assessment system described throughout this document requires teachers to be more involved in designing, administering, scoring, and interpreting assessments than in the past. To fulfill their crucial role in such a system, it is likely that teachers will need to expand their "assessment literacy."

Ideally, all staff, including administrators and teachers, should participate in an introductory workshop designed to familiarize participants with the benefits of a multiple-measures, standards-based assessment system, and with the general characteristics of the different types of assessments that make up the assessment system in their school or district.

An introductory workshop on assessment-related issues, however, is not sufficient. Training sessions geared toward specific aspects of development and implementation are also necessary. For example, teachers need to learn how to effectively plan for and implement a variety of different types of assessments (e.g., projects, portfolios, written on-demand assessments) in their classrooms. Of particular importance is helping teachers learn to effectively support students through assessment experiences, especially complex cumulative assessments such as projects and portfolios. The ACE/C-TAP experience has shown that teachers benefit from professional development resources (e.g., literature, workshops, collaborative conversations) that help them effectively "coach" students to do the following: identify and understand the requirements of on-demand assessment items; plan and organize long-term work; reflect on and evaluate their work vis-à-vis standards; and refine and improve various types of work (e.g., writing, hands-on work products) in light of constructive feedback.

In addition to learning how to effectively implement new types of assessments, teachers and other staff will likely need help to reliably evaluate students' responses to such assessments. While teachers have always designed and conducted assessments for classroom purposes, they typically judge the quality of student work based on their own criteria, criteria which they have not always applied consistently across students within their classes. In addition, different teachers covering similar content in different courses have not always used the same standards or criteria when judging students' achievement. Some have reputations as "hard graders," while others are known as "easy graders." A standards-based assessment system requires teachers and others responsible for scoring to use a common set of criteria when evaluating completed assessments. This means that teachers may need some assistance in learning to use common, agreed-upon performance standards, scoring scales (i.e., rubrics) and benchmark performances to make consistent judgments about the quality of student work both within and across classes.

Finally, teachers may also need support in learning to use assessment results to diagnose the learning needs of individual students or groups of students and to plan future instruction accordingly. This issue is discussed in more detail later in this chapter.

Resources for Professional Development

There are numerous resources that schools and districts can draw upon to support the professional development needs described earlier. Several of these resources are described below.

- Written or visual materials (e.g., books, journal articles, educational videos) can help support the development of specific skills in students and teachers, and also provide new ideas for working through a variety of issues related to the development and implementation of a new assessment system. A partial list of such resources can be found in Appendix D. Additional documents are available from a host of professional organizations, regional education laboratories, private educational consulting firms, and individual consultants.

- Professional associations and statewide or national conferences related to specific content or career areas can help school and district staff to clarify their understanding of specific standards and provide ideas for linking instruction and assessment to those standards.

- Internal or external consultants (e.g., from research-oriented organizations) not directly involved in the development process can contribute to objective monitoring and help a district keep the "big picture" in focus when implementing a new assessment system over time. Such consultants can provide expertise and can also promote reflection and conversation about critical pedagogical and technical issues that might help schools and districts build their own capacity to improve.

- Teachers and administrators who have already implemented (or who are farther along in developing) a similar assessment system can serve as mentors and "critical friends," sharing complementary experiences, insights, and expertise through formal and informal channels (e.g., in-service workshops, structured mentoring programs, networking opportunities). Connecting to other teachers, schools, and communities engaged in similar efforts can provide a district or school with a comparative perspective on local efforts, as well as much needed information and support. It can also provide a forum for evaluating progress toward meeting goals and identifying solutions to problems that threaten to impede progress.

- Reviewing students' assessment work, either informally or as part of formal school-wide or district-wide scoring activities, can also serve as a valuable professional development opportunity. Reviewing such work

can help teachers recognize evidence of standards-related achievement and learn to reliably distinguish between different levels of performance. It can also lead to insights about how to refine assessments to better elicit evidence of student achievement in the future. In addition, ACE/C-TAP teachers report that reviewing student assessment work in relation to standards stimulates reflection on their own instructional practices and generates discussion about teaching strategies that may better support students in meeting learning goals.

Facilitating Change Without Jeopardizing Student Learning

While it is important to promote change during the development and implementation of a new assessment system, some procedures must be in place to ensure that learning is not cut short for students who are participating in yet "unproved" practices. There are a variety of ways that schools and districts can help minimize the risk to students as new or revised assessments, curricula, and instructional strategies are tried out. Several are discussed below.

An important first step in minimizing the risk to students is identifying any gaps between old and new expectations for learning. Once standards are developed or adopted, a school or district can carefully review curriculum, instruction, and assessments to check their alignment with the educational outcomes specified in the standards. This analysis should identify strengths that can be built upon, as well as gaps between current policy and practice and the outcomes represented in the standards. Development efforts can then be targeted toward filling gaps to help ensure that students have opportunities both to learn and demonstrate targeted knowledge and skills.

For example, many sets of standards now emphasize problem-solving skills as an important educational outcome. Teaching students problem-solving skills, however, requires curricular and instructional strategies that recognize students as constructors of meaning rather than mere recipients of information. It also requires assessments that call for students to use their knowledge in non-routine ways. If curriculum, instruction, and assessment used in a school or district have focused primarily on the transmission of knowledge and routine skills, then staff must alter their practices if students are to achieve learning goals related to problem-solving. More specifically, school or district staff will need to identify or develop the following: 1) the specific problem-solving skills to be taught, 2) appropriate strategies for teaching and assessing problem-solving skills, and 3) professional

development opportunities aimed at helping teachers guide students as they develop their capacity as independent problem solvers.

A second way to minimize risk to students involves keeping a close eye on the change process as it progresses. While individuals should be encouraged to plunge into the development and implementation process, they must also be encouraged to monitor their efforts regularly to identify both positive and potentially destructive consequences, and to reorganize their efforts as necessary (Fullan with Stiegelbauer, 1991). Analyzing students' responses to individual assessments can indicate places where expectations may be unclear or where specific assessments fail to elicit the type or depth of performance desired. Specific shortcomings in individual assessments, related scoring procedures, curriculum, or instruction can then be altered in an effort to improve the overall system. Teachers can also check to see whether too many learning outcomes have been assigned to a single course, and make recommendations for reexamining and reprioritizing the standards across the curriculum.

In addition, the entire assessment system should be examined periodically for the degree to which it produces unnecessary or redundant information. Some redundancy is desirable (i.e., information on similar aspects of student achievement collected from different types of assessments). However, if information from one or more assessments is not proving useful, the assessment(s) should be modified or dropped.

A third way in which schools and districts can minimize risk to students involves limiting the student population exposed to the assessment changes, especially during the early phases of development and implementation. One way to do this is to phase in a new assessment system slowly over time. By developing and implementing a system in phases, a school or district can limit the number of students affected by new assessments and instructional techniques until such techniques have been tried out repeatedly and proven effective. Ideas for phasing in an assessment system over time are discussed in more detail below.

Developing and Implementing a New Assessment System in Phases

Developing and implementing a standards-based assessment system is a complex process involving multiple steps, changes in established practice, and, as described above, potential risks to student learning. To help

participants manage the complexity, change, and risk involved, it is wise to develop and implement a new assessment system in phases. Table 7.2 lists several steps that schools and districts can take to develop and implement an assessment system in phases. Each step is discussed in more detail below.

Table 7.2 **Steps for Developing and Implementing a New Assessment System in Phases**

- Building on existing assessments and instructional practices
- Developing and implementing new assessments sequentially (i.e., one at a time)
- Starting small (i.e., in a single program, career area, or academic department) and then expanding

Building on Existing Assessments and Instructional Practices

When developing a new assessment system, it is not necessary, and probably not desirable, for a school or district to discard its old system (i.e., whatever assessment practices it has in place) and immediately replace it with a new, entirely unfamiliar system. Instead, development and implementation can proceed in phases, starting with familiar practices as a foundation. During the first phase of development, for example, schools or districts can identify and refine existing assessments and instructional practices that already align with targeted standards. During this phase they can also determine what new assessments, instructional practices, and curriculum are needed to complete the assessment system over time. Not until this initial phase is complete, should a school or district begin the next phase of actually developing and/or selecting new assessments for implementation. By building on existing practices, schools and districts acknowledge successful assessment efforts already underway, limit the amount of new development required, and allow time for reflection and discussion about the gaps between the old and new systems.

Developing and Implementing New Assessments Sequentially

When developing assessments for a new system, it is usually best to develop and implement the assessments sequentially, or in other words, one at a time. Each assessment can be tried out and refined until acceptable levels of validity and reliability are achieved, and until teachers feel confident in their ability to effectively prepare students for, and support them through,

the assessment process. Developing and trying out one assessment at a time helps school and district staff to develop competence with a system's assessments at a comfortable pace. It also helps minimize risk to students by limiting the number of new, yet "unproved" practices experimented with at any one time.

When choosing to develop and implement new assessments sequentially, it is usually best to start with an assessment that is relatively easy to integrate into an existing instructional program. For example, C-TAP teachers who heavily emphasize project-based learning and hands-on application of knowledge in their classes, often prefer to introduce the project assessment first, since many of the requirements of the project assessment (e.g., project planning and monitoring, the creation of a major product or event, self-evaluation) are already familiar to students. Other C-TAP teachers have found it easiest to start with the portfolio assessment because the career preparation and career-technical skills and knowledge assessed by the portfolio are usually a part of the existing curriculum. In addition, the individual entries of the portfolio (e.g., Career Development Package, Writing Samples) can be introduced one at a time so that teachers and students can grow comfortable with each part of the assessment gradually. The sequence in which a teacher introduces the entries can also facilitate adjustment to the new assessment. Teachers can start with a relatively straightforward entry (e.g., Career Development Package) before moving on to more complex, difficult entries (e.g., work samples, writing samples).

Starting Small and Then Expanding

Another strategy for phasing in an assessment system is to start small, generate a track record of success, and then expand efforts to a wider scale. In this context, starting small generally means initiating development and implementation efforts within a single program, career area, or academic department. The assessment system can be experimented with and refined within the program, career area, or academic department (e.g., one assessment at a time if desired) until it consistently produces the outcomes intended. Once the system, including related instructional practices, is working well on a small scale it can be expanded to other areas within a school or district.

Starting small and generating a track record of success allows a school or district to test the effectiveness of a system on a small scale before expending the time and resources necessary to use the system across an

entire school or district. Again, it also helps minimize risks to students by limiting the number of students exposed to new assessments and related changes in instruction and curriculum. Expanding assessment efforts to a wider scale can have various benefits, including the generation of comparable achievement information across programs and opportunities to make professional development efforts more efficient (i.e., teachers across different programs and departments can be trained to use new assessments at the same time).

Many schools and districts using C-TAP have introduced and begun to institutionalize the assessment system by starting small and then expanding efforts. Perhaps because of their structure and dependence on collaboration among teachers, career academies (i.e., schools within schools that focus on career preparation) have been most successful in expanding the use of C-TAP assessments beyond a single teacher. Typically, one or more career-technical teachers in an academy learn to use the C-TAP system, and then introduce the system to other academy colleagues. Eventually each teacher in the academy takes responsibility for implementing some part of C-TAP, as illustrated in the sample career academy portfolio schedule in Appendix E. After all assessments are well established in the initial career academy, school staff can, and often do, consider expanding the C-TAP system to another career academy or to an academic department outside the academy. The teachers in the initial academy serve as resources to new teachers adopting and learning to use the C-TAP assessments.

Table 7.3 shows the timetable that one health careers academy in Southern California followed to expand the implementation of C-TAP assessments within the academy. Prior to their involvement with C-TAP, the five teachers participating in this academy (teachers of health, science, English/language arts, history, and mathematics) met weekly to coordinate their curricular program. Since adopting C-TAP, they have used these weekly meetings to gradually organize the implementation of the C-TAP portfolio and project assessments. They also meet regularly with an advisory board consisting of industry representatives, community college faculty, and parents, to discuss the C-TAP program, students' work placements, students' performance, and general school-to-work issues.

Table 7.3 Expansion of C-TAP throughout a Health Career Academy Grades 10-12

Year 1

- Health teacher gets initial C-TAP training; initiates senior projects.

Year 2

- Health teacher continues projects; initiates portfolios after getting additional C-TAP training.
- Health teacher organizes "project presentation evenings" for parents.

Year 3

- Health teacher continues projects and portfolios; initiates administration of written scenarios.
- First English/language arts teacher gets involved (i.e., helping with writing samples in the portfolio); three other English teachers follow.
- Students present projects and portfolios to School Board, underclassmen, and parents (Note: School board members and attending teachers are impressed).
- Health and English/language arts teachers get additional C-TAP training; history and science teachers are introduced to the C-TAP system and receive training.
- Health teacher participates in C-TAP scoring.

Year 4

- Health teacher continues projects; creates and administers own curriculum-based written scenarios.
- Portfolios continue with each teacher assuming a specific responsibility for portfolio implementation (e.g., English/language arts teacher responsible for 12th grade writing sample, mathematics teacher responsible for 10th grade resume).
- More presentations to the School Board and other guests.
- All teachers get additional C-TAP training.
- Health teacher participates in C-TAP item writing and scenario benchmarking and scoring sessions.

In schools without career academies, implementation of C-TAP often begins with an individual teacher or small group of teachers within a career-technical program or department. As with career academies, the teacher(s) learn to use the C-TAP assessments and then help colleagues within their program or department learn to implement the system. Sometimes the career-technical teachers invite peers from one or more academic subject areas to collaborate in C-TAP assessment efforts, as described earlier in this chapter.

The following vignette describes one district's efforts to introduce C-TAP assessments into its career education programs.

How One District Began Implementing C-TAP

One relatively large Southern California district has been involved in a serious change effort since 1988, when *Second to None* and other literature on educational reform inspired a desire for change among the district's administrators and teachers. The three high schools in the district (two comprehensive and one continuation) house three career preparation programs: business education, family consumer science (formerly home economics), and health and medical services. These programs currently offer a certificate of mastery, and are considering offering a certificate of completion (i.e., less rigorous requirements, but still reflective of acceptable performance in a coordinated program).

The adoption and implementation of C-TAP in the district coincided with the actual development of the career preparation programs themselves. C-TAP was first incorporated into the business education program. Administrators (including the Director of Career Preparation and the Assistant Superintendent for Curriculum and Instruction) and business education teachers worked together for two years to design a well-articulated sequence of courses across grades 9-12. They used the existing draft Model Curriculum Standards in business education to guide their efforts. As they established course objectives matched to standards, they also set draft performance standards which provided teachers with a common set of expectations with which to begin work. At the same time, they sought an assessment system that would be compatible with a standards-based, student-centered approach to learning. They envisioned using portfolios and projects as their major assessment vehicles. C-TAP seemed to meet their needs because of its rigorous portfolio and project components and its well-defined school-to-work connections. The development team received C-TAP training and worked closely with the California Department of Education and C-TAP project staff as they worked to incorporate the assessment system into their business education program. They also set up an advisory group to help review student work and to secure community internship sites for students. The advisory group also helped establish the performance standards. The assessment components are scored with C-TAP rubrics, and students' performances are a major factor in determining whether they will receive a certificate of mastery in business education.

The district has been recognized in recent years for its certificate program in business education, partly because of its C-TAP portfolios. The success of this program stimulated the Family Consumer Science program to adopt C-TAP; and now the Health and Medical Services faculty is in the process of planning how to incorporate C-TAP into the newly forming program, which will be integrated with the science department.

Meeting the Needs of All Students

Developing and implementing a standards-based assessment system must include efforts to ensure that all students, including students with special needs (e.g., English learners, students from "minority" cultures, students with diagnosed learning or developmental needs) are assessed equitably and receive the support needed to achieve targeted standards. Table 7.4 summarizes some of the steps that schools or districts can consider as they attempt to meet the needs of all students. The first step, which involves the use of assessment data, describes a strategy that can be used for all students. The other four steps describe strategies that are particularly useful for special needs students. Each step is discussed in more detail below.

Table 7.4 Steps in Meeting the Needs of All Students

- Using assessment data to identify patterns in student achievement and to inform instructional planning
- Inviting special needs experts to participate in assessment development and implementation efforts
- Developing multiple ways for students to represent what they know and can do
- Mediating student performances on assessments
- Ensuring that lack of English proficiency is not confused with lack of subject matter knowledge when evaluating student work

Using Assessment Data to Identify Patterns in Student Achievement and Inform Instructional Planning

As mentioned in previous chapters, analyzing assessment data (e.g., student work, scores on entire assessments or components of assessments) can help identify strengths and weaknesses in the performance of individual students and groups of students. The results of such analysis provides an essential foundation for planning ways to improve the achievement of all students vis-à-vis standards.

For example, assessment data can be evaluated to identify patterns of good or poor performance (in relation to specific content) across students at the classroom, program or department, school, or district level. Skills that are performed poorly by large numbers of students may point to the need for

changes in curriculum or instruction. Once such patterns are evident, specific strategies should be designed to help students learn and effectively demonstrate the knowledge and skills in question.

Similarly, assessment data can be analyzed to identify systematic differences in achievement between different groups of students. For example, are English learners doing as well as native English speakers on assessments? Are the performances of boys and girls on par with each other? If significant disparities between groups of students are found, steps can be taken to determine the reasons for the disparities, and to identify specific strategies for improving the achievement of low-performing student groups. These strategies are likely to include changes in instruction (e.g., providing additional language instruction, "scaffolding" learning experiences to a greater extent), and could include changes in the assessments themselves (e.g., increasing the different ways in which students may demonstrate knowledge and skills, increasing opportunities for support and guidance during the assessment process).

At the classroom level, teachers can review student assessment work-in-progress and engage in dialogue with students to better pinpoint needs for additional instruction and support for individual students. They can then develop specific personalized strategies for helping such students meet targeted learning goals.

Inviting Special Needs Experts to Participate in Development and Implementation

An important step in supporting English learners, students from minority cultures, and students with diagnosed learning or developmental needs through the assessment process is to invite staff who work with these students regularly to participate in assessment development and implementation efforts. Utilizing these individuals' expertise from the start helps ensure that the special needs of these students are considered at all points during the assessment development and implementation process.

For example, teachers who are proficient in the use of English language development techniques, Specially Designed Academic Instruction in English (SDAIE) techniques, and strategies for structuring and supporting the learning of special education students are invaluable resources for modifying instruction and assessments to meet the needs of diverse students. School counselors who have worked with a variety of special needs students are also excellent resources. They can help identify the home-school norms of

students that may conflict with planned assessment and instruction, as well as help support such students as they work on assessments. Both specialist teachers and counselors can also share their expertise with peers, helping all staff improve learning and assessment for special needs students.

Developing Multiple Ways for Students to Represent What They Know and Can Do

If an assessment is to capture what all students have learned, it needs to be flexible, allowing for different ways of representing or demonstrating knowledge and skills. Assessments that cannot be modified to provide such flexibility need to be complemented by assessments that can. This is essential for all students.

In addition, special needs students must not be evaluated solely on the basis of traditional on-demand assessments that require students to process language quickly and that provide few contextual cues to support student understanding. Such assessments rarely provide adequate or reliable information about what these students know and can do. A multiple-measures system that utilizes a variety of assessments, including those that provide opportunities for students to refine their work over time with guidance (i.e., cumulative assessments), is one way to provide more reliable information about the abilities of special needs students, as well as other students.

Mediating Student Performances on Assessments

Mediating (or scaffolding) the administration of assessments is another way to support special needs students through the assessment process. When teachers or others mediate the administration of an assessment they provide input (e.g., information, modeling, feedback) to students during the assessment process. The input is usually designed to help clarify students' understanding of the questions or task(s) at hand, and to improve the overall quality of their finished work. Mediation allows schools and districts to determine the best work students can produce when provided with some teacher support. This technique has been prevalent in special education for some time (Feuerstein, 1979; Samuda et al., 1989). It is also quite common in cumulative assessments that have gained popularity in recent years (e.g., projects and portfolios).

When developing and implementing a standards-based assessment system, a school or district will need to decide on the level of mediation that

is acceptable for the various assessments in its system. They should aim to support students with special needs without making assessments too easy or unmeaningful as measurement instruments. The level of mediation deemed acceptable by a school or district is likely to vary by assessment type. Generally speaking, many schools or districts will allow less mediation for highly standardized written on-demand assessments than for more complex cumulative assessments. Acceptable strategies for mediating the administration of written on-demand assessments might include repeating instructions, reading a question aloud for a student, allowing the use of a dictionary during testing, or, in some cases, altering the amount of time allowed for completion. A less appropriate means for mediating an on-demand assessment might be discussing the meaning of specific questions in depth. Examples of acceptable mediation strategies for cumulative assessments include helping brainstorm standards-related topics for a project, providing students with specific feedback and allowing them to improve their work based on the feedback.

For accountability purposes, all mediation provided to a student during an assessment should be documented so that interested parties (e.g., teachers, parents, employers) will understand exactly what a student can do under what circumstances. A school, district, or state should be able to gather performance data on each student and to disaggregate data for students who required mediation beyond what is normally permitted (Ysseldyke, 1994; Olsen, et al., 1994). Documentation of mediation can be taken into account when examining the student's score.

Ensuring that Lack of English Proficiency is Not Confused with Lack of Subject Matter Knowledge when Evaluating Student Work

When evaluating the work of English learners, it is all too easy to confuse lack of English language proficiency with lack of subject matter knowledge. There are several steps that schools and districts can take to help ensure that this does not happen. First, unless language proficiency is being assessed, it is important that the scoring criteria used to evaluate student work focus primarily on the content-specific knowledge and skills being assessed and not on language skills.

Second, when evaluating the work of an English learner, it helps to utilize scorers who have some knowledge of the student's primary language. A scorer who is familiar with a student's first language may be able to understand information or ideas a student is trying to express more readily

than a scorer without familiarity with the language. For example, teachers familiar with certain Asian languages may interpret sentences where articles or the verb "to be" is missing more readily than teachers unfamiliar with such languages. Teachers knowledgeable about Spanish spelling patterns may be able to recognize misspelled words (e.g., pipel for people, polait for polite) more readily than a scorer unfamiliar with Spanish. Districts can improve the evaluation of student work by periodically providing opportunities for teachers with knowledge of students' various languages and cultures to work with each other to examine student work.

Of course, when communication is one of the criteria for evaluating student performances, elements of language proficiency are being legitimately assessed. Aspects of communication skill are germane to readiness for certain kinds of jobs as well. English learners also need the type of feedback available from assessment of their English communication abilities.

Establishing Community-Wide Support

Establishing community support is critical when developing and implementing a new assessment system. Without such support, a district or school risks engendering opposition to or outright rejection of different parts of the system or the entire system itself. To help establish community support, those leading the development and implementation efforts must help community members (e.g., parents, business and industry representatives, elected officials, members of the tax-paying public) to understand the new assessment system and to have faith that any problems with the system (e.g., with individual assessments, assessment scoring procedures, reporting of student achievement) will be detected and resolved before they affect student learning. Table 7.5 summarizes two steps that schools and districts can take to help achieve these ends and thereby establish community support for a multiple-measures, standards-based assessment system.

Table 7.5 Steps in Establishing Community-Wide Support

- Keeping the community informed about assessment plans and progress
- Providing opportunities for community members to participate in development efforts

Keeping the Community Informed About Assessment Plans and Progress

Virtually every school or district has existing mechanisms through which educational issues can be explained and discussed (e.g., the school board, the PTA, parent advisory groups, school site councils, special task forces, ongoing committees, newsletters, open houses, "back to school nights"). As an assessment system is planned, developed, and tried out, schools and districts can utilize these mechanisms to communicate with the community about assessment plans and progress.

Efforts to communicate with the community about a new assessment system should begin when the system is first conceived and continue through its development and ongoing use. The goal of such efforts should be to help community members understand various aspects of the assessment system and the development and implementation processes, including the following: the standards (i.e., the expectations for learning) that guide assessment and instruction; the goals and benefits of the assessment system; the characteristics of specific assessments within the system; the ways in which curriculum and instruction may need to change (or are changing) to support student learning and success on assessments; the methods used to evaluate students' assessment work; the methods used to report student achievement; and the challenges anticipated or encountered during the development and implementations processes. To accomplish this goal, schools and districts will need to provide community members with information, as well as with opportunities to discuss the information they receive and to ask questions and raise concerns. By providing information about an assessment system and responding to community members' questions and concerns, schools and districts can do much to quell unfounded suspicions or misconceptions that might form when individuals are uninformed or lack sufficient information about an assessment system.

There are many ways in which schools and districts can provide the community with information about an assessment system. For example, as new assessments are developed, a school or district can present community members with sample items or tasks, and explain how these items or tasks are linked to targeted standards, curricula, and instructional practices. Once assessments have been tried out and scored, a school or district can present sample student responses, along with scoring criteria, to show the community how assessments can elicit evidence of students' standards-based achievement. In viewing samples of student work over time, members of the

community may be able to see improvements in student learning that can be attributed to changes in instruction and assessment.

It should be noted that school staff are an important part of the community that needs to be kept informed about assessment plans and progress. For example, if a district is trying out a new assessment in one school first, staff in the district's other schools (i.e., staff that will eventually use the system) should be kept informed about assessment development and implementation efforts. Similarly, if a school is trying out a new assessment with one grade or class, then all other staff members in the school should be kept informed about the development and implementation progress. By keeping all staff informed, a school or district helps avoid creating a sense that an elite group is single-handedly spearheading an effort that will ultimately be thrust upon others.

Providing Opportunities for Community Members to Participate in Development Efforts

Schools and districts can also establish community support by providing opportunities for community members to actively participate in the assessment development process. The same mechanisms through which a school or district can communicate information about a new assessment system (e.g., the school board, the PTA, parent advisory groups, school site councils) can also be used to involve community members in the assessment development process. Several examples of ways in which community members can participate in development efforts are described below.

First, as mentioned in Chapter 1, community members can help develop and refine the standards upon which an assessment system will be based. Some community members (e.g., business and industry representatives) can participate in writing the standards by helping to identify and describe the specific knowledge and skills that students should learn and demonstrate. Other community members can review and discuss draft standards once they are written, providing feedback on the language used or the content covered. Before asking community members for feedback, however, it is important for schools or districts to ensure that all participants understand the similarities and differences between current and proposed learning objectives, the rationale for increased and decreased emphases of particular sets of knowledge and skills, and any possible changes that the standards will necessitate in instruction and assessment.

Community members can also serve on advisory committees that provide ongoing input to schools regarding curriculum, instruction, and assessment. For example, C-TAP teachers at one school met regularly with advisory committees in different career areas to review progress as the teachers adapted and implemented the C-TAP system (for local use). These advisory committees included both parents and business/industry representatives, which helped ensure that the C-TAP assessments and resulting student work were reviewed from a variety of perspectives. Parents, for example, focused on the assessments' effects on students and often questioned procedures that business/industry representatives took for granted (e.g., specific grading policies and procedures for providing feedback on performance). The business/industry representatives provided suggestions that made the assessments more meaningful to future employers, and helped teachers prioritize instruction and assessment goals. Throughout the process, members of the advisory committees provided early signs of adverse reactions and gaps in understanding that assessment developers could then address.

Finally, strategies for actively involving community members in the development process can also be applied at the school level (i.e., within the school community). Some teachers and administrators can be asked to participate on the leadership team that will oversee assessment development efforts. Teachers and administrators who are not part of the leadership team, including those who may not use the new assessment system immediately, can help develop or review standards and assessments. They can also help articulate scoring criteria and evaluate the student work that results from new assessments, either informally or during more formal benchmark and scoring sessions. In these ways, support for a new assessment system can be cultivated within the school community.

Regardless of the specific strategies that schools or districts use to involve the community, the process for developing a new assessment system should be genuinely open to review if community support for the system is to be achieved. Schools and districts should keep in mind, however, that it will be impossible to incorporate all community concerns and ideas into the development process. A summary of community input and what was done (or not done) in response to that input can then be disseminated by schools and districts. In this way, each community group gets its opinion voiced and goes on record supporting or not supporting particular standards, assessments, or other aspects of the overall system.

Coordinating Local and State Assessment Efforts

Responsibility for student assessment can be shared between the state and local districts. A vision of shared responsibility suggests the need for some basic forms of coordination between district and state assessment efforts. Table 7.6 summarizes two general steps that districts can take to help coordinate their assessment efforts with those of the state. Each strategy is described in more detail below.

Table 7.6 **Steps in Coordinating Local and Statewide Assessment**

- Linking local and state standards
- Incorporating statewide assessments into a local assessment system

Linking Local and State Standards

A district can begin to coordinate its assessment efforts with those of the state by forging links between the district's and the state's standards, where they exist. As mentioned in Chapter 1, local standards should reflect, at a minimum, the same content (i.e., knowledge and skills) and rigor emphasized in the standards underlying various statewide assessments. Linking local and state standards helps to ensure that the curriculum, instruction, and assessments provided at the local level will support student success on statewide assessments. It also makes it possible to use results from statewide assessments within a local assessment system. (This point is described further in the next section.)

A first step in forging links between local and state standards is identifying the similarities and differences between the two sets of standards. One way to do this is to create a table, list, or graphic that illustrates the relationship of local standards to state standards. By examining such a table, list, or graphic, a district can learn if local standards adequately reflect the knowledge and skills covered in state standards. If not, the district should change or make additions to the local standards as necessary.

If local standards have not yet been developed, a district can use the state standards themselves as a guide for developing local standards, thus forging a direct link between the two sets of standards.

Incorporating Statewide Assessments into a Local Assessment System

If a district links its standards to state standards, as recommended previously, it is likely that the district can use the assessment information collected, scored, and reported by the state as part of the district's local assessment system. For this reason, when considering the types of assessments to include in a local assessment system, a district should examine statewide assessments to determine the potential role(s) they might play at the local level.

For example, the California Department of Education (CDE) administers a variety of statewide assessments, including required standardized tests and optional end-of-course high school examinations within selected academic and career-technical areas (i.e., the Golden State Examinations and Assessments in Career Education). A district whose own standards align with the content underlying these statewide assessments could use results from one or more of these state tests (i.e., student scores) as measures of student achievement within their own local system. This could result in considerable cost savings to a district by limiting the number of new assessments that need to be developed and administered at the local level. A district could then focus its attention on developing and implementing assessments that measure knowledge and skills not covered by statewide assessments and that probe the depth of students' knowledge and abilities in ways that standardized state tests cannot.

In addition to using assessment information collected, scored, and reported by the state, a district that incorporates state assessments into its local assessment system can take advantage of a variety of materials and training that support both state assessments (e.g., GSE, ACE) and local assessments (e.g., C-TAP), including the following:

- GSE and ACE Guides for Teachers: These guides explain features of the GSE and ACE testing programs, and provide sample test questions, including general scoring criteria and acceptable student responses for written-response items. They are available for each subject area for which a GSE or ACE assessment is given.
- C-TAP Teacher and Student Guidebooks: These guidebooks describe the requirements for the C-TAP assessment components (i.e., portfolio, project, written scenario). The student version also includes examples, organizers (checklists), and hints for completion and explains how each assessment component will be evaluated. The teacher version describes

strategies for implementing each C-TAP assessment and evaluating student work.

- **C-TAP Guides to Evaluating Student Work:** These guides explain and illustrate how to evaluate student responses to C-TAP assessments (i.e., portfolio, project, and written scenarios) using holistic and dimensional scoring guides and benchmark performances. They are available for selected C-TAP assessments within selected career-technical areas.
- **GSE and ACE Scoring Activities:** Teachers from various academic and career-technical areas can participate in GSE and ACE scoring activities. During such activities, teachers are trained to reliably evaluate student work using scoring rubrics and benchmark performances. This training, which is free to districts, can be a valuable source of professional development, deepening teachers' understanding of assessment-related issues and their ability to effectively distinguish between different levels of student performance.

Summary

There are several steps that schools and districts can take to support the overall development and implementation of a multiple-measures, standards-based assessment system.

First, a school or district can facilitate development and implementation efforts by ensuring that organizational conditions supportive of change are in place. Among the steps that schools and districts can take to strengthen organizational conditions for change are the following: committing both time and resources to development and change over years; establishing a strong leadership team that is able to shepherd an assessment system through numerous stages of modification and refinement; encouraging collaboration among staff in an effort to meet shared goals; providing professional development opportunities that focus on the knowledge and skills needed to develop and/or implement an assessment system effectively; and facilitating change without jeopardizing student learning as new or revised assessments, curricula, and instructional strategies are tried out and modified.

Second, schools or districts can develop and/or implement a new assessment system in phases in order to help participants manage the complexity, change, and risks involved in the process. For example, a school or district can start by identifying and refining existing assessments and

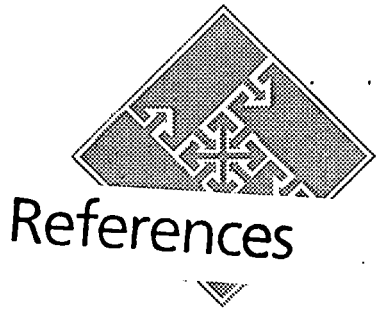
instructional practices that align with targeted standards. Next a school or district can determine what new assessments and practices will be needed to complete the system over time. New assessments can then be developed sequentially (i.e., one at a time) until acceptable levels of validity and reliability are achieved. This sequential development will help teachers become competent with each assessment at a comfortable pace, and will help minimize risk to students by limiting the number of unproved practices being used at any one time. Schools or districts can also test the effectiveness of a system by initiating assessment development efforts on a small scale (i.e., in one program, career area, or academic department), and then generating a track record of success before expanding efforts to other areas in a school or district.

Third, during assessment development and implementation efforts, schools and districts must work to ensure that all students, especially students with special needs (e.g., English learners, students from "minority" cultures, students with diagnosed learning or developmental needs) can be assessed equitably and receive the support needed to achieve targeted standards. Among the steps that schools and districts can take to achieve these ends are the following: analyzing assessment data to identify patterns in student performance and to then plan ways to improve instruction and student achievement; inviting special needs experts to participate in assessment development and implementation efforts as a way of ensuring that the needs of special students are considered throughout the process; developing multiple ways (i.e., a variety of on-demand and cumulative assessments that are as flexible as possible) for all students to represent and demonstrate what they know and can do; mediating student performances on assessments; and ensuring that lack of English proficiency is not confused with lack of subject matter knowledge when evaluating student work.

Fourth, from the time a new assessment system is conceived through its development and ongoing use, schools and districts must work to establish community support for the system. Along these lines, schools can utilize existing mechanisms for communication (e.g., the school board, the PTA, parent and industry advisory groups, school site councils, newsletters, open houses) to provide community members with information about the assessment system, as well as with opportunities to discuss that information and to ask questions and raise concerns. In addition, schools and districts can provide opportunities for community members to actively participate in the assessment development process (i.e., helping develop and/or review standards and assessment tasks, sitting on advisory committees that monitor assessment progress and results). By disseminating information about an

assessment system and providing opportunities to discuss and participate in development efforts, schools and districts can do much to quell unfounded suspicions or misconceptions that might form when individuals are uninformed or lacking sufficient information.

Finally, districts can benefit by coordinating their own assessment efforts with those of the state. Districts can begin this process by forging links between district and state content and performance standards (where they exist), making sure that district standards reflect the same content and rigor as the state standards. Creating such links will help ensure that the curriculum, instruction, and assessments provided at the local level support student success on statewide assessments. In addition, if a district links its standards to state standards, it is likely that the district can use assessment information collected, scored, and reported by the state as part its own local assessment system. This can result in considerable cost savings to a district by limiting the number of new assessments that need to be developed and administered at the local level.



References

Ananda, S., Rabinowitz, S., Carlos, L., & Yamashiro, K. (1995). *Skills for tomorrow's workforce.* San Francisco: Far West Laboratory for Educational Research and Development (now WestEd).

Anastasi, A. (1982). *Psychological testing* (5th ed.). New York: Macmillan Publishing Company, Inc.

Bailey, J., & McTighe, J. (1996). Reporting achievement at the secondary level: What and how. In T. R. Guskey (Ed.), *Communicating student learning: 1996 Association for Supervision and Curriculum Development yearbook.* Alexandria, VA: Association for Supervision and Curriculum Development.

Baron, J. (1996). Developing performance-based student assessments: The Connecticut experience. In J. Baron and D. Palmer Wolfe (Eds.), *Performance-based student assessment: Challenges and possibilities. The ninety-fifth yearbook of the National Society for the Study of Education.* Chicago: The University of Chicago Press.

California Department of Education - Agricultural Education Unit and Career-Vocational Education Division in cooperation with the University of California at Davis. (1993). *Draft agriculture performance standards and integrated activities, grades 9-12.* Sacramento, CA: Author.

California Department of Education. (1995). *State school-to-career plan.* Sacramento, CA: Author.

California Department of Education - Superintendent's Challenge Initiative. (1995). *Career preparation standards: Draft interim content and performance standards in business education* (Computer Science and Information Systems Career Path Standards). Sacramento, CA: Author.

California Department of Education. (1996). *Golden State examination in economics: A guide for teachers and students.* Sacramento, CA: Author.

California Department of Education - High School Teaching and Learning Office. (1996b). *Industrial and technology education career path guide and model curriculum standards.* Sacramento, CA: Author.

California Department of Education. (1997, September). Combining scores from multiple measures to identify a standards-based level of individual student schievement. In *Standards, students, and success.* Session V at the "Schools In" Conference, Sacramento, CA.

California Department of Education - School and District Accountability Division. (1998). *Multiple measures: Models for combining measures to determine whether students meet grade-level standards.* Sacramento, CA.

Career-Technical Assessment Program for the California Department of Education. (1995). *Career Technical Assessment Program (C-TAP) Teacher Guidebook.* San Francisco: Far West Laboratory for Educational Research and Development (now WestEd).

Council of Chief State School Officers Workplace Readiness Assessment Consortium. (1995). *Consensus framework for workplace readiness assessment.* Washington, DC: Author.

Far West Laboratory for Educational Research and Development (now WestEd) (1995). *New directions: Reframing assessment.* San Francisco: Author.

Feuerstein, R. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device, theory, instrument, and techniques.* Baltimore, MD: University Park Press.

Fullan, M. with Stiegelbauer, S. (1991). *The new meaning of educational change.* New York: Teachers College Press.

Gandal, M. (1996). *Making standards matter.* Washington, DC: American Federation of Teachers.

Gronlund, N. (1985). *Measurement and evaluation in teaching* (5th ed.). New York: Macmillan Publishing Company.

Haladyna, T. (1994). *Developing and validating multiple-choice test items.* Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Herman, J., Aschbacher, P., & Winters, L. (1992). *A practical guide to alternative assessment.* Alexandria, VA: Association for Supervision and Curriculum Development.

Kane, Michael. (1994). Validating the performance standards associated with passing scores. *Review of educational research*, 64(3).

Katz, L., & Chard, S. (1989). *Engaging children's minds: The project approach.* Norwood, NJ: Ablex Publishing Corporation.

Kenney, E., & Perry, S. (1994). Talking with parents about performance-based report cards. *Educational leadership,* 52(2), 24-27.

Koelsch, N., Estrin, E., & Farr, B. (1995). *Guide to developing equitable performance assessments.* San Francisco: Far West Laboratory for Educational Research and Development (now WestEd).

McLaughlin, M., & Shepard, L., with J. O'Day. (1995). *Improving education through standards-based reform: A report by the National Academy of Education Panel on Standards-Based Education Reform.* Stanford, CA: National Academy of Education.

National Research Council. (1996). *National science education standards.* Washington, DC: National Academy of Sciences.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

National Center on Education and the Economy and the University of Pittsburgh (New Standards Project). (1997). *Performance standards: English-language arts, mathematics, science, applied learning (Volume 3: High School).* Washington, DC: Author.

Olsen, L., Chang, H., Salazar, D., Leong, C., Perez, Z. M., McLain, G., & Raffel, L. (1994). *The unfinished journey: Restructuring schools in a diverse society.* San Francisco: California Tomorrow.

Peckron, K. (1996). Beyond the A: Communicating the learning progress of gifted students. In T. R. Guskey (Ed.), *Communicating student learning: 1996 Association for Supervision and Curriculum Development yearbook.* Alexandria, VA: Association for Supervision and Curriculum Development, 1996.

Resnick, L. (1996). *Performance puzzles: Issues in measuring capabilities and certifying accomplishments.* Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Rogers, L. (1996). *The California freshwater shrimp project: An example of environmental project-based learning.* Berkeley, CA: Heyday Books.

Samuda, R. J., Kong, S. L., Cummins, J., Lewis, J., & Pascual-Leone, J. (1989). *Assessment and placement of minority students.* Kinston/Toronto: Intercultural Social Sciences.

School-to-Work Opportunities Act of 1994, 20 U. S. C. A. 6101 (1994).

Secretary's Commission on Achieving Necessary Skills (SCANS). (1991). *What work requires of schools: A SCANS report for America 2000.* Washington, DC: U. S. Department of Labor.

Stiggins, R. (1994). *Student-centered classroom assessment.* New York: Macmillan College Publishing Company.

Ysseldyke, J. (1994). *Recommendations for making decisions about the participation of students with disabilities in statewide assessment programs: A report on a working conference to develop guidelines for statewide assessments and students with disabilities.* Synthesis Report 15. Alexandria, VA: National Association of State Directors of Special Education and Minneapolis, MN: National Center on Educational Outcomes.

Appendix A

An Example of How Student Work Can Illustrate a Performance Standard

The following performance standard and brief excerpts from student work are from the New Standards Project's high school performance standards. They show how the presentation of student work can facilitate understanding of specific pieces of a performance standard. To see the longer excerpts of student work and more examples relating the work to the standard, see pages 108–117 in *Performance Standards: English-Language Arts, Mathematics, Science, Applied Learning. Volume 3: High School* (National Center on Education and the Economy and University of Pittsburgh, 1997.)

The standard, "Design a Product, Service, or System," is one part of a three-part problem-solving standard. This part of the standard reads as follows:

Design a Product, Service, or System

The student designs and creates a product, service, or system to meet an identified need; that is, the student:

- develops a design proposal that:
  - shows how the ideas for the design have been developed;
  - reflects awareness of similar work done by others and of relevant design standards and regulations;
  - justifies the choices made in finalizing the design with reference, for example, to functional, aesthetic, social, economic, and environmental considerations;
  - establishes criteria for evaluating the product, service, or system; and
  - uses appropriate conventions to represent the design;
- plans and implements the steps needed to create the product, service, or system; and
- makes adjustments as needed to conform with specified standards or regulations regarding quality or safety; and
- evaluates the product, service, or system in terms of the criteria established in the design proposal, and with reference to:
  - information gathered from sources such as impact studies, product testing, or market research; and
  - comparisons with similar work done by others.

To meet this standard, one student designed and built an electric car as part of a team. The remainder of this Appendix describes the student's project and shows how selected excerpts from pieces of the student's work provide evidence related to the performance standard. Names of individuals, schools, and other locations have been blacked out to protect the privacy of the student and other individuals.

The Project: ElectroHawk 1

Students were required to complete an application project that would develop their skills in gathering and using information, communication, and problem solving, and to help them to become self-directed learners. The students defined the project and acquired a mentor from outside the school to assist them. The students were supervised by a teacher throughout the process of developing a proposal and planning a presentation of the project. The student whose work is featured in this Appendix designed an electric car for a local competition.

Circumstances of Performance

The student worked as a member of a team to get most of the work done. The student was also the actual driver of the car in competition. The team worked with an adult mentor and a teacher advisor. The students were required to maintain a journal to record the time they spent on the project. The work culminated in a presentation to interested adults and peers.

Excerpts of Student Work Provided

Excerpts are provided from three pieces of student work related to the project: the proposal paper, a timeline, and a journal. Comments on the excerpts are also provided, showing how different parts of the standard are reflected in the student's work.

The proposal explains the genesis of the project. The Public Utilities District (P.U.D.) provided the school with an electric motor, a speed control, and two batteries as the basis for designing and building an electric or solar-electric vehicle for entry in a competition with other schools in the local area.

Evidence of the process used for the design of the vehicle can be found in the proposal, timeline, and journal. The proposal paper records the plan the student envisaged early in the process. This plan is reflected in the timeline. The journal provides insight into the reality of the design process, especially the way in which the students responded to problems they encountered as the design took shape.

Two Excerpts from Proposal Paper (Student Work)

Application Project Proposal Paper

Have you ever wanted to go for a ride into the future? Or maybe drive an almost non-polluting vehicle? For my application project, I propose to build a full size, fully drivable, fully operational solar/electric car. I am currently, and will continue to build, and improve an electric vehicle. I, along with the aid of 4 other students, and the watchful eye of Mr. [redacted] and Mr. [redacted], am currently building this vehicle in the [redacted] Technology Department. The vehicle, along with the many tests and upgrades, should be completed by the end of July.

B → Once we have the entire chassis finished we can begin mounting and wiring all of the electrical components such as speed control, throttle, and batteries. Our vehicle is very compact, and finding adequate space will be difficult. We also need to wire up the vehicle, and from the schematic, it does not look easy.

H → After everything is wired up, and in place we will begin going over all of the rules and regulations to make sure that we are legal and able to race. There will be a practice day when all of the competing vehicles will turn out at [redacted] speed Way to take practice runs, as well as have a judge look over our vehicle for anything we may be missing.

I → Finally, after everything is completed we will begin doing tests and trials. Our main goal of running the various tests will be to find any flaws in the structure that may be present and get the vehicle running at it's most efficient levels. We will also begin lightening the vehicle at this point to see what the least amount of material is needed to make the vehicle hold together.

Commentary on Excerpts from Proposal Paper

B The proposal paper records some of the design issues that the student envisaged would require resolution. These are reflected in the timeline.

H The proposal paper and journal contain several references that demonstrate attention to relevant regulations and to matters related to safety.

I The students devoted a lot of time and energy to testing their design and to trying out strategies to improve its performance and efficiency. The strategies included analysis of records of performance.

Excerpt from Timeline (Student Work)

F

Application Project
Timeline

March 30, 1995: Completed
By this date the mock-up will be completed and work on the metal chassis will commence.

April 14, 1995: Completed
By this date the chassis will be completed and work on the rear suspension will commence.

April 21, 1995: Completed
By this date the rear suspension will be completed and work on the front suspension will commence.

May 9, 1995:
By this date work on the front suspension will be completed and work on wiring the car will commence.

May 6, 1995:
Work on wiring the car will be completed and safety checks will be performed. As well as checks to make sure our vehicle can satisfy the rules.

May 9, 1995:
By this date all safety parameters will be met, and performance testing of the vehicles systems will begin.

May 13, 1995:
The vehicle will be taken to [redacted] speedway to get looked over by a judge that will check to make sure that we meet all the necessary guidelines.

Commentary on Excerpt from Timeline

F

The timeline records the planned steps for turning the design into a reality while the journal entries record the ways in which those steps were achieved in practice and the modifications to the process the students made along the way.

Excerpts from Journal (Student Work)

B →

3/31 (2 hrs.)

This evening I finally finished my goal statement for this project after 3 rewrites. Several times I forgot to include little details here and there, but now I have finished and am ready to go. I also found out today that info for the front axle of the car didn't come and is now over a week late, and if we don't get an order in by the end of this week I will get a little concerned due to the time factor which is quickly becoming an enemy to us as race day approaches. Oh well, it just means longer hours.

B →

4/11 (5 hrs.)

Today we worked very hard to try and get a rear axial and suspension finished but instead we had to settle for a nearly finished rear axial. I expect we should be finished with the rear axial by our next meeting. I also started work on a very unique front axle system conceived by Mr. [redacted] using the same concept as the 3 wheel "banana" bikes at Seaside Or. I did manage to get a full mock-up of the system built. We may also still use the front axial kit. Mr. [redacted] is going to try and order one as soon as possible.

B →

4/14 (4 hrs.)

Today we made up our minds that we wanted the vehicle driveable by 4/22. And to do this we needed to devise a plan of attack. We made the decision to work this Saturday. I was given the duty to try and get [redacted] Pizza to sponsor us by giving us a couple of pizzas for lunch. We also decided to work late Tuesday and Thursday since we needed to be done the following Saturday, which is when time trails and first inspection of the vehicle. It is not absolutely necessary to have our vehicle ready on this date, but we would still like to make a showing. As far as work goes we finished the rear axial as well as the part that the axial attaches to the car. We also built the mock-up of the battery box. There isn't a whole lot of room for it, but we will do what we can.

Commentary on Excerpts from Journal

B

The journal records several instances in which the students found it necessary to adjust their priorities in order to deal with unforeseen problems and to meet deadlines.

Appendix B

Models for Combining Multiple Measures

(Adapted from California Department of Education, 1998)

The following three models represent approaches for combining multiple measures to set grade-level performance standards when results from two or three measures of achievement are available for students at a grade level in a specific subject.

Schools and districts can adapt the models presented here to fit their own circumstances. For example, they may need to adapt the models to reflect the specific measures of achievement they use and the number of different performance levels they desire within each measure.

(NOTE: A Model A does exist, but is not presented here because it involves only one measure of achievement and does not, in this context, provide useful information regarding how to combine results from multiple measures.)

Model B

This model uses two measures of achievement with only two levels of performance for each measure. It is considered a conjunctive model because to meet the grade-level standard, a student must score within the top level of performance for each measure of achievement. A strong performance on one of the measures of achievement cannot compensate for a weak performance on the other measure. Tables B.1 and B.2 provide specific examples of this approach, using the following two measures of achievement: a norm-referenced test and a class grade in Table B.1 and a norm-referenced test and a writing assessment in Table B.2.

Table B.1 A conjunctive model using results from a norm-referenced test (two levels of performance) and a class grade (two levels of performance) to determine whether students meet grade-level standards.

		Score on NORM-REFERENCED TEST*	
		1 – 49	50+
Class Grade	A – C		MGLS
	D – F		

MGLS = Meets Grade-Level Standards

* All norm-referenced test scores are stated in terms of national percentile ranks.

In the example above, a student meets the grade-level performance standard if he or she scored at or above the 50th percentile on the norm-referenced test and earned a grade of "C" or better in the relevant class (subject).

Table B.2 A conjunctive model using results from a norm-referenced test (two levels of performance) and a writing assessment (two levels of performance) to determine whether students meet grade level standards.

		Score on NORM-REFERENCED TEST*	
		1 – 49	50+
Score on WRITING ASSESSMENT	4 – 6		MGLS
	1 – 3		

MGLS = Meets Grade-Level Standards

* All norm-referenced test scores are stated in terms of national percentile ranks.

In the example above, a student meets the grade-level performance standard if he or she scored at or above the 50th percentile on the norm-referenced test and scored at least a "4" on the writing assessment.

Model C

This model uses two measures of achievement, each with more than two levels of performance. Using this model, grade-level performance standards can be defined operationally by looking simultaneously at the different levels of performance on each of the two measures of achievement, and deciding whether each possible combination of results meets or does not meet the standards. Deciding which combinations of results meet grade-level standards can be accomplished using a consensus process involving teachers, parents, and administrators and by looking at students' work and/or assessment items for each combination.

Model C represents a compensatory approach for combining multiple measures because it allows a school or district to conclude (when setting grade-level performance standards) that a superior performance on one measure of achievement can compensate for a weaker performance on the other measure. Tables B.3 and B.4 provide specific examples of this approach, using the following measures of achievement: a norm-referenced test and a class grade in Table B.3 and a norm-referenced test and a writing assessment in Table B.4.

Table B.3 A compensatory model for combining results from a norm-referenced test (six levels of performance) with a class grade (five levels of performance) to determine whether students meet grade-level standards.

Score on
NORM-REFERENCED TEST*

Class Grade	1 – 29	30 – 39	40 – 49	50 – 59	60 – 69	70+
A		MGLS	MGLS	MGLS	MGLS	MGLS
B			MGLS	MGLS	MGLS	MGLS
C				MGLS	MGLS	MGLS
D						
F						

MGLS = Meets Grade-Level Standards

* All norm-referenced test scores are stated in terms of national percentile ranks.

Table B.4 A compensatory model for combining results from a norm-referenced test (six levels of performance) with a writing assessment (six levels of performance) to determine whether students meet grade-level standards.

Score on WRITING ASSESSMENT	Score on NORM-REFERENCED TEST*					
	1 – 29	30 – 39	40 – 49	50 – 59	60 – 69	70+
6		MGLS	MGLS	MGLS	MGLS	MGLS
5			MGLS	MGLS	MGLS	MGLS
4				MGLS	MGLS	MGLS
3					MGLS	MGLS
2						
1						

MGLS = Meets Grade-Level Standards

* All norm-referenced test scores are stated in terms of national percentile ranks.

As these examples demonstrate, Model C allows for a variety of ways to meet grade-level performance standards. The white areas, labeled MGLS, represent the score combinations that result in students meeting grade-level standards. Notice, in both examples, that students with relatively low norm-referenced test scores (e.g., 30 – 39 or 40 – 49) can still meet grade-level expectations if they do well on the other measure of achievement (i.e., the class grade in Table B.3 or the writing assessment in B.4).

Model D

This model uses three measures of achievement, each with more than two levels of performance. As with Model C, grade-level performance standards can be defined (through a consensus process) by deciding which combinations of results do and do not meet desired standards of achievement. Table B.5 provides a specific example of this approach, using the following three measures of achievement: a norm-referenced test, a class grade, and a writing assessment.

Table B.5 **A compensatory model using three measures with different levels to determine combinations of assessments that meet grade-level standards.**

Class Grade	Writing Assessment Score	Score on NORM-REFERENCED TEST* 1 – 29	30 – 39	40 – 49	50 – 59	60 – 69	70+
A	6		MGLS	MGLS	MGLS	MGLS	MGLS
	5		MGLS	MGLS	MGLS	MGLS	MGLS
	4		MGLS	MGLS	MGLS	MGLS	MGLS
	3			MGLS	MGLS	MGLS	MGLS
	2						
	1						
B	6		MGLS	MGLS	MGLS	MGLS	MGLS
	5		MGLS	MGLS	MGLS	MGLS	MGLS
	4			MGLS	MGLS	MGLS	MGLS
	3				MGLS	MGLS	MGLS
	2						
	1						
C	6		MGLS	MGLS	MGLS	MGLS	MGLS
	5			MGLS	MGLS	MGLS	MGLS
	4				MGLS	MGLS	MGLS
	3					MGLS	MGLS
	2						
	1						
D	6						
	5						
	4						
	3						
	2						
	1						
F	6						
	5						
	4						
	3						
	2						
	1						

MGLS = Meets Grade-Level Standards

* All norm-referenced test scores are stated in terms of national percentile ranks.

As the example demonstrates, there can be different ways to meet grade-level standards, and performance on one measure can compensate for performance on another, but only within certain limits. Students in this example cannot meet the grade-level standards if they score below the 30th percentile on the norm-referenced test, or score below "3" on the writing assessment, or earn a grade below a "C."

NOTE: Combining the results of three measures can yield more sensitive and accurate classification of students, partially because students have multiple opportunities to demonstrate their proficiency.

Appendix C

Sample Portfolio Schedules Involving Collaboration Among Teachers

Because of the variety and relative independence of its entries, the C-TAP portfolio assessment lends itself to collaboration among teachers (i.e., within programs, across programs, at the same grade level, at different grade levels). When implementing portfolios, teachers can work together in a variety of ways, including sharing ideas for supporting students through the assessment process, dividing up responsibility for entries, establishing portfolio storage procedures, and monitoring and evaluating student progress.

This appendix provides two examples of schedules for implementing portfolio assessments, each of which involves collaboration among teachers. Example 1 shows a schedule for implementing key elements of the portfolio over the first half of a school year. Opportunities for collaboration (e.g., planning, discussing progress) are interspersed throughout the schedule. A schedule like this can vary considerably at different grade levels and in different programs, depending, in part, on course length and the amount of time students spend in class each day and/or week. Although teachers' specific roles are not outlined in the schedule, it is very important that they be clear to all collaborating teachers from the outset.

Example 2 shows an abbreviated version of a portfolio schedule from a Health Careers Academy that has been using C-TAP portfolios for some time. This schedule highlights the collaborative implementation of portfolio entries across different disciplines (i.e., different courses). At this point, coordination among the five participating teachers from different disciplines is well-established. (NOTE: A similar schedule allocating the instruction of standards-based knowledge and skills across different courses can also be developed. Basic levels of knowledge and beginning levels of proficiency in specific skills could be the focus of an introductory course. More advanced courses could increase the knowledge and skills requirements in specified ways.)

Example 1

General Collaborative C-TAP Portfolio Schedule
(August–December)

August 29	Joint teacher planning on portfolios (reviewing general plans of previous spring)
September 9-13	Introducing portfolios to students
September 13	Introducing portfolios to parents by letter
October 1	Introduction to resume writing (*Student Guidebook*; direct instruction)
October 30	Drafts of resumes due
November 5	Teachers meet to discuss progress and review schedule
November 15	"Final drafts" of resumes due
November 20	Introduction to job application process (*C-TAP Student Guidebook*)
	Students obtain sample application forms from work sites (classroom discussion follows)
December 6	Completed job applications due
December 12	Students plan for first work sample (identify topic, outline format)
December 12	Teachers meet to discuss progress, plan
December 20	Work sample draft due

Example 2 **C-TAP Portfolio Schedule from a Health Careers Academy**

Portfolio Schedule	When	What Part	Where, who is keeping information
Science Teacher	November	10th Writing Sample	Rm. C152 English/Language Arts Teacher
		11th Writing Sample	Rm. C164 History Teacher
Health Teacher	December	11th & 12th Resume	Rm. C105 Health Teacher
	February	12th Work Samples (4)	
	April	11th & 12th Job Applications	
	April-May	10th, 11th, 12th Letter of Recommendation	
	May	12th Letter of Recommendation	
English/ Language Arts Teacher	March-May	12th Writing Sample	Rm. C105 Health Teacher
Mathematics Teacher	February-March	10th Job Applications	Rm. C152 English/Language Arts Teacher
		10th Resume	Rm. C152 English/Language Arts Teacher
History Teacher	February-March	11th Work Samples (2)	Rm. C164 History Teacher
		10th Work Samples (2)	Rm. C152 English/Language Arts Teacher

Appendix D

Anastasi, A. (1996). *Psychological testing* (7th ed.). Englewood Cliffs, NJ: Prentice Hall.

Berk, Ronald A. (1986). A Consumer's guide to setting performance standards on criterion-referenced tests. *Review of educational research,* 56(1).

Brandt, R. (Ed.). (1992). *Readings from Educational leadership: Performance assessment.* Alexandria, VA: Association for Supervision and Curriculum Development.

California Assessment Collaborative. (1993). *Charting the course: Toward instructionally sound assessment.* San Francisco, CA: Far West Laboratory for Educational Research and Development (now WestEd).

Career-Technical Assessment Program for the California Department of Education. (1998). *Career-Technical Assessment Program (C-TAP) student guidebook.* San Francisco, CA: WestEd.

Career-Technical Assessment Program for the California Department of Education. (1998). *Career-Technical Assessment Program (C-TAP) teacher guidebook.* San Francisco, CA: WestEd.

Career-Technical Assessment Program for the California Department of Education. (1996). *Career-Technical Assessment Program (C-TAP) guides to evaluating student work.* San Francisco, CA: WestEd.

CRESST Line, quarterly newsletter of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), UCLA Center for the Study of Evaluation, 301 GSE and IS, mailbox 951522, Los Angeles, CA 90095-1522.

Estrin, E. (1993). *Alternative assessment: Issues in language, culture, and equity.* Knowledge Brief No. 11. San Francisco, CA: Far West Laboratory for Educational Research and Development (now WestEd).

Evaluation Comment, quarterly publication of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), UCLA Center for the Study of Evaluation, 301 GSE and IS, mailbox 951522, Los Angeles, CA 90095-1522.

Fullan, M. with Stiegelbauer, S. (1991). *The new meaning of educational change.* New York: Teachers College Press.

Gifford, B. and O'Connor, M. (1992). *Changing assessments: Alternative views of aptitude, achievement and instruction.* Boston, MA: Kluwer Academic Publishers.

Gronlund, N. (1994). *Measurement and assessment in teaching* (7th ed.). Englewood Cliffs, NJ: Prentice Hall.

Haladyna, T. (1994). *Developing and validating multiple-choice test items.* Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Herman, J., Aschbacher, P., and Winters, L. (1992). *A practical guide to alternative assessment.* Alexandria, VA: Association for Supervision and Curriculum Development.

Jamentz, K. (1998). *Standards: From document to dialogue.* San Francisco, CA: WestEd.

Kane, Michael. (1994). Validating the Performance Standards Associated with Raising Scores. *Review of Educational Research,* 64(3).

Koelsch, N., Estrin, E. and Farr, B. (1995). *Guide to developing equitable performance assessments.* San Francisco, CA: Far West Laboratory for Educational Research and Development (now WestEd).

Lieberman, A. and Miller, L. (1990). *Teacher development in professional practice schools.* Teachers College Record, 92(1), 105-122.

Mitchell, R. (1992). *Testing for learning.* New York: The Free Press.

National Forum on Assessment. (1995). *Principles and indicators for student assessment systems.* Cambridge, MA: National Center for Fair and Open Testing (FairTest).

Newman, F., Secada, W., and Wehlage, G. (1995). *A guide to authentic instruction and assessment: Vision, standards, and scoring.* Madison, WI: Center for Education Research.

Regional Educational Laboratory Network Program on Science and Mathematics Alternative Assessment. (1994). *A toolkit for professional developers: Alternative assessment.* Portland, OR: Northwest Regional Educational Laboratory.

Regional Education Laboratories. (1998). *Improving classroom assessment: A toolkit for professional developers.* Portland, OR: Northwest Regional Educational Laboratory.

Stiggins, R. (1994). *Student-centered classroom assessment.* New York: Macmillan College Publishing Company.

Symposium: Equity in Educational Assessment (whole issue). (1994, Spring). *Harvard educational review,* 64(1).

Wiggins, G. (1993). *Assessing student performance: Exploring the purpose and limits of testing.* San Francisco, CA: Jossey-Bass Publishers.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

NOTICE

REPRODUCTION BASIS

☐ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☑ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").