Models of Standards Implementation: Implications for the Classroom

by

Robert J. Marzano

Mid-continent Regional Educational Laboratory
Aurora, Colorado

2550 S. Parker Road, Suite 500
Aurora, Colorado 80014
(303)337-0990 - (303)337-3005 - Fax

December 1998

This publication is based on work sponsored wholly, or in part, by the Office of Educational Research and Improvement, Department of Education, under Contract Number RJ96006101. The content of this publication does not necessarily reflect the views of OERI or any other agency of the U.S. Government.

TABLE OF CONTENTS

CHAPTER 1

THE STANDARDS MOVEMENT

The purpose of this monograph is to describe the various ways that standards and standards-based education are being addressed throughout the country. It is certainly not an exaggeration to say that the topic of standards permeates American education. Indeed, virtually every state now has standards in core content areas or is in the process of generating those standards. More specifically, according to a 1997 report by the American Federation of Teachers (AFT) (see Gandal, 1997), 49 states either have or are in the process of setting standards. The only state that is not officially setting standards at the level of the state department of education is Iowa; however, Iowa educators are expected to establish content standards at the local district level.

Although there is a high and relatively uniform level of activity across the country around standards, there is great diversity in the approaches to standards implementation. Each approach affects the classroom teacher in different ways. Consequently, it is imperative that a school or district carefully plan for standards implementation so that the effects on the classroom are a function of design and not a function of happenstance. Three specific models of standards implementation are presented in the following chapters, along with the various ways these approaches impact classroom teachers and ultimately students.

Before presenting these approaches, we first consider a brief history of the modern standards movement.

A Brief History of the Modern Standards Movement

The education community can trace the start of the modern standards movement to the publication of *A Nation At Risk* in 1983. Researcher Laurie Shepard (1993) states that this widely read and controversial report caused a dramatic shift in the rhetoric of education reform, so that it came to embody a concern for the basic safety of our nation. It is hard to overestimate the impact of this much quoted statement from *A Nation At Risk*: "The educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a nation and a people. . . We have, in effect, been committing an act of unthinking, unilateral, educational disarmament" (National Commission on Excellence in Education, 1983, p. 5). Undoubtedly, this report caused American society to develop a deep concern for the future and quality of education in this country.

The growing concern surrounding the credibility of our educational system spurred President Bush and the state governors to gather in Charlottesville, Virginia, for the first education summit in September 1987. At this conference, they defined and agreed upon six broad goals, which were subsequently published as *The National Education Goals Report: Building a Nation of Learners* (National Education Goals Panel [NEGP], 1991). Two of these goals (3 and 4) related to specific academic standards:

> Goal 3: By the year 2000, American students will leave grades four, eight, and twelve having demonstrated competency in challenging subject matter, including English, mathematics, science, history, and geography; and every school in America will ensure that all students learn to use their minds well, so they may be prepared for responsible citizenship, further learning, and productive employment in our modern economy.
>
> Goal 4: By the year 2000, U.S. students will be first in the world in science and mathematics achievement. (p. 4)

The summit, in turn, became a catalyst for the establishment of content area standards by national subject-matter organizations. Many of those subject-matter groups turned to the National Council of Teachers of Mathematics (NCTM) for guidance because of the quality of its document *Curriculum and Evaluation Standards for School Mathematics*, published in 1989. At the present time, standards have been defined for most of the content areas taught in our nation=s schools, and many of the efforts to identify these standards were funded by the U.S. Department of Education.

Figure 1.1 contains a listing of the works produced by groups that were either funded by the U.S. Department of Education, or that identify their efforts as representative of the national consensus in their subject areas.

Although these subject-area documents were intended to stand as a de facto set of national standards, it was not long before individual states proceeded to develop their own standards documents. Some might question why the states would choose to define their own standards when "national" documents had already been created: It may very well have to do with this nation=s attitude that school policy and curricula should be regulated at the state rather than the federal level. In the words of Fred Temper, an associate superintendent in the California Department of Education, "I guess like most states we'd like to feel that we can set our own standards" (in Olson, 1995a, p. 15).

The movement for individual states to create their own standards was given substantial backing at the second education summit in Palisades, New York, in 1996. Led by President Clinton, the state governors committed to the effort of designing state standards (National Governors Association, 1996). This commitment by state governors reflects the opinions of both members of the educational community and private individuals who believe that the standards movement will either succeed or fail at the state level. As education reporter Lynn Olson (1995a) notes:

Science	National Research Council. (1996). *National Science Education Standards.* Washington, DC: National Academy Press.
Foreign Language	National Standards in Foreign Language Education Project. (1996). *Standards for Foreign Language Learning: Preparing for the 21st Century*. Lawrence, KS: Allen Press, Inc.
English Language Arts	National Council of Teachers of English and the International Reading Association. (1996). *Standards for the English Language Arts.* Urbana, IL: National Council of Teachers of English.
History	National Center for History in the Schools. (1994). *National Standards for History for Grades K-4: Expanding Children=s World in Time and Space*. Los Angeles: Author. National Center for History in the Schools. (1994). *National Standards for United States History: Exploring the American Experience*. Los Angeles: Author. National Center for History in the Schools. (1994). *National Standards for World History: Exploring Paths to the Present.* Los Angeles: Author. National Center for History in the Schools. (1996). *National Standards for History: Basic Edition.* Los Angeles: Author.
Arts	Consortium of National Arts Education Associations. (1994). *National Standards for Arts Education: What Every Young American Should Know and Be Able to Do in the Arts.* Reston, VA: Music Educators National Conference.
Health	Joint Committee on National Health Education Standards. (1995). *National Health Education Standards: Achieving Health Literacy.* Reston, VA: Association for the Advancement of Health Education.
Civics	Center for Civic Education. (1994). *National Standards for Civics and Government.* Calabasas, CA: Author.
Economics	National Council on Economic Education. (1996, August). *Content Statements for State Standards in Economics, K-12* (Draft). New York: Author.
Geography	Geography Education Standards Project. (1994). *Geography for Life: National Geography Standards.* Washington, DC: National Geographic Research and Exploration.
Physical Education	National Association for Sport and Physical Education. (1995). *Moving into the Future, National Standards for Physical Education: A Guide to Content and Assessment.* St. Louis: Mosby.
Mathematics	National Council of Teachers of Mathematics. (1989). *Curriculum and Evaluation Standards for School Mathematics.* Reston, VA: Author.
Social Studies	National Council for the Social Studies. (1994). *Expectations of Excellence: Curriculum Standards for Social Studies.* Washington, DC: Author.

Figure 1.1. AOfficial@ Standards Documents.

The U.S. Constitution makes it clear: States bear the responsibility for educating their citizens. They decide how long students continue their education and how

> the schools are financed. They control what is taught, what is tested, which textbooks are used, and how teachers are trained.
>
> Thus, despite all the talk about national education standards, it is the 50 individual states that ultimately will determine what students should know and be able to do. (p. 15)

As mentioned previously, most states either have completed or are close to completing their own educational standards, although not all of the state efforts have been favorably reviewed. To illustrate, a report published by the American Federation of Teachers in 1997 (Gandal, 1997) revealed the following:

1. The states have maintained a strong commitment to standards-based reform.

2. Most of the states still need to refine some standards in order to define the foundation of a common core of learning.

3. Overall, the states are still having difficulty setting strong English and social studies standards.

Although the standards movement has had a strong reception nationwide, some reports indicate that classroom teachers have not been dramatically affected. For example, a report by the polling organization Public Agenda indicates that, as of 1998, the majority of teachers surveyed say that they do not take student performance on standards into account when grading students (Public Agenda, 1998).

Some claim that the influence of the modern standards movement is significant but indirect. For example, researchers Robert Glasser and Robert Linn (1993) assert that the significance of the standards movement in American education may only become apparent in retrospect:

> In the recounting of our nation's drive toward educational reform, the last decade of this century will undoubtedly be identified as the time when a concentrated press for national education standards emerged. The press for standards was evidenced by the efforts of federal and state legislators, presidential and gubernatorial candidates, teacher and subject-matter specialists, councils, governmental agencies, and private foundations. (p. xiii)

The Rationale for the Standards Movement

Given the amount of energy around the modern standards movement, a logical question is what is the evidence that we need standards? While *A Nation at Risk* was certainly a clarion for the fact that public education had significant problems, it did not specifically identify standards-based education as the solution to those problems. Rather, it was only after careful scrutiny of the current system that it became clear that standards-based education held the potential of alleviating at least two major weaknesses in American education: (a) the lack of a well-articulated curriculum, and (b) an emphasis on educational "inputs" as opposed to educational "outputs." We first consider the issue of curriculum.

Lack of a Well-articulated Curriculum

The noneducator looking at the public schools would most likely conclude that they operate by a well-articulated curriculum. Indeed, most school districts can provide "curriculum guides" that list detailed explanations of what is taught not only in every subject area, but often at every grade level. This could lead one to assume that a well-honed curriculum exists within any given school district. Upon a closer examination of the curricular structure of schools, however, one frequently discovers that the image of a well-articulated course of study transitioning from one grade level to the next is little more than illusion. E. D. Hirsch, vocal critic of this nation=s educational system and author of the popular book *Cultural Literacy: What Every American Needs to Know* (Hirsch, 1987), focuses on this point in his latest book, *The Schools We Need: Why We Don't Have Them* (Hirsch, 1997):

> We know, of course, that there exists no national curriculum, but we assume, quite reasonably, that agreement has been reached locally regarding what should be taught to children at each grade level C if not within the whole district, then certainly within an individual school. . . But. . .the idea that there exists a coherent plan for teaching content within the local district, or even within the individual school, is a gravely misleading myth. (p. 26)

Hirsch continues by explaining that the idea of a coherent curriculum is a commonly held assumption. In fact, the notion of a local curriculum is held by most educators as a matter of faith. To exemplify this, Hirsch provides the following anecdote:

> Recently, a district superintendent told me that for twenty years he had mistakenly assumed each of his schools was determining what would be taught to children at each grade level, but was shocked to find that assumption entirely false; he discovered that no principal in his district could tell him what minimal content each child in a grade was expected to learn. (pp. 26-27)

Although Hirsch=s proposed solutions to the problems existing in K-12 education are not necessarily sound, he does have a point. Current research confirms that what is set forth in curriculum guides commonly does not translate into classroom procedure. For example, a number of studies (Doyle, 1992; Stodolsky, 1989; Yoon, Burstein, & Gold, n.d.) have

demonstrated that even when educators use highly structured textbooks, each teacher makes independent and idiosyncratic judgments about which points to emphasize, which to delete, and what supplementary material to add. Researchers Stevenson and Stigler (1992) illustrate this in their book *The Learning Gap*:

> Daunted by the length of most textbooks and knowing that the children's future teachers will be likely to return to the material, American teachers often omit some topics. Different topics are omitted by different teachers, thereby making it impossible for the children's later teachers to know what has been covered at earlier grades C they cannot be sure what their students know and do not know. (p. 140)

The lack of consistency in individual classroom curricula is also apparent in the research on how teachers utilize time. To illustrate, researcher David Berliner=s (1979, 1984) study of the content that teachers emphasize within reading and language arts showed that one fifth-grade teacher allocated 137 minutes a day to instruction in this area, while another only utilized 68 minutes. Likewise, at the second-grade level, one teacher devoted 47 minutes a day to reading and language arts instruction, while another set aside 118 minutes C 2 2 times more instructional time per day than the first instructor.

Finally, researcher Charles Fisher and his colleagues (Fisher et al, 1980) have provided the following anecdotes and commentary on variations in the curriculum:

> . . . in one second-grade class the average student received 9 minutes of instruction over the whole school year in the arithmetic associated with the use of money. This figure can be contrasted with classes where the average second grader was allocated 315 minutes per school year in the curriculum content area of money. As another example, in the fifth grade some classes received less than 1,000 minutes of instruction in reading comprehension for the school year (about 10 minutes per day). This figure can be contrasted with classes where the average student was allocated almost 5,000 minutes of instruction related to comprehension during the school year (about 50 minutes per day).
>
> The differences in time allocations at the level of "reading" and "mathematics" and at the level of specific subcontent areas are substantial. These differences in how teachers allocate time are related to differences in student learning. Other things being equal, the more time allocated to a content area, the higher the academic achievement. (Fisher, et al, 1980, p. 16)

In practice, American schools do not appear to have clearly delineated what should be addressed at each grade level. The intent behind the modern standards movement is that standards will compel teachers to focus on specific content at specific grade levels.

Inputs versus Outputs

Another significant issue that standards-based education is designed to address is the traditional educational focus on "inputs" versus "outputs." This problem is addressed by Chester Finn, former Assistant Secretary of Education. Finn (1990) describes the change in perspective facilitated by the standards movement as an emerging paradigm for education:

> Under the *old* conception (dare I say paradigm?), education was thought of as process and system, effort and intention, investment and hope. To improve education meant to try harder, to engage in more activity, to magnify one's plans, to give people more services, and to become more efficient in delivering them.
>
> Under the *new* definition, now struggling to be born, education is the result achieved, the learning that takes root when the process has been effective. *Only* if the process succeeds and learning occurs will we say that *education* happened. Absent evidence of such a result, there is no education C however many attempts have been made, resources deployed, or energies expended. (p. 586)

Finn explains that the deficiencies inherent in the old "input" paradigm of schooling became obvious in the mid-1960s. At that time, Congress commissioned the U.S. Office of Education to conduct a study on the quality of American education. Researcher James Coleman, chief author of the resulting and much celebrated AColeman Report,@ summarized the significance of his study as follows:

> The major virtue of the study as conceived and executed lay in the fact that it did not accept the [traditional] definition, and by refusing to do so, has had its major impact in shifting policy attention from its traditional focus on comparison of inputs (the traditional measures of school quality used by school administrators: per-pupil expenditures, class size, teacher salaries, age of building and equipment, and so on) to a focus on output, and the effectiveness of inputs for bringing about changes in output. (Coleman, 1972, pp. 149-150)

Finn explains that this report caused irreparable damage to the pre-existing "input" paradigm and began the drive toward educational outputs. Obviously, the output manifestation that is most viable to the majority of Americans is student achievement, and herein lies the connection to the standards movement.

Three Basic Approaches

As a result of our work with schools and districts across our seven-state region and the country, researchers and consultants at the Mid-continent Regional Educational Laboratory have identified a number of ways that a school, a district, or even an entire state might implement standards. These approaches can be thought of as fitting into three basic categories: (a) external tests, (b) performance tasks and portfolios, and (c) reporting on individual standards. These are briefly summarized in Figure 1.2.

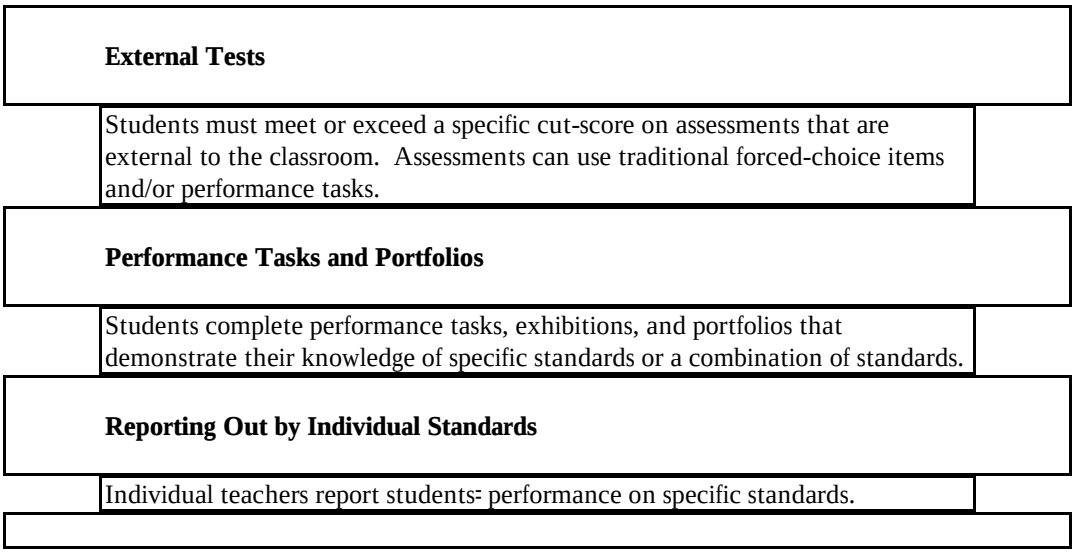


Figure 1.2. Three Approaches to Standards.

The external test approach is described in Chapter 2; the performance task and portfolio approach is described in Chapter 3; reporting out by individual standards is reported in Chapter 4. Before describing these approaches, it is important to note that they are not mutually exclusive. That is, use of one approach does imply that another cannot be used simultaneously. In short, a school, district, or state could (and perhaps should) utilize combinations of these various approaches to design a standards-based system that is specific to their needs. Finally, Chapter 5 addresses some issues that must be considered regardless of the implementation model that is employed.

CHAPTER 2

THE EXTERNAL TEST APPROACH

When a state, district, or school adheres to the external test approach as its method of implementing standards, it views the score or scores from certain tests as the foremost or only indication of whether students have met a specific standard. The use of a test as the main approach to implementing standards is often the first, and sometimes only, means considered by state departments of education. This is exemplified by President Clinton's comments to the National Governors Association at the Second Education Summit on March 27, 1996, in Palisades, New York.

> I believe every state, if you're going to have meaningful standards, must require a test for children to move, let's say, from elementary to middle school, or from middle school to high school, or to have a full-meaning high school diploma. And I don't think they should measure just minimum competency. You should measure what you expect these standards to measure. (pp. 6-7)

Traditional Tests

To most educators and noneducators the term Atest@ conjures up images of students sitting at desks responding to multiple-choice items. As we shall see, there are other types of tests students can take. However, for now we will consider this traditional form of testing as one of the options within the external test approach to standards implementation. It is helpful to conceptualize traditional exams as falling into two structural categories and three functional categories. The structural categories are norm-referenced tests (NRTs) and criterion-referenced tests (CRTs). Norm-referenced tests rate student performance against the performance of other students C usually a national sample C and most often report student performance in percentile scores. CRTs compare student performance to a pre-determined "cut-score" C the minimum number of questions a student must answer correctly in order to pass. For instance, if a student does not achieve a pre-determined minimum score on the mathematical portion of a CRT, this means that the student has not met the criterion, or cut-score, for that particular subtest. Both NRTs and CRTs make substantial use of multiple-choice items. CRTs are almost always the type of test used to implement standards for one fairly straightforward reason: These tests provide a clear pass/fail outcome.

The three functional categories for traditional tests are: off-the-shelf tests, state tests, and district tests. As indicated by the name, "off-the-shelf" refers to tests that a school or district has purchased from a testing company. They are given this name because they can be bought "off-the-shelf" much as one might purchase any other item from a store (Bond, Friedman, & van der Ploeg, 1994). Several leading companies publish this kind of test. Some of the more widely used off-the-shelf tests include:

California Achievement Tests

Comprehensive Test of Basic Skills
Iowa Test of Basic Skills
Metropolitan Achievement Tests
Sequential Tests of Educational Progress
SRA Achievement Series
Stanford Achievement Tests

(Cannell, 1988)

The results of many of these tests can be reported in either a norm-referenced or criterion-referenced manner (McMillan, 1997).

State-mandated tests make up the second functional category. Measurement expert Peter Airasian (1994) notes that this kind of test has been common in the United States since the mid-1980s. Airasian writes that "the aim of these tests is to centralize educational decision making at the state level and to prod teachers and pupils to work harder to pass the tests" (p. 369).

While state-mandated tests have emerged as the preferred testing method within the standards movement, state-level tests have recently experienced a considerable amount of criticism. This is exemplified in researcher Monty Neill=s 1997 study of state assessment systems entitled *Testing Our Children: A Report Card on State Assessment Systems*. In his report, Neill makes the following observations:

- While most states have implemented content standards, many state exams are not based on standards and many important areas in their standards are not assessed.
- By relying too heavily on multiple-choice items on tests, states fail to provide an adequate variety of methods in which students may demonstrate learning.
- States are generally weak in offering suitable performance development in interpreting and putting to use the results of state assessments. (Neill, 1997, pp. 5-6)

Neill notes that, overall, only seven states currently use assessment systems that do not require at least some substantial improvements to be acceptable as tools for making decisions about students= performance on standards. These states are Colorado, Connecticut, Kentucky, Maine, Missouri, New Hampshire, and Vermont. Neill further explains that all other state assessment systems will require significant modifications, and 15 states need "a complete overhaul" of their assessment systems. (Neill, 1997, p. 7)

Although almost every state has assessments, not all states currently require students to obtain a specific score on these tests to graduate from high school, although legislative rhetoric in the majority of states indicates this requirement may eventually be implemented. Again, Neill (1997) reports that currently 17 states require that students receive a certain score on a particular test as part of their high school graduation requirements. Two states have a test already in place, but they allow districts to use acceptable alternatives. Four states plan to implement a test in 1997,

and two states have a test, but do not require that a certain score be received before a student is allowed to graduate. Figure 2.1 summarizes information from Neill-s study of state testing practices.

	1	2	3	4		1	2	3	4
AL	X				MT				
AK			X		NC	X			
AR			X		ND				
AZ					NE				
CA					NH				
CO					NJ	X			
CT					NM	X			
DE		X	X		NV	X			
FL	X				NY	X			
GA	X				OH	X			
HI	X				OK				
IA					OR				
ID					PA				
IL					RI				
IN					SC	X			
KS					SD				
KY					TN	X			
LA	X				TX	X			
MA			X		UT				
MD	X				VA	X			
ME					VT				
MI				X	WA				
MN		X			WI				
MO					WV				X
MS	X				WY				

1 = State has a high school graduation test.
2 = State has a high school graduation test, but will accept an alternative.
3 = State plans to have a high school graduation test.
4 = State has a high school test, but it is not required for a diploma. (Adapted from Neill, 1997)

Figure 2.1. State Testing Practices.

District-designed tests comprise the third functional category of external tests. McMillan (1997) offers a warning that individual districts commonly do not possess the technical proficiency necessary to design their own tests, although many test publishers can customize reports for individual districts. As described by McMillan, "These reports indicate specific skills and may include standards that are set by the state or district" (p. 88). Often, then, the so-called district-

designed tests are actually customized versions of the off-the-shelf variety. In fact, some measurement experts recommend that districts contract with testing companies to design and construct district-level tests, as opposed to attempting such a task on their own (Airasian, 1994; McMillan, 1997).

The Problem With Traditional Tests

While traditional tests play a definite role in the implementation of standards, they do have some inherent weaknesses that educators must be made aware of if the tests are to be used effectively. Neill reports that many of the tests required by states for high school graduation rely on multiple-choice and other traditional types of items. He explains:

> Unfortunately, most states rely too heavily on multiple-choice items and fail to use a reasonable range of assessment methods. Excluding writing assessments, of the 50 states, 26 rely entirely or nearly entirely on multiple-choice. Another 16-18 rely mostly on multiple-choice (have less than half their scores derived from constructed-response items; in two states, the proportions were not clear, but appear to be around the one-half point). Only 6-8 states have less than half multiple-choice items. (Neill, 1997, p. 15)

The main flaw of an over-dependence on multiple-choice items is that this kind of test assesses only a narrow range of skills. Moreover, some studies have noted that multiple-choice tests do not take into account students' abilities to apply or think critically about knowledge (see Marzano, 1990; Marzano & Costa, 1988). The many weaknesses inherent in standardized tests have been chronicled by assessment expert Ruth Mitchell (1992):

> No matter how sophisticated the techniques, however, multiple-choice tests corrupt the teaching and learning process for the following reasons:
>
> 1. Even at their best, multiple-choice tests ask students to select a response. Selection is passive; it asks students to recognize, not to construct, an answer. The students do not contribute their own thinking to the answer.
>
> 2. Multiple-choice tests promote the false impression that a right or wrong answer is available for all questions and problems. As we know, few situations in life have a correct or incorrect answer.
>
> 3. The tests tend to rely on memorization and the recall of facts or algorithms. They do not allow students to demonstrate understanding of how algorithms work.
>
> 4. The form of multiple-choice tests means that test makers select what can easily be tested rather than what is important for students to learn.

5. Multiple-choice tests do not accurately record what students know and can do, either positively or negatively, as a personal example shows. In 1974, I passed by four points over the cut score the German- language examination to qualify for the Ph.D., and I am on record somewhere as having a reading knowledge of German. But I cannot read any word of German that does not look like an English or Latin cognate. My answers were either guesses or choices based on probabilities. If the graduate examiners had really wanted to know if I could read German, I should have been required to translate a passage.

6. The tests trivialize teaching and learning. If all classroom activity C the books, the lectures, the discussions, the exercises, the homework C ends up in a few bubbles taking no more than an hour, then what is all the fuss about? The end is incommensurate with the means. Students know that much of passing multiple-choice tests is test wisdom C how to guess productively, what items to omit C and they invest only enough effort to get by. (pp. 15-16)

Performance Assessments

To counteract the weaknesses existing in tests that rely on multiple-choice items, many state assessment systems are now utilizing performance tasks. These tasks require that students think through and construct their own responses rather than select from a number of pre-determined options, as is the case with multiple-choice items. (We will consider performance tasks in greater depth in Chapter 3.) Because of the processes required to work through them, performance tasks are often called "constructed-response" items. The following example is a mathematics task designed by the National Assessment of Educational Progress (NAEP, 1992) and administered to a nationwide representative sample of eighth graders:

> Treena won a 7-day scholarship worth $1,000 to the Pro Shot Basketball Camp. Round-trip travel expenses to the camp are $335 by air or $125 by train. At the camp she must choose between a week of individual instruction at $60 per day or a week of group instruction at $40 per day. Treena's food and other expenses are fixed at $45 per day. If she does not plan to spend any money other than the scholarship, what are all choices of travel and instruction plans that she could afford to make? Explain your reasoning. (Dossey, Mullis, & Jones, 1993, p. 116)

Researchers John Dossey, Ina Mullis, and Chancey Jones (1993), describe the types of thinking and reasoning required to complete this task:

> The solution to this task requires students to use everyday consumer sense to determine Treena's fixed expenses and analyze the various choices she has for travel (plane or train) and instruction (individual or group). Students also must compare the total cost for each of the four alternatives to which this analysis leads

> to the $1,000 value of Treena's scholarship, in order to conclude which choices meet the given conditions. (p. 116)

One of the aspects that makes performance tasks such a powerful assessment tool is that often they require students to explain their reasoning. This provides some insight into the logic systems used by the students in developing their responses C information that cannot easily be assessed through multiple-choice items. This quality has generated overwhelming support for the use of performance tasks as supplements to traditional tests or as viable alternatives to traditional tests (see Archbald & Newmann, 1988; Baron, 1991; Baron & Kallick, 1985; Berk, 1986a, 1986b; Frederiksen & Collins, 1989; Marzano, 1990; Marzano & Costa, 1988; Mitchell, 1992; Resnick, 1987a, 1987b; Resnick & Resnick, 1992; Shepard, 1989; Stiggins, 1994; Wiggins, 1989, 1991, 1993a, 1993b; Winograd & Perkins, 1996).

External Tests and the Classroom Teacher

Regardless of whether an external test is developed by a district, a state department, or a national publishing company, and regardless of whether it contains multiple-choice items, performance tasks, or both, it is essential that classroom teachers be made aware of the exact content covered in that test so that the content might be included as a routine part of classroom work. In other words, classroom teachers should directly cover the content included in external tests. Unfortunately, this idea is contrary to a misguided principle of the modern American educational system that "it is unethical to teach to a test." Actually, one of the main principles of the modern standards movement is that teachers should, in fact, teach to tests (Wiggins, 1989). The question, then, is how did the idea arise that such a practice is unethical?

Researcher Grant Wiggins explains that this misconception is rooted in an unfounded assumption that effective assessment requires a component of secrecy. In his book *Assessing Student Performance: Exploring the Purpose and Limits of Testing* (1993b), Wiggins states, "This assumption is so common that we barely give it a second thought; the tests that we and others design to evaluate the success of student learning invariably depend upon secrecy. (p. 72)

The assertion that test content must be kept hidden from students is flawed from at least two perspectives. First, it defies an individual=s basic right to know the criteria upon which he or she will be judged, especially if the resulting judgments will have high stakes implications for that individual. Wiggins firmly supports the right of students to have prior knowledge of how and over what they will be tested.

> Why would we take for granted that students do not have a right to full knowledge and justification of the form and content of each test and the standards by which their work will be judged? The student's (and often the teacher's) future is at stake, yet neither has an opportunity to question the aptness or adequacy of the test, the keying of the answers, or the scoring of the answers. Why would we assume that any test designer C be it a company or a classroom teacher C has a prior right to keep such information from test takers (and often test users)? Why

> would we assume, contrary to all accepted guidelines of experimental research, that test companies (and teachers) need not publish their tests and results after the fact for scrutiny by independent experts as well as the test taker? Maybe the better advice to test makers is that offered twenty years ago by performance assessment researchers Robert Fitzpatrick and Edward Morrison: "The best solution to the problem of test security is to keep no secrets." (p. 73)

The idea that test content should be kept secret is also false from the perspective that test secrecy makes sense only if a given test adequately represents a student's abilities in the subject area being tested. If this were the case, test secrecy would have the potential of being fair, since competence in a subject-matter would ensure a high score on the test. Current research, however, has shown that this is most often not the case.

Over the last two decades, a relatively new area of measurement theory referred to as "generalizability theory" has been used to analyze educational testing practices. Although it is complex in practice, this theory is designed to find how generalizable a student's score on a particular test is to his overall competence in the subject matter tested (Brennan, 1983; Feldt & Brennan, 1993). Research shows that while an individual might receive a high score in a subject when tested in one manner, he could receive a low score when the same subject matter is tested in a different fashion. To illustrate, in a series of studies by Richard Shavelson and colleagues (Shavelson & Baxter, 1992; Shavelson, Gao & Baxter, 1993; Shavelson & Webb, 1991; Shavelson, Webb, & Rowley, 1989), students were given the same science test three times with each version using a different format (i.e., hands on, computer simulated, and descriptions written after a hands-on experiment). The researchers found that a student might do quite well on a test using one format, but perform poorly when tested using the other two formats. It should be emphasized that Shavelson=s research did not address multiple-choice items, but focused mainly on performance tasks. The results of these studies have caused Shavelson and his colleagues to conclude that an individual performance test is not an accurate indicator of how proficiently a student can perform in a particular content area. In fact, measurement experts (e.g., Lane, Liu, Ankenmann, & Stone, 1996; Linn, 1994; Shavelson, Gao, & Baxter, 1993) now assert that it takes between 10 and 36 performance tasks to accurately assess a student within a single content area. Therefore, a student=s score on one performance task is not a generalizable indicator of that student=s competence in the subject area assessed by that task.

The lack of generalizability of performance tasks has produced strange anomalies regarding the competence of individual students. To illustrate this point, consider the following story, which appeared in the Wall Street Journal, regarding a high school senior who had completed a performance task designed by the local state department of education:

> Jonathan, 17 years old, was declared a novice at writing. But by that time, he had already been accepted at the Massachusetts Institute of Technology, California Institute of Technology, Carnegie Mellon University, Rensselaer Polytechnic Institute and the University of Illinois. He had earned a perfect score on the American College Testing, or ACT, exam, which Mid-western states favor over

the Scholastic Aptitude Test; he had scored a near-perfect 770 out of 800 on the verbal portion of the SAT; he had accumulated a 3.993 grade point average; he was a National Merit Scholar, had a perfect grade in advance-placement English, and was on his way to graduating at the head of his class. (Wall Street Journal, March, 1997)

Tests which rely heavily on multiple-choice items are also at risk of not being generalizable, especially if fewer than six items are used to assess a topic (see Linn & Gronlund, 1995, p. 442). Unfortunately, a close look at tests using multiple-choice items reveals that many topics are assessed with only a few items. An example of this is seen in Figure 2.2, which is an analysis of the number of items used to assess certain topics in a popular standardized test.

As Figure 2.2 illustrates, the number of items assessing a particular topic ranges between 3 and 11. One might well ask if a student's score on a subtest designed to assess her understanding of health and safety using only three items is a generalizable indicator of her knowledge of the subject. Probably not! It should also be noted that the test from which the item count in Figure 2.2 was taken is one of the most popular standardized tests currently available. This does not mean that this particular test is flawed. In fact, it is quite difficult to design a single test that provides a generalizable measure of a student=s competence in a subject area.

	Content Area	**# of Items**
Word Analysis	Initial Sounds: words	5
	Letter Substitutions	5
	Word Building: vowels	5
	Vowel Sounds	11
	Silent Letters	3
	Affixes	3
Math Concepts	Number Systems	4
	Whole Numbers	3
	Geometry	5
	Measurement	6
	Fractions and Money	4
	Number Sentences	6
	Estimation	3
Social Studies	History	4
	Geography	6
	Economics	8
	Political Science	6
	Sociology and Anthropology	7
Science	Nature of Science	9
	Life Science	6
	Earth and Space	5
	Physical Science	8
	Health and Safety	3

Figure 2.2. Items Per Topic on a Standardized Test.

The research on generalizability makes it clear that external tests very often address highly specific information. As a result, a teacher must be acutely aware of the content and format of the items appearing on these tests and must be sure to cover this information in the classroom. Tests developed by testing companies commonly aid classroom teachers in this endeavor by providing fairly detailed descriptions of what content is assessed in their subtests. For example, the publishing company Harcourt, Brace, Jovanovich will supply, upon request, information regarding the content of the *Stanford Achievement Tests*. For the intermediate level mathematical computation subtests, they report that the following topics will be assessed:

- addition with whole numbers
- subtraction with whole numbers
- multiplication with whole numbers
- division with whole numbers
- computation with fractions and decimals
- estimates

Unfortunately, the publishers do not provide as much detail regarding the content covered in other areas such as science or the social sciences. (For a detailed discussion of the Stanford Achievement Tests, see Airasian, 1994.) Likewise, the Riverside Publishing Company offers the following information regarding the *Sources of Information* subtest within the Iowa Test of Basic Skills, Level 8:

- locating specific places on maps
- determining directions on maps
- determining distances on maps

They do not, however, provide this much detail regarding some of their other subtests such as social studies. (For a detailed discussion of the Iowa Test of Basic Skills see McMillan, 1997.)

Unfortunately, as yet, this level of information does not appear to be readily available for tests developed by state departments of education. Research for this monograph has uncovered little in the way of specific guidance regarding the content covered in state department-designed tests, even though those state departments have contracted with national testing companies to develop their tests. In fairness, though, the current lack of success in identifying the specific content on state-level standards tests might be due to the fact that only a small sample of state departments have been contacted. In addition, some states that were contacted were still designing their tests and might not have felt ready to provide the requested information. At any rate, state departments and districts that have developed their own standards tests should make available the specific content addressed in those tests. It is very important that classroom teachers approach the local agencies that have developed the tests and press for detailed information regarding the content covered in them.

CHAPTER 3

PERFORMANCE TASKS AND PORTFOLIOS

The second means of implementing standards uses both performance tasks and student portfolios that are **developed over time** as the primary method of assessing students' competence in standards. In this approach, students might have to complete a performance task illustrating their knowledge of mathematics to meet specific mathematics standards. Before considering this approach in depth, it is important to describe in more detail the nature of performance assessments.

Performance Tasks

Classroom teachers often use the terms *performance assessment* and *authentic assessment* interchangeably, while some educators assert that there is actually a difference between the two. Evaluation specialist Carol Meyer (1992) defines performance assessment as a situation in which students must construct responses to demonstrate that they can apply learning. Authentic assessments also require students to construct responses to show the application of knowledge, but the given situation is more "real life." Researchers Fred Newmann, Walter Secado, and Gary Wehlage (1995) provide the examples of authentic tasks in geometry and social studies shown in Figure 3.1.

Authentic Geometry Task

Design packaging that will hold 576 cans of Campbell's Tomato Soup (net weight, 10 3 oz.) or packaging that will hold 144 boxes of Kellogg's Rice Krispies (net weight, 19 oz.). Use and list each individual package's real measurements; create scale drawings of front, top, and side perspectives; show the unfolded boxes/containers in a scale drawing; build a proportional, three-dimensional model.

Authentic Social Studies Task

Write a letter to a student living in South Central Los Angeles conveying your feeling about what happened in that area following the acquittal of police officers in the Rodney King case. Discuss the tension between our natural impulse to strike back at social injustice and the principles of nonviolence.

Note: From *A Guide to Authentic Instruction and Assessment: Vision, Standards and Scoring* (pp. 24-25), by F. M. Newmann, W. G. Secado, G. G. & Wehlage, G. G. 1995, Madison, WI: Wisconsin Center for Educational Research, University of Wisconsin.

Figure 3.1. Sample Authentic Assessments.

The line between authentic tasks and performance tasks is decidedly unclear. Is the mathematical task described in Chapter 2 about Treena's scholarship to the Pro Shot Basketball camp a performance task or an authentic task? Is it a situation that students might encounter in

real life? Could the task in Figure 3.1, which asks students to design packages to hold differing quantities of various products, be a real-world problem? In the final analysis, performance tasks and authentic tasks are so much alike in actual practice that the difference between them is negligible. As a result, in this monograph, the term performance task is used to signify any task in which students must apply knowledge and defend their reasoning regardless of whether it is a situation that might occur in "real life."

One of the most fascinating qualities of performance tasks is that, as students practice these tasks, their ability to do them can increase dramatically. To illustrate, studies conducted at McREL have shown that student ability to do performance tasks can be improved if teachers make systematic use of such tasks in the classroom. In one elementary school, for example, McREL researchers gave mathematics performance tasks to all first-grade through fifth-grade students in September. Two skills were assessed in each of these tasks: problem-solving ability and the ability to communicate mathematically. The percentage of students who received a score that was satisfactory or better than satisfactory for these skills is reported in columns A and C of Figure 3.2.

Ethnicity	**A** **Pretest Problem Solving**	**B** **Post-test Problem Solving**	**C** **Pretest Communication**	**D** **Post-test Communication**
Asian (25)	16.0% (4)	68.0% (17)	12.0% (3)	44.0% (11)
African American (130)	8.5% (11)	50.8% (66)	12.3% (16)	32.3% (42)
Hispanic (31)	3.2% (1)	77.4% (24)	0% (0)	48.4% (15)
White (116)	29.3% (34)	78.4% (92)	24.1% (28)	62.9% (73)
Grand Total (302)	16.6%	65.6%	15.6%	46.7 %

Note: From R. J. Marzano and J. S. Kendall, 1992, unpublished data, Aurora, CO: Mid-continent Regional Educational Laboratory. Copyright 8 1992 by McREL. Reprinted with permission.

Figure 3.2. Pretest and Post-test Results on Performance Tasks.

That September, 16.6% of the students developed satisfactory or better responses in the problem-solving portion of the tasks, and 15.6% in the communication portion. Quite naturally, the teachers participating in the study wanted a higher percentage of their students to receive satisfactory scores. Consequently, over the course of that school year, the teachers gave their students performance tasks of their own design and offered at their own pace. The teachers made a concentrated effort to continually ask the students to explain what they did as they attempted these tasks.

At the end of the year, the students were given another performance task, which again was assessed for problem-solving and communication. These post-test scores, listed in columns B and D in Figure 3.2, showed a dramatic improvement in the percentage of satisfactory or better scores. The percentage of students receiving at least satisfactory scores in problem-solving rose from 16.6% in September to 65.6%, and the percentage in communicating mathematically jumped from 15.6% to 46.7%. Even more impressive were the gains reported in the performance of Hispanic and African-American students. For instance, 50% of African-American students received a satisfactory or better score for problem-solving on the post-test, as opposed to 8.5% who had done so in September. The results of this study indicate that one can expect a student=s ability on performance tasks to increase if teachers systematically focus on these tasks in the classroom.

This does not suggest that teachers should drill students on the precise performance tasks that will appear on their standards test, whether that test be required at the district or state level. Instead, students should be given practice in performing tasks similar to the ones that might appear on an external standards test. Essentially, this allows students to receive practice in the general skills common to all performance tasks rather than on the specific content of any one task. In the following section entitled "Performance Tasks, Portfolios, and the Classroom Teacher," these general skills will be explained in detail.

Portfolios

Performance tasks and portfolios are symbiotically linked because a portfolio is often a collection of performance tasks. Researchers Lauren Resnick and Daniel Resnick (1992) define portfolios as follows:

> A variant of the performance assessment is the *portfolio assessment*. In this method, frequently used in the visual and performing arts and other design fields, individuals collect their work over a period of time, select a sample of the collection that they think best represents their capabilities, and submit this portfolio of work to a jury or panel of judges. (p. 61)

Likewise, researcher Mark Reckase (1995) defines a portfolio as "a purposeful collection of student work that exhibits to the student (and/or others) the student's efforts, progress, or achievement in (a) given area(s)." (p. 21)

By their basic design, portfolios lend themselves most naturally to subjects that involve products like writing and the arts, although lately there have been efforts to define the recommended portfolio contents for subject areas that are not necessarily product oriented. For example, mathematics teacher Pam Knight (1992) has constructed the following list of items that might be included in a mathematics portfolio:

- samples of word problems in various stages of development, along with the student's description of his or her thinking during the various problem-solving stages;
- the student's self-evaluation of his or her understanding of the mathematical concepts that have been covered in class, along with examples;
- the student's self-evaluation of his or her competence in the mathematical procedures, strategies, and algorithms that have been covered in class along with examples.

When performance tasks and portfolios are used as the main evidence of a student=s competency in district standards, the student is required to complete a number of performance tasks that are then organized into a portfolio. A useful comparison is the difference between external tests (discussed in Chapter 2) and this method. In the first approach, students must "pass" an external assessment of some kind to graduate from one level to another. In this sense, the external tests and the performance task and portfolio approach are similar because both require a form of "exit assessment" before a student is permitted to move to the next level. As we have already discussed, both might also involve performance tasks. With the external test approach, however, the exit assessment occurs at a pre-determined time. Basically, the student must take and pass a test in order to move from one level to the next. Although that test may include or even be entirely composed of performance tasks, it is still a test that is given in a relatively short time period and at a specific point (e.g., a few hours on a certain day or series of days set aside for the administration of the external tests). The main difference between the performance task approach to external tests and the use of performance tasks and portfolios, as described in this chapter, is that the latter method occurs over a much longer period of time C possibly even years before the student is ready to actually present her portfolio of performance tasks.

The Popularity of Performance Tasks and Portfolios

The performance task and portfolio model of standards implementation is probably the most popular approach currently in use. In fact, much of the literature on standards-based education assumes the use of this method. For instance, this approach is emphasized by researcher Joseph McDonald and his colleagues (McDonald, Smith, Turner, Finney, & Barton, 1993) in their discussion of the innovations brought about as a result of Theodore Sizer's Coalition of Essential Schools. The Coalition was born of the studies carried out by Sizer and his colleagues between 1979 and 1984 (Sizer, 1985; Powell, Farrar, & Cohen, 1985; Hampel, 1986). One of the basic tenets of Coalition philosophy is that students should not be awarded diplomas until they have demonstrated their competence. This principle is manifested as an emphasis on performance tasks and portfolios in many Coalition schools.

Central Park East Secondary School (CPESS) is one of the most often-cited examples of the performance task and portfolio approach. Researchers Linda Darling-Hammond and Jacqueline Ancess (1994) report that CPESS, located in East Harlem, enrolls about 500 students between grades 7 and 12. Approximately 85 percent of these individuals are from Latino and African-American families. Sixty percent of the students at the school qualify for free or reduced-price lunch, and twenty-five percent are eligible for special education programs. The foundation of the approach used at CPESS is the completion of 14 projects organized into a portfolio. These are:

1. A postgraduate plan
2. An autobiography
3. A report on school/community internship
4. A demonstration of an awareness of ethics and social issues
5. A demonstration of an appreciation of fine arts and ethics
6. A demonstration of an awareness of mass media
7. A demonstration of the importance and utility of "practical" skill areas such as medical care, independent living, legal rights, and securing a driver's license
8. A demonstration of an understanding of geography
9. A demonstration of competence to work in a language other than English, as a speaker, listener, reader, or writer
10. A demonstration of facility with the scientific method
11. A demonstration of competence in mathematics
12. A demonstration of an understanding and appreciation of a broad array of literature
13. A demonstration of an understanding of history and how it affects our lives today
14. A demonstration of participating in any team or individual, competitive or non-competitive sport or activity (pp. 14-15)

Darling-Hammond and Ancess report that there is no single way to complete or present these projects. In fact, a student might use one performance task to fulfill two or more of the 14 topics

described above. To exemplify this, Darling-Hammond and Ancess offer an account of how Marlena, a student at CPESS, approached the required 14 topics. She completed one portfolio for three separate science internships that she took at Brookhaven National Laboratory, Hunter College, and Columbia University over a two-year period. Marlena developed another mathematics portfolio that included mathematical models of rainfall created under differing assumptions. Yet another portfolio was built for her media project, which contained a sophisticated, evidence- based analysis of race, gender, and class stereotyping occurring in prime-time television. Marlena=s history portfolio followed the chronology of segregated education in the United States, which she then applied to contemporary debates about Afrocentric schools. This project also served as her entry for ethics and social issues (see #4 above).

Teachers at CPESS evaluate each portfolio according to a 20-point scoring grid, which is subsequently translated into a more qualitative descriptive scale: distinguished (18-20), satisfactory (15-17), and minimally satisfactory (12-14). Projects receiving a score lower than twelve must be resubmitted.

Project scores are recorded in a special section of a student=s transcript. Figure 3.3 depicts the portfolio section of Marlena's transcripts.

Performance Tasks, Portfolios, and the Classroom Teacher

When a school or district implements standards through the use of performance tasks and portfolios, it often allows the students a great deal of choice regarding the exact subject-matter content that will be covered. CPESS, for instance, specifies only general subject areas in its list of 14 required topics. As a result, the classroom teacher=s role becomes that of guide and coach as students design their performance tasks. A teacher interacting with Marlena, for instance, might have helped her identify the specific geography content (see #8) on which she desired to focus her studies, the genres and exact titles of literature (see #12) that Marlena wished to explore and analyze, and so on. The teacher, using the performance task and portfolio approach, then becomes much more a resource and guide to the students than a provider of content.

Central Park East Secondary School
1573 Madison Avenue, New York, NY 10029 (212) 860-8936

TRANSCRIPT OF PORTFOLIOS

Please refer to the Curriculum Bulletin for Portfolio requirements. A Portfolio is graded on the basis of all items within it as well as knowledge and skill defended before the student=s Graduation Committee. Listed below is the title of the student=s major work in each area as well as the cumulative grade. Individual portfolio items are available on request.

Dist = Distinguished work Sat = Satisfactorily met requirements
MinSat = Minimally met requirements FP = Final project (in-depth study)

THE PORTFOLIO

	Grade	Date (completed project)
Post Graduate Plan	Sat	12-90
Autobiography	Sat	12-13-91
Practical Skills & Knowledge (Life Skills)	Dist	3-1-92
Internship (Brookhaven National Lab and Hunter College, NY)	Dist	1-3-91
Ethics, Social Issues, & Philosophy (Controversy of Afrocentric schools)	Dist	2-28-92
Literature (Influences on Malcolm X=s life)	Sat +	3-92
History (Events affecting the controversy of Afrocentric schools)	Sat	2-28-92
Geography (Geography of the West Indies)	Sat +	6-5-92
Language other than English (Spanish: English only vs. dual language)	Sat +	1-3-92
Mathematics (Mathematical models: lines and sines)	Dist	3-16-92
Science & Technology (Construction of Expression vectors with Phosphatases 1 & 2A)	Dist	4-92
Fine Arts & Aesthetics (Opera: ADie Fliedermaus@ and AThe Marriage of Figaro@)	Sat	12-13-91
Mass Media (Entertainment or News? Our Children=s Education)	Sat +	2-24-91
Physical Challenge (Aerobics)	MinSat	6-17-92
Review Date		

Note: From *Graduation by Portfolio at Central Park East Secondary School* (p. 20), by L. Darling-Hammond and J. Ancess, 1994, New York: National Center for Restructuring Education, Schools, and Teaching, Columbia University. Reprinted with permission from the National Center for Restructuring Education, Schools, and Teaching (NCREST).

Figure 3.3. Portion of Transcript Devoted to Projects at Central Park East Secondary School.

In addition to fulfilling these roles as the students design their performance tasks and portfolios, the teacher is also responsible for teaching and reinforcing three basic skill areas that are common components of most performance tasks. These components are illustrated in the following performance task designed for use in a history class:

> For the next two weeks we will be studying American military conflicts of the past three decades, in particular the Vietnam War. You will form teams of two and pretend that you and your partner will be featured in a news magazine television special about military conflict. Your team has been asked to help viewers understand the basic elements of the Vietnam War by relating them to a situation that has nothing to do with military conflict but has the same basic elements. You are free to choose any nonmilitary situation you wish. In your explanation, the two of you must describe how the nonmilitary conflict fits each of the basic elements you identified in the war. You will prepare a report, with appropriate visuals, to present to the class in the way you would actually present it if you were doing your feature on the news magazine special. You will be assessed on and provided rubrics for the following:

1. Your understanding of the specific details of the Vietnam War.

2. Your ability to identify the similarities and differences between the Vietnam War and the nonmilitary conflict you selected.

3. Your ability to design and deliver a report.

4. Your ability to work as an effective member of a team.

As outlined by the directions to the students, this task is intended to assess four areas, only the first of which addresses actual history content. The remaining three deal with areas which are almost always inherent in a performance task: (1) thinking and reasoning, (2) communication, and (3) lifelong learning. These are the general skills mentioned in the preceding section that are frequently embedded in performance tasks. It stands to reason that by offering students practice in these skills, teachers can enhance students= performance in many, if not most, of the performance tasks they will encounter later.

1. Thinking and Reasoning

More than 80 years ago, educational philosopher John Dewey (1916) wrote that "the sole direct path to enduring improvement in the methods of instruction and learning consists in centering upon the conditions which exact, promote, and test thinking" (p. 6). In the past few decades, the National Science Board Commission on Precollege Education in Mathematics, Science and Technology (1983), the College Board (1983), and the National Education Association (Futrell, 1987) have put forth a call for the enhancement of thinking and reasoning in American education. This need to further students' abilities to think and reason is also clearly set forth under Goal 3 of the six national education goals articulated at the first education summit in Charlottesville, Virginia. (See Chapter 1 for a discussion of the National Goals.) As previously stated, Goal 3 specifically targeted the areas of English, mathematics, history, science, and geography. Furthermore, it declared that "every school in America will ensure that all students learn to use their minds well so they may be prepared for responsible citizenship, further learning, and productive employment in our modern economy." (NEGP, 1991, p. ix)

There are several methods that a classroom teacher might use to help develop students= thinking and reasoning skills. For example, Quellmalz (1987) has identified the following four categories of thinking and reasoning that can easily be adapted to classroom instruction: analyzing, comparing, inferring, and evaluating. Perkins (1992) has identified seven areas of reasoning: explaining, exemplifying, applying, justifying, comparing and contrasting, contextualizing, and generalizing. Marzano and his colleagues (Marzano, 1992; Marzano, Pickering, Arredondo, Blackburn, Brandt, & Moffett, 1992) have identified fifteen individual areas of thinking and reasoning.

Even though researchers in education and psychology agree on the importance of teaching thinking and reasoning, there is not agreement about the exact list of specific thinking and

reasoning skills, as the examples above illustrate. Recently, researchers at McREL attempted to identify a definitive list of thinking and reasoning skills. They assumed that if general thinking and reasoning skills do, in fact, exist, they should be found in the national standards documents. To illustrate, if the standards documents in mathematics, science, and history all mention the thinking and reasoning skill of problem-solving as important, then one might conclude that problem-solving is a general thinking and reasoning skill that cuts across these subject areas. The national standards documents, therefore, represent a source from which general thinking and reasoning skills can be gleaned if they, in fact, exist.

To study thinking and reasoning in the national documents, McREL researchers focused their attention on the following:

1. Science:
   - *Benchmarks for Science Literacy* (Project 2061, 1993)
   - *National Science Education Standards* (National Research Council, 1996)

2. Mathematics:
   - *Curriculum and Evaluation Standards for School Mathematics* (National Council of Teachers of Mathematics, 1989)

3. Social Studies:
   - *Expectations of Excellence: Curriculum Standards for Social Studies* (National Council for the Social Studies, 1994)

4. Geography:
   - *Geography for Life: National Geography Standards* (Geography Education Standards Project, 1994)

5. History:
   - *National Standards for History: Basic Edition* (National Center for History in the Schools, 1996)

6. Civics:
   - *National Standards for Civics and Government* (Center for Civic Education, 1994)

7. Physical Education:
   - *Moving into the Future, National Standards for Physical Education: A Guide to Content and Assessment* (National Association for Sport and Physical Education, 1995)

8. Health:
   - *National Health Education Standards: Achieving Health Literacy* (Joint Committee on National Health Education Standards, 1995)

9. The Arts:
   - *National Standards for Arts Education: What Every Young American Should Know and Be Able to Do in the Arts* (Consortium of National Arts Education Association, 1994)

10. Foreign Language:
    - *Standards for Foreign Language Learning: Preparing for the 21st Century* (National Standards in Foreign Language Education Project, 1996)

11. The English Language Arts:
    - *Standards in Practice: Grades K-2* (Crafton, 1996)
    - *Standards in Practice: Grades 3-5* (Sierra-Perry, 1996)
    - *Standards in Practice: Grades 6-8* (Wilhelm, 1996)
    - *Standards in Practice: Grades 9-12* (Smagorinski, 1996)

12. The World of Work:
    - *What Work Requires of Schools: A SCANS Report for America 2000* (The Secretary=s Commission on Achieving Necessary Skills, 1991)
    - *Workplace Basics: The Essential Skills Employers Want* (Carnevale, Gainer & Meltzer, 1990)

Note that for a few subject areas multiple documents were used. Specifically, in science, the ?official@ standards document is certainly the *National Science Education Standards* published by the National Research Council. However, *Benchmarks for Science Literacy* was also analyzed because of its wide acceptance as a reference document in the field of science.

Four grade-interval-specific documents were analyzed for the English language arts, as opposed to the more general document published by the National Council of Teachers of English (NCTE) and the International Reading Association (IRA), entitled *Standards for the English Language Arts* (1996). This was done at the recommendation of NCTE (Myers, 1997), since the more specific documents were designed to articulate benchmarked skills and abilities, and the general document was not.

Areas 1 through 11 above are subject matters traditionally considered basic by most state departments of education, as evidenced by the fact that most states have or are in the process of identifying standards in these specific areas or combinations of these areas (e.g., civics, history, and geography might be combined into one subject area).

In addition to the documents that address the core subject areas, two documents that were analyzed reflected what the "world of work" (e.g., employers) considers important skills to be enhanced in K-12 education (see 12 above).

All documents were analyzed for thinking and reasoning skills that were stated explicitly and implicitly. (For a detailed discussion of the protocols used in the analysis, see Kendall and Marzano, 1997.) In all, McREL identified six general thinking and reasoning skills that are mentioned in a majority of the subject areas. They are listed below, along with the percentage of subject areas in which they are cited.

1. Utilizes mental processes that are based on identifying similarities and differences (100%)

2. Applies problem-solving and troubleshooting techniques (83%)

3. Understands and applies basic principles of argumentation (83%)

4. Applies decision-making techniques (75%)

5. Understands and applies basic principles of hypothesis testing and scientific inquiry (58%)

6. Understands and applies basic principles of logic and reasoning (50%)

As indicated above, the extent to which the thinking and reasoning processes were addressed across the twelve subject areas ranged from a high of 100% to a low of 50%. For each of these thinking and reasoning skill areas, specific understandings and abilities were identified in each of the four grade-level intervals: Level 1: grades K-1; Level II: grades 3-5; Level III: grades 6-8; Level IV: grades 9-12. To illustrate, Figure 3.4 presents the Level III benchmarks for the thinking and reasoning skill of "under-standing and applying basic principles of hypothesis testing and scientific inquiry."

(21,233;NHI,66;NSI,145)

1. Understands that there are a variety of ways people can form hypotheses, including basing them on many observations, basing them on few observations, and constructing them on only one or two observations

(MI,75;NSI,148,171)

2. Verifies results of experiments

(2E,299;NHI,66:NSI,145)

3. Understands that there may be more than one valid way to interpret a set of findings

(2E,299;NSI,171)

4. Questions findings in which no mention is made of whether the control group is very similar to the experimental group

(SSE,149;NSE,145;NSI,171)

5. Reformulates a new hypothesis for study after an old hypothesis has been eliminated

(MI,78,81,143;NSI,145,171)

6. Makes and validates conjectures about outcomes of specific alternatives or events regarding an experiment

Figure 3.4. Level III (Grades 6-8) Benchmarks for Hypothesis Testing and Scientific Inquiry.

Note that each benchmark in Figure 3.4 is accompanied by a detailed code called a "citation log." (These benchmarks in their entirety are available on the World Wide Web C Uniform Resource Locator: www.mcrel.org/standard.html.)

This level of detailed analysis allowed McREL researchers to answer the question, To what extent do different subject areas place emphasis on these various thinking and reasoning skills? Figure 3.5 presents the results of our analysis of the percentage of citations within the twelve subject areas that were devoted to the specific thinking and reasoning skills.

	Similarities & Differences	Problem Solving	Argumentation	Decision Making	Hypothesis Testing	Logi
ience (27.2%)	8.3%	11.5%	22.9%	3.1%	32.3%	21.8%
istory (13.0%)	32.6%	26.1%	15.2%	15.2%	8.7%	2.2%
athematics (11.3%)	17.5%	50.0%	7.5%	0.0%	20.0%	5.0%
cial Studies (9.1%)	28.1%	6.3%	28.1%	28.1%	3.1%	6.3%
ıe Arts (8.2%)	46.4%	32.1%	7.1%	14.3%	0.0%	0.0%
vics (7.4%)	23.1%	38.5%	7.7%	30.8%	0.0%	0.0%
ork (6.8%)	12.5%	54.2%	20.8%	0.0%	4.2%	8.3%
reign Language (4.0%)	92.9%	0.0%	7.1%	0.0%	0.0%	0.0%
eography (3.7%)	30.8%	7.7%	7.7%	30.8%	23.1%	0.0%
ealth (3.7%)	46.2%	15.4%	7.7%	30.8%	0.0%	0.0%
ysical Education .1%)	45.5%	18.2%	0.0%	36.4%	0.0%	0.0%
nguage Arts (2.8%)	50.0%	0.0%	0.0%	10.0%	20.0%	20.0%

rinted by permission of the Mid-continent Regional Educational Laboratory.

ıre 3.5. Subject Matter Emphases on Various Critical Thinking and Reasoning Skills.

The twelve subject areas are listed in the first column of Figure 3.5 ranked by the percentage of total references to thinking and reasoning attributed to that subject area. For example, of all references to thinking and reasoning in all the documents analyzed, the science documents accounted for 27.2%, history, 13%, mathematics, 11.3%, and so on. It is probably advisable not to place too much importance on these percentages, since documents differed in length, and in some subject areas more than one document was analyzed. Science is an example where two documents were used. However, it is interesting to note some strong patterns. Language Arts, which had the most documents C four, accounted for the lowest percentage of references to thinking and reasoning. Additionally, three subject areas (science, history, and mathematics) accounted for over half (51.5%) of all references made to thinking and reasoning.

The more defensible inferences from Figure 3.5 can be made by studying the patterns of emphasis within individual subject areas. This can be done by analyzing the percentages of references for each of the six thinking and reasoning skill areas within specific subjects. For example, one can infer that science places major emphasis on hypothesis testing and scientific inquiry since 32.3% of its references to thinking and reasoning were specific to this one skill area. Additionally, science places heavy emphasis on argumentation and the use of logic, and some emphasis on problem solving and identifying similarities and differences. It places relatively minor emphasis on decision-making.

Using the patterns of emphasis depicted in Figure 3.5, one might form the following conclusions about the six thinking and reasoning skill areas:

1. Identifying similarities and differences should receive some attention in all subject areas and be stressed in history, social studies, the arts, foreign language, geography, health, physical education, and the language arts.

2. Problem solving should occur in all subject areas except for foreign language and the language arts, and be stressed in science, social studies, and subjects that address the world of work.

3. Argumentation should receive some attention in all subject areas except physical education and the language arts, and be stressed in science, social studies, and subjects that address the world of work.

4. Decision making should receive some attention in science, history, the arts, and language arts. It should be stressed in social studies, civics, geography, health, and physical education.

5. Hypothesis testing and scientific inquiry should receive some attention in social studies and subjects that address the world of work. It should be stressed in science, mathematics, geography, and the language arts.

6. Logic should receive some attention in history, mathematics, social studies, and subjects that address the world of work. It should be stressed in science and the language arts.

These conclusions can provide general guidance for future curriculum design. Using these findings and conclusions, educators can infuse the teaching of thinking and reasoning into virtually all subject areas in a systematic fashion that is consistent with the basic structure of those subject areas.

Regardless of whether a teacher favors the thinking and reasoning skills identified in the McREL study or those identified by Quellmalz, Perkins, and others, it is imperative that the students receive precise instruction in the steps involved in specific thinking and reasoning processes; otherwise, there is a good chance that they will not understand what they are expected to do when asked to utilize a particular process. This was dramatically demonstrated in a study by the National Assessment of Education Program (NAEP). A representative sample of 17-year-olds was given specific information regarding the diet of frontiersmen and then asked to compare what these people ate to their own diet. Assuming that the students understood their own diet, this task primarily was asking them to draw on their abilities to compare. Shockingly, only 27% of the students in the representative sample received a score of "adequate" or better for this task (Mullis et al., 1990). When the responses of the remaining 73% of the students who received lower than an adequate score were analyzed, the researchers found that these students did not actually compare their own diet to that of the frontiersmen, but simply listed which foods each group ate. It appeared that these students did not understand that, in order to compare, they should:

1. Identify by which characteristics they will compare the items and explain why these characteristics are significant.

2. Explain how the items are similar and different according to the characteristics they have used.

2. Communication Skills

Inherent in performance tasks and portfolios is an "exhibition" of knowledge. Exhibitions are simply presentations of student work. As defined by assessment expert Grant Wiggins, an exhibition requires students to present the fruit of their work (in Willis, 1996). Education reporter Scott Willis (1996) offers the following description of exhibitions:

> Typically [exhibitions are] multimedia in nature: Students may have to write a paper, make an oral presentation, build a model or create computer graphics, and respond spontaneously to questions. Often, exhibitions are a "culminating" performance. The audience for exhibitions may include teachers, classmates, younger students, parents, or other community members. (p. 1)

Quite obviously, one must possess communication skills to effectively exhibit knowledge. Consequently, teaching and reinforcing basic communication skills should help improve student performance on these tasks. Marzano, Pickering, and McTighe (1993) have listed specific communication skills which are often embedded in performance tasks and portfolios:

- Expresses ideas clearly.
- Effectively communicates with diverse audiences.
- Effectively communicates in a variety of ways.
- Effectively communicates for a variety of purposes.

Each of these skills are briefly highlighted. Clarity of expression is an internal component of every form of communication, whether it takes the shape of a written essay, an oral report, an audiotaped report, or other. In every case, there must be a clear main point or theme that is backed up with appropriate supporting detail. If one of the aforementioned communication skills were to be identified as superordinate to the rest, clarity of expression might well be the one.

The ability to communicate with diverse audiences is also an important aspect of effective communication. For students, such audiences might include peers, parents, experts, novices, the general public, and school board members. The select audiences with which a student can effectively communicate will increase as that individual matures. For instance, a primary student might only be able to communicate with parents and teachers, while a high school student would be expected to communicate effectively with a broad spectrum of audiences. According to current theory in rhetoric (Durst and Newell, 1989), sensitivity to the level of knowledge of a particular audience and an understanding of the interests of its members are essential if one is to communicate with that audience. Sensitivity to audience also includes appropriately matching the tone and style of communication so that it can best be received by a given audience. To overlook these important aspects can result in communication that is logically cohesive, but not enjoyed or fully understood by the audience.

Skilled communicators also use a variety of communication forms. Most schools only emphasize two of these forms: writing and speaking. Because we live in an information-oriented society, however, there are a number of other forms which are both useful and appropriate:

- Oral reports
- Videotapes
- Written reports
- Panel discussions
- Dramatic enactments
- Outlines
- Debates
- Graphic representations
- Newscasts
- Discussions
- Audiotapes

- Flowcharts
- Slide shows

All of these are viable communication tools, although sometimes students may wish to convey emotion as well as information. In these instances, they may select other modes of communication:

- Collages
- Dances
- Plays
- Songs
- Paintings
- Pictures
- Sculptures

To be an effective communicator, then, one must be proficient in a variety of communication forms.

Lastly, effective communicators must also be able to communicate for many different purposes. For example, communications meant to inform, to persuade, to generate questions, or to elicit sympathy, anger, humor, pride, or joy, will each take slightly different forms. Research has noted that people who are able to write for specific purposes have some understanding of certain rhetorical conventions (Durst and Newell, 1989), and effective communicators know about and are able to apply these conventions.

3. Lifelong Learning Skills

Lifelong learning skills involve those competencies that, as implied by their name, are used throughout life in many different situations. These competencies are often associated with the world of work and include:

- Demonstrating the ability to work toward the achievement of group goals
- Demonstrating effective interpersonal skills
- Restraining impulsivity
- Seeking multiple perspectives
- Setting and managing progress toward goals
- Persevering
- Pushing the limits of one's abilities

(For a more detailed list of lifelong learning skills, see Costa [1984], and Marzano [1992].)

Lifelong learning skills received national attention when the report *What Work Requires of Schools: A SCANS Report for America 2000* was published by the Secretary's Commission on Achieving Necessary Skills (SCANS) in 1991. The commission devoted one year to "talking to

business owners, to public employees, to the people who manage employees daily, to union officials, and to workers on the line and at their desks. We have talked to them in their stores, shops, government offices, and manufacturing facilities" (p. v). The majority of those surveyed believed that American students must learn the skills and abilities C including those listed above C required to make them productive members of the work force. The American Society for Training and Development (ASTD) published a complementary report to the SCANS work entitled *Workplace Basics: The Essential Skills Employers Want* (Carnevale, Gainer, & Meltzer, 1990). The individual skills identified in this piece were almost exactly the same as those named by the SCANS report.

Parents also have emphasized the importance of lifelong learning skills. The polling firm Public Agenda surveyed a representative sample of parents about what they believed should be taught in the public schools, and published their findings as *First Things First: What Americans Expect From Public Schools* (Farkas, Friedman, Boese, & Shaw, 1994). This report stated that 88% of the parents surveyed thought that schools should teach and reinforce work-related skills, such as punctuality, dependability, and self-discipline.

Educators seem to have reached the same conclusions regarding lifelong learning skills. Specifically, the American Association of School Administrators surveyed 55 prominent educators C the "Council of 55" C about what students must learn in school in order to be adequately prepared for the 21st century. The council identified interpersonal skills and the ability to be part of a team as vital components to success in the next century (Uchida, Cetron, & McKenzie, 1996).

To summarize, lifelong learning skills seem to be of great value to those constituent groups related both directly and indirectly to education.

Helping Students Create Performance Tasks

How can a teacher help reinforce thinking and reasoning skills, communication skills, and lifelong learning skills in the classroom? One of the best ways is to help students design their own performance tasks incorporating these elements. Marzano, Pickering, and McTighe (1993) have developed a useful process for doing this. It involves the following steps:

<u>Step 1</u>:
Have students identify a question related to something in the current unit of study that interests them. When students construct their own performance tasks, they usually do so by identifying a question that interests them. We have found that providing students with questions that are cued to the thinking and reasoning processes discussed previously can generally aid this process. These "student-oriented" questions are listed in Figure 3.6.

To demonstrate how these questions might be used, consider the following scenario: A student has been studying John F. Kennedy and has been asked to create a performance task. The student would begin by selecting a question from the list in Figure 3.6 that he would like to

explore. After considering his options, the student identifies the question "Do you have a hypothesis about a future event@ found under hypothesis testing and scientific inquiry. Specifically, the student realizes that he has a hypothesis about what might have happened if JFK had not been assassinated.

Step 2:
Help students develop a first draft of the task that utilizes one or more of the reasoning processes listed in Figure 3.6. The teacher helps the student use the basic question to develop a first draft of the performance task. The first draft of the task about John F. Kennedy might read as follows:

> I'm going to examine what might have happened if John F. Kennedy had not been assassinated. I will identify what I believe would have happened and provide evidence for my prediction.

Step 3:
Help students identify effective communication skills that will be incorporated into the performance task. Next, the student begins to think about which communication skills he might use as part of his performance task. Previously, we considered the following communication skills, which are often included in performance tasks:

- Expressing ideas clearly
- Effectively communicating with diverse audiences
- Effectively communicating in a variety of ways
- Effectively communicating for a variety of purposes

The student uses this list or one generated by the district, school, or classroom teacher to select the skill that best fits his task. For this discussion, assume that the student has identified Aexpressing ideas clearly@ as the best match for his chosen task.

Thinking and Reasoning Process	Related Questions
1. Identifying similarities and differences	Do you want to determine how things are similar and different? Do you want to organize things into groups? Do you want to identify the rules or characteristics that have been used to form groups? Do you see a relationship that no one else sees? What is the abstract pattern or theme that is at the heart of the relationship?
2. Problem-solving and trouble shooting	Do you want to describe how some obstacle can be overcome?

	Do you want to improve on something?
3. Argumentation	Is there a position you want to defend on a particular issue? Are there differing perspectives on an issue you want to explore?
4. Decision-Making	Is there an important decision that should be studied or made?
5. Hypothesis Testing and Scientific Inquiry	Is there a prediction you want to make and then test? Do you have a hypothesis about a past or future event that you want to explore? Do you have a new theory or idea that you want to explore?
6. Logic	What rule or rules are operating in this situation? Based on these rules, what can be concluded? Are any rules not being followed in this situation?

Figure 3.6. Questions for Thinking and Reasoning Processes.

Step 4:

Helping students select a lifelong learning skill that they will incorporate into their performance task. Lifelong learning is the final area students might consider when generating their performance tasks. As indicated by the previous discussion, the options available within this category of skills and abilities are quite numerous. Assuming that the student has decided to work with others in the John F. Kennedy task, he might elect to incorporate the lifelong learning skill "the ability to work toward group goals" into his task.

Step 5:

Help students rewrite the task so that it clearly identifies all skill areas. After the student has selected all the elements to be incorporated into the task, he rewrites the task to make these elements explicit. The student studying John F. Kennedy might rewrite his task in the following way:

I'm going to examine what might have happened if John F. Kennedy had not been assassinated. I will make a prediction and provide evidence that supports it. Working with two other people who have identified similar topics, I will gather information from various sources. While working with my research partners, I will keep track of how well I monitor my behavior in the group. After I have collected enough information, I will make an oral report, taking special care to express my ideas clearly.

Step 6:
Once students have completed the task, provide feedback on each element of the task.
Receiving feedback regarding the skills embedded in a performance task is a critical component for students engaging in such a project. If they are to effectively learn from the experience, four components appear to be involved in the task we have been following:

1) An understanding of the crucial elements of John F. Kennedy's presidency

2) The ability to generate and defend a hypothesis

3) The ability to express ideas clearly

4) The ability to work toward group goals

Students should receive individual and specific feedback for each of these four areas. Chapter 4 details techniques by which a teacher might assess a task like the John F. Kennedy sample. However, the most effective method is to outline levels of performance C often called a rubric C for each of the task components. The student or group of students working on the J.F.K. task would consequently receive four rubric scores, one for each area of the task. Each of the scores would be granted independently of the other three.

CHAPTER 4

REPORTING OUT BY INDIVIDUAL STANDARDS

Perhaps the most radical approach to implementing standards is to report student progress on individual standards in each class. This drastically changes the nature and format of report cards. To illustrate, consider the sample report card in Figure 4.1 (pp. 41-42).

In this example, each teacher has assessed an overall course grade for the student, but has provided scores for individual standards as well. Notice also that the teachers have utilized a four-point scale with the corresponding skill levels of novice, basic, proficient, and advanced. The student could also be graded on a three-point scale, a five-point scale, and so on. The actual number of points is not so critical as the fact that all teachers are utilizing the same scale.

The report card in Figure 4.1 has a dual purpose in that it not only offers letter grades with which students and parents are already familiar, but it also rates a student on specific standards. It is important to understand that with this approach, there may very well be some repetition of standards from course to course, and most reasoning, communication, and lifelong learning standards (discussed in Chapter 3) are commonly the focus of core courses because these standards usually span all content areas. In Figure 4.1, for example, logic and reasoning are covered under both science and mathematics.

Consequently, a report card which rates students' performance according to specific standards requires a transcript that does the same. Figure 4.2 shows a sample transcript based on standards.

Notice that the scores in the first column represent an average for each standard, indicating that the students have been previously assessed on individual standards. The number of times each standard assessment has occurred is listed in the column to the right the average score. In this example, probability (mathematics standard 5) has been assessed three times with an average score of 1.7. This transcript also lists the highest score received for the standard (3), the lowest score (1), and the most current score (3). The most recent standards scores are of great interest to some. (For a discussion, see Guskey 1996a.) As indicated by the name, this score is representative of the most recent classroom assessment of the student's performance on standards. When implementing the use of this kind of transcript, a district or school must decide whether to report overall performance on a set of standards using all the scores, or just the most recent ones. The sample transcript in Figure 4.2 reports both.

Nobel County School District 1: George Washington High School Student Progress Report

Name: Al Einstein
Address: 1111 E. McSquare Dr.
City: Relativity, Colorado 80000
Grade Level: 11

Course Title	Grade
Algebra II and Trigonometry	C-
Advanced Placement Physics	A +
U.S. History	B-
American Literature	C +
Physical Education	B-
Chorus	B +
Geography	B-

Current GPA: 2.81
Cumulative GPA: 3.23

Standards Ratings

Algebra II and Trigonometry

		Novice (1)	Basic (2)	Proficient (3)	Advanced (4)
Mathematics Standard 1:	Numeric Problem-Solving	-------(1)			
Mathematics Standard 2:	Computation	-------(1)			
Mathematics Standard 3:	Measurement		----------(2)		
Mathematics Standard 4:	Geometry		----------(2)		
Mathematics Standard 5:	Probability		----------(2)		
Mathematics Standard 6:	Algebra	-------(1)			
Mathematics Standard 7:	Data Analysis		----------(2)		
Reasoning Standard 5:	Decision-Making		----------(2)		
Lifelong Learning Standard 4:	Self-regulation	-------(1)			
Overall Mathematics:	**1.6**				

Advanced Placement Physics

		Novice (1)	Basic (2)	Proficient (3)	Advanced (4)
Science Standard 1:	Earth and Space				----------(4)
Science Standard 2:	Life Sciences			----------(3)	
Science Standard 3:	Physical Sciences				----------(4)
Science Standard 4:	Science and Technology				----------(4)
Reasoning Standard 4:	Princ. Of Scientific Inquiry				----------(4)
Lifelong Learning Standard 1:	Working with Groups			----------(3)	
Overall Science:	**3.7**				

U.S. History

		Novice (1)	Basic (2)	Proficient (3)	Advanced (4)
History Standard 1:	Civilization and Society		----------(2)		
History Standard 2:	Exploration and Colonization			----------(3)	
History Standard 3:	Revolution and Conflict				----------(4)
History Standard 4:	Industry and Commerce		----------(2)		
History Standard 5:	Forms of Government		----------(2)		
Reasoning Standard 3:	Identifying Similarities & Differences			----------(3)	
Lifelong Learning Stand. 3:	Leadership Skills			----------(3)	
Overall History:	**2.7**				

Nobel County School District 1: George Washington High School Student Progress Report

Name: Al Einstein

American Literature

		Novice	Basic
		(1)	(2)
Lang. Arts Standard 1:	The Writing Process	----------(2)	
Lang. Arts Standard 2:	Usage, Style, and Rhetoric	----------(3)	
Lang. Arts Standard 3:	Research: Process & Product	--------(3)	
Lang. Arts Standard 4:	The Reading Process	----------(2)	
Lang. Arts Standard 5:	Reading Comprehension	----------(3)	
Lang. Arts Standard 6:	Literary/Text Analysis	----------(2)	
Lang. Arts Standard 7:	Listening and Speaking	----------(2)	
Lang. Arts Standard 8:	The Nature of Language	----------(2)	
Lang. Arts Standard 9:	Literature	-------(1)	
Reasoning Standard 1:	Principles of Argument	----------(4)	
Lifelong Learning Standard 5:	Reliability & Responsibility	----------(2)	
Overall Lang. Arts:	**2.4**		

Physical Education

	Novice	Basic
	(1)	(2)

Phys. Ed. Standard 1: Movement Forms: Theory and Practice ----------(2)
Phys. Ed. Standard 2: Motor Skill Development ----------(3)
Phys. Ed. Standard 3: Physical Fitness: Appreciation ----------(3)
Phys. Ed. Standard 4: Physical Fitness: Application ----------(2)
Reasoning Standard 6: Decision-Making ----------(3)
Lifelong Learning Standard 1: Working with Groups ----------(2)
Overall Phys. Ed.: 2.5

Chorus

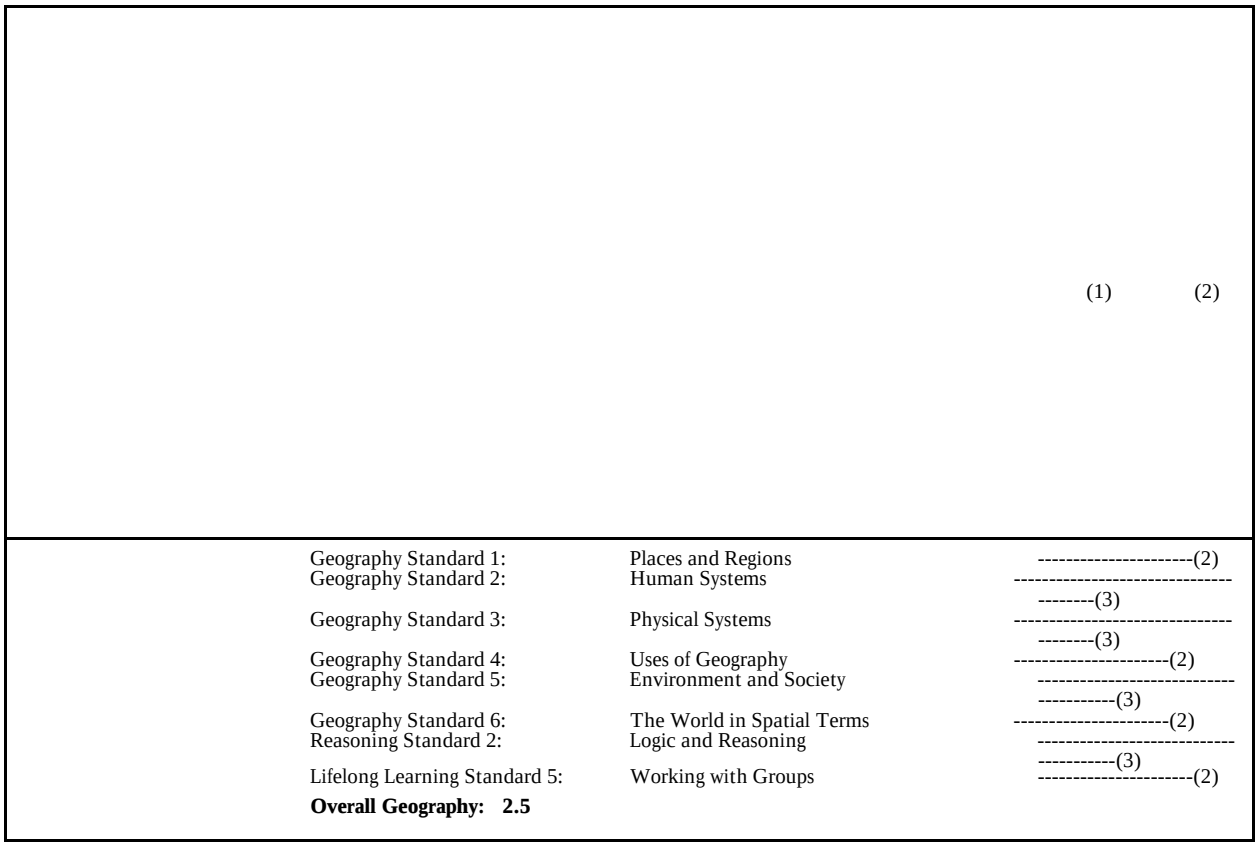
Novice Basic

(1) (2)

Music Standard 1: Vocal Music ----------(3)
Music Standard 2: Instrumental Music ----------(3)
Music Standard 3: Music Composition ----------(3)
Music Standard 4: Music Theory ----------(2)
Music Standard 5: Music Appreciation ----------(4)
Reasoning Standard 3: Identifying Similarities & Differences ----------(3)
Lifelong Learning Standard 2: Working with Individuals ----------(3)
Overall Music: 3.0

Geography

Novice Basic



Note: From R. J. Marzano and J. S. Kendall. (1996). *A Comprehensive Guide to Designing Standards-Based Districts, Schools, and Classrooms.* Copyright 8 1996 by McREL Institute. Reprinted with permission.

Figure 4.1. Sample Report Card: Reporting Student Performance by Grade and by Standard.

Subject and Standards Rated Average	Average Rating	Number of Ratings	Most Recent Rating	Highest Rating	Lowest Rating
Subject: MATHEMATICS					
Standard 1: Numeric Problem-Solving	1.7	3	3	3	1
Standard 2: Computation	1.3	3	2	2	1
Standard 3: Measurement	2.7	3	2	3	2
Standard 4: Geometry	1.5	2	2	2	1
Standard 5: Probability	1.7	3	3	3	1
Standard 6: Algebra	1.0	2	1	1	1
Standard 7: Data Analysis	3.0	1	3	3	3
Overall Mathematics	**1.84**	**17**	**2.28**	**3**	**1**
Subject: SCIENCE					
Standard 1: Earth and Space	4.0	4	4	4	4
Standard 2: Live Sciences	3.5	2	4	4	3
Standard 3: Physical Sciences	3.5	4	4	4	2
Standard 4: Science and Technology	3.75	4	4	4	3
Overall Science	**3.69**	**14**	**4.0**	**4**	**2**
Subject: HISTORY					
Standard 1: Civilization & Hmn. Society	2.75	4	3	3	2

Standard 2: Exploration & Colonization	3.0	3	3	3	3
Standard 3: Revolution and Conflict	3.75	3	3	4	3
Standard 4: Industry and Commerce	2.3	3	3	3	1
Standard 5: Forms of Government	3.0	2	2	4	2
Overall History	**2.96**	**15**	**2.8**	**4**	**1**
Subject: GEOGRAPHY					
Standard 1: Places and Regions	2.0	2	1	3	1
Standard 2: Human Systems	3.75	4	3	4	3
Standard 3: Physical Systems	2.5	4	3	3	2
Standard 4: Uses of Geography	3.5	2	4	4	3
Standard 5: Environment and Society	3.0	3	4	4	2
Standard 6: The World in Spatial Terms	2.5	2	3	3	2
Overall Geography	**2.88**	**17**	**3.0**	**4**	**1**
Subject: LANGUAGE ARTS					
Standard 1: The Writing Process	2.6	7	3	3	2
Standard 2: Usage, Style and Rhetoric	3.0	9	4	4	2
Standard 3: Research: Process & Product	2.8	5	4	4	2
Standard 4: The Reading Process	2.6	5	2	3	2
Standard 5: Reading Comprehension	3.6	9	2	4	2
Standard 6: Literary/Text Analysis	2.8	6	3	3	2
Standard 7: Listening and Speaking	3.5	10	4	4	3
Standard 8: The Nature of Language	3.0	3	4	4	2
Standard 9: Literature	2.0	3	2	2	2
Overall Language Arts	**2.88**	**57**	**3.1**	**4**	**2**

Subject and Standards Rated Average	Average Rating	Number of Ratings	Most Recent Rating	Highest Rating	Lowest Rating
Subject: THE ARTS/MUSIC					
Standard 1: Vocal Music	2.0	2	3	3	1
Standard 2: Instrumental Music	3.3	3	3	4	3
Standard 3: Music Composition	2.0	2	2	2	2
Standard 4: Music Theory	3.0	2	2	4	2
Standard 5: Music Appreciation	4.0	3	4	4	4
Overall Music	**2.86**	**12**	**2.8**	**4**	**1**
Subject: PHYSICAL EDUCATION					
Standard 1: Movement Forms: Theory & Practice	2.3	3	2	3	2
Standard 2: Motor Skill Development	2.0	4	3	3	1
Standard 3: Physical Fitness: Appreciation	3.75	4	4	4	3
Standard 4: Physical Fitness: Application	2.0	4	3	3	1
Overall Physical Education	**2.5**	**15**	**3.0**	**4**	**1**
Subject: REASONING					
Standard 1: The Principles of Argument	3.7	10	4	4	2
Standard 2: Logic and Reasoning	3.0	10	4	4	2
Standard 3: Identifying Similarities & Differences	3.0	12	4	4	2
Standard 4: Principles of Scientific Inquiry	3.6	3	4	4	3
Standard 5: Techniques of Problem-Solving	3.8	13	4	4	3
Standard 6: Techniques of Decision-Making	3.2	13	4	4	2
Overall Reasoning	**3.4**	**61**	**4.0**	**4**	**2**
Subject: LIFELONG LEARNING SKILLS					
Standard 1: Working with Groups	2.8	17	3	3	2
Standard 2: Working with Individuals	3.01	17	4	4	2
Standard 3: Leadership Skills	2.7	14	3	3	2
Standard 4: Self-Regulation	2.6	13	3	3	1
Standard 5: Reliability and Responsibility	3.0	17	3	3	3
Overall Lifelong Learning Skills	**2.82**	**78**	**3.2**	**4**	**1**
All subject areas combined	**2.87**	**286**	**3.1**	**4**	**1**

Note: From R. J. Marzano and J. S. Kendall. (1996). *A Comprehensive Guide to Designing Standards-Based Districts, Schools, and Classrooms*. Copyright 8 1996 by McREL Institute. Reprinted with permission.

Figure 4.2. Sample Transcript: Reporting Student Performance by Standard.

The Inherent Danger in Changing Report Cards

For a school or district to adopt a new report card and transcript format like the ones shown in Figures 4.1 and 4.2, there is often an element of risk. Education reporter Lynn Olson (1995b) provides an accounting of what happened in a Rhode Island school district as an example of the dynamics that can occur when a district tries to change its traditional grading and reporting practices. Administrators, parents, and volunteer community members in the district worked for two years to create a reporting system that evaluated students on fairly specific information and

skills. To the shock of this report card committee, some parents in the district had a strong negative reaction to the new system, even though it had been subjected to extensive study and testing. In the following passage, Olson (1995b) dramatizes this group:

> The three women seated around Dona LeBouef's butcher-block kitchen table look more like a bevy of P.T.A. moms than a rebel army. Dressed in coordinated shirts and pants and denim jumpers, they're articulate and polite. Classical music plays softly in the background as they sip their coffee and review the weapons in their campaign: a large sheaf of photocopied newspaper articles and editorials, old report cards, and petitions.
>
> Their target is pilot report cards introduced by the public school system here last fall that eliminated traditional letter grades in the elementary schools. The new format, tested citywide, was designed to more accurately reflect the teaching going on in the classroom and to provide families with more detailed information about their children. School officials thought parents would be pleased. They were wrong. (p. 23)

Olson explains that the elimination of the traditional A-B-C grading format was upsetting to the parents. Basically, the new report cards looked too different from what the parents were used to. Although there was sufficient evidence to indicate that the new system was more accurate and informative than the one previously in place, a relatively small group of parents was able to rally the support of 1,300 community members who signed a petition protesting the new grading format. As Olson (1995b) notes:

> At issue is one of the most sacred traditions in American education: the use of letter grades to denote student achievement. The truth is that letter grades have acquired an almost cult-like importance in American schools. They are the primary shorthand tool for communicating to parents how children are faring. (p. 24)

In spite of the friction which might be felt when changing to a reporting system that focuses on individual standards, the struggle and effort are apparently worth it. Wiggins (1994) writes: "Using a single grade with no clear and stable meaning to summarize all aspects of performance is the problem. We need more, not fewer, grades and more efficient kinds of grades if the parent is to be informed.@ (p. 29)

One of the advantages to the report card pictured in Figure 4.1 is that it provides students and teachers with both a familiar A-B-C overall grade and less-familiar specific scores on specific standards. This kind of record-keeping will obviously require changes in classroom practice. We have detailed four steps that classroom teachers can follow if they wish to adopt this very specific form of reporting:

Step 1. Organize Your Content Around Standards

The first step a classroom teacher must take to begin reporting student performance according to specific standards is to identify which standards will be addressed within a given grading period, as well as the specific content within each standard. This task is greatly facilitated if the school, district, or state department of education has already identified educational benchmarks for specific grades. Benchmarks define the knowledge or skill that should be addressed as part of a specific standard. For example, Figure 4.3 provides the grade 6 through 8 benchmarks for the Florida state science standard titled "The student understands the basic principles of atomic theory."

2. The student understands the basic principles of atomic theory.

- the student describes and compares the properties of particles and waves
- the student knows the general properties of the atom (a massive nucleus of neutral neutrons and positive protons surrounded by a cloud of negative electrons) and accepts that single atoms are not visible
- the student knows that radiation, light, and heat are forms of energy used to cook food, treat diseases, and provide energy

(State of Florida, Department of State, 1996)

Figure 4.3. Florida State Science Standard and Benchmarks for Grades 6 Through 8.

Benchmarks as specific as these give teachers guidance on the actual content that should be covered in class. Consequently, the sunshine standard would serve as a guide for an eighth-grade teacher in Florida reporting student achievement on individual standards, because it would help him make decisions regarding the specific content to address. We have observed that the more information a teacher receives regarding the specific content to be covered, the better.

Step 2. Plan the Types of Assessment That Will Be Used for the Various Standards

Ultimately, judgments must be made regarding student performance on the particular standards addressed in an instructional unit. This requires gathering information regarding each student=s performance on each standard. *Assessment* is the common term for the act of "gathering information about student performance." Unfortunately, many educators have a very narrow view of assessment: When asked to assess students, they immediately interpret this to mean design a test. Actually, almost any means of gathering information on student achievement can be considered assessment.

Recent years have seen a dramatic increase in the means of assessment suitable for classroom use. We have observed that different types of assessments are best suited to different kinds of content and have depicted this in Figure 4.4.

	Forced-Choice Items	Essay Questions	Performance Tasks	Teacher Observation	Student Self-Assessment
specific declarative ge	H	H	H	M	H
specific procedural ge	L	H	H	H	H
g and reasoning skills	L	H	H	M	H
nication Skills	L	H	H	L	H
learning skills	L	M	M	H	H

Figure 4.4. Types of Assessment for Different Types of Standards.

Figure 4.4 provides a score of High (H), Medium (M), or Low (L) for each of several main types of classroom assessment regarding their usefulness for rating different types of knowledge and skill. As depicted in Figure 4.4, different forms of classroom assessment are effective for some types of knowledge, but not for others. Each type is considered individually.

Forced-Choice Items

Stiggins (1994) gives the following definition for forced-choice items:

> This is the classic objectively scored paper and pencil test. The respondent is asked a series of questions, each of which is accompanied by a range of alternative responses. The respondent's task is to select either the correct or the best answer from among the options. The index of achievement is the number or proportion of questions answered correctly. (p. 84)

Although up to this point we have limited our consideration of forced-choice tests to those utilizing multiple-choice items, Stiggins discusses four types of forced-choice items: (1) multiple-choice items, (2) true/false items, (3) matching exercises, and (4) short answer

fill-in-the-blank items. Stiggins (1994) explains that short answer, fill-in-the-blank items are categorized here because they allow for only a single right or wrong answer.

Teachers often use forced-choice items in combination with essay items to develop homework assignments, quizzes, midterm examinations, and final examinations. As one might guess, these items play a substantial role in classroom assessment. Some educators mistakenly believe that forced-choice items should be discarded entirely and replaced with formats that require students to use their own knowledge and understanding to construct personal responses. These educators are not acknowledging the fact that forced-choice tests do play an important role in assessment. To understand this role, one must first recognize the difference between two main categories of knowledge C declarative and procedural. This distinction is considered key by many cognitive psychologists (Anderson, 1982, 1983, 1990a, 1990b, 1993, 1995; Fitts, 1964; Fitts & Posner, 1967; Frederiksen, 1977; Newell & Simon, 1972; Norman, 1969; Rowe, 1985; van Dijk, 1980).

Declarative knowledge is best described as information and often contains component parts. Knowledge of an "average," for instance, requires a basic understanding of the concept of distributions, the concept of a range of scores, and so on. Procedural knowledge, on the other hand, pertains to skills, strategies, and processes. Calculating the average of a group of scores, for example, requires the basic computation skills of addition, subtraction, multiplication, and division.

These two categories of knowledge are highly interactive, so students risk having gaps in their knowledge if one type is neglected. Returning to the example of averages, a student might be able to calculate an average from a set of data, but not have any concept of what the average could tell him about the data set. Likewise, a student might understand the information provided by an average, but still not be able to perform the mathematical computations necessary to calculate that average.

Figure 4.4 shows forced-choice items to be straightforward and effective ways to assess student understanding of declarative knowledge, especially factual knowledge. An example of this is a recommendation made by the National Center for Research on Evaluation Standards and Student Testing (CRESST) at UCLA. It recommends that students provide short answers about the following content before completing a complex essay question dealing with the Lincoln/Douglas debate:

- Popular sovereignty
- Dred Scott
- Missouri Compromise
- Bleeding Kansas
- States- rights
- Federalism
- Underground railroad
- Abolitionists

(Baker et al., 1992, pp. 15-16)

These short-answer items give students the freedom to deal with the broader aspects of the time period in their essays without including too much detail in an effort to demonstrate a thorough knowledge of the topic.

Forced-choice items can also serve as a tool for assessing some types of procedural knowledge, specifically those procedures that utilize a series of specific steps in a specific order. These procedures are termed algorithms. The four computational processes C addition, subtraction, multiplication, and division, are all algorithms that could be assessed effectively using forced-choice items. More complex procedures, however, cannot be assessed with forced-choice items. For instance, the process of writing cannot be effectively assessed in a forced-choice format.

Essay Questions

Essay questions have become a time-honored staple for classroom teachers (Durm, 1993). As indicated in Figure 4.4, they can be employed effectively to assess not only declarative and procedural knowledge, but the use of thinking and reasoning skills as well. When used to assess declarative knowledge, they are often intended to determine how well a student understands concepts and generalizations C the Abig ideas@ C and the relationships among them.

A method that maximizes the effectiveness of essay questions is to give students information to which they can react and construct responses. The example presented in Figure 4.5 is from a history exam administered by CRESST, in which students are provided with original transcripts from the Lincoln/Douglas debate.

With this information as the groundwork presented to all students, the following essay item is given:

> Imagine that it is 1858 and you are an educated citizen living in Illinois. Because you are interested in politics and always keep yourself well-informed, you make a special trip to hear Abraham Lincoln and Stephen Douglas debating during their campaigns for the Senate seat representing Illinois. After the debates you return home, where your cousin asks you about some of the problems that are facing the nation at this time.
>
> Write an essay in which you explain the most important ideas and issues your cousin should understand. Your essay should be based on two major sources: (1) the general concepts and specific facts you know about American History, and especially what you know about the history of the Civil War; (2) what you have learned from the readings yesterday. Be sure to show the relationships among your ideas and facts. (Baker et al., 1992, p. 23)

Stephen A. Douglas

Mr. Lincoln tells you, in his speech made at Springfield, before the Convention which gave him his unanimous nomination, that C

"A house divided against itself cannot stand."
"I believe this government cannot endure permanently, half slave and half free."
"I do not expect the Union to be dissolved, I don't expect the house to fall; but I do expect it will cease to be divided."
"It will become all one thing or all the other."

That is the fundamental principle upon which he sets out in this campaign. Well, I do not suppose you will believe one word of it when you come to examine it carefully, and see its consequences. Although the Republic has existed from 1789 to this day, divided into Free States and Slave States, yet we are told that in the future it cannot endure unless they shall become all free or all slave. For that reason he says. . .

Abraham Lincoln

Judge Douglas made two points upon my recent speech at Springfield. He says they are to be the issues of this campaign. The first one of these points he bases upon the language in a speech which I delivered at Springfield which I believe I can quote correctly from memory. I said there that "we are now far into the fifth year since a policy was instituted for the avowed object, and with the confident promise, of putting an end to slavery agitation; under the operation of that policy, that agitation had not only not ceased, but had constantly augmented." "I believe it will not cease until a crisis shall have been reached and passed. 'A house divided against itself cannot stand.' I believe this Government cannot endure permanently, half slave and half free." "I do not expect the Union to be dissolved" C I am quoting from my speech C "I do not expect the house to fall, but I do expect it will cease to be divided. It will become all one thing or the other. Either the opponents of slavery will arrest the spread of it and place it where the public mind shall rest, in the belief that it is in the course of ultimate extinction, or its advocates will push it forward until it shall become alike lawful in all the States, North as well as South.". . .

Note: From *Political Debates Between Abraham Lincoln and Stephen A. Douglas*, by Cleveland, 1902, in *CRESST Performance Assessment Models: Assessing Content Area Explanations* (pp. 43-47), by E. L. Baker, P. R. Aschbacher, D. Niemi, and E. Sato, 1992, Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), UCLA.

Figure 4.5. Excerpts from the Lincoln/Douglas Debate.

Note that to complete this task, students must discuss general concepts and relationships among ideas, neither of which can be assessed adequately using forced-choice items. This exemplifies the fact that essay questions are most appropriate for dealing with big ideas and ideas within the area of declarative knowledge. Forced-choice items, on the other hand, lend themselves much more readily to assessing lower level factual information.

Figure 4.4 also shows that essay items can be a fairly effective means of assessing procedural knowledge. In this case, students are asked to explain or critique a procedure, as in the CRESST chemistry example which follows:

> Imagine you are taking a chemistry class with a teacher who has just given the demonstration of chemical analysis you read about earlier.

> Since the start of the year, your class has been studying the principles and procedures used in chemical analysis. One of your friends has missed several weeks of class because of illness and is worried about a major exam in chemistry that will be given in two weeks. This friend asks you to explain everything that she will need to know for the exam.
>
> Write an essay in which you explain the most important ideas and principles that your friend should understand. In your essay you should include general concepts and specific facts you know about chemistry, and especially what you know about chemical analysis or identifying unknown substances. You should also explain how the teacher's demonstration illustrates important principles of chemistry.
>
> Be sure to show the relationships among the ideas, facts, and procedures you know. (Baker et al., 1992, p. 29)

Granted, a more direct assessment of the student's understanding of the chemical analysis procedures would require a student to actually demonstrate these procedures, but the essay question can still give a classroom teacher important insight into the student=s skills and mental processes. In fact, Shavelson and his colleagues (Shavelson & Baxter, 1992; Shavelson, Gao, & Baxter, 1993; Shavelson & Webb, 1991; Shavelson, Webb, & Rowley, 1989) have discovered that this indirect assessment of procedural skills is highly correlated with more straightforward, hands-on types of assessment.

Essay questions can also effectively assess thinking and reasoning skills. In Chapter 3, we considered six skills taken from the national standards documents. When a student uses declarative knowledge in conjunction with these reasoning processes to construct an essay, he must prove himself competent in both the declarative knowledge and the aforementioned thinking and reasoning processes. Assume, for instance, that a teacher is planning an essay test based on the information found in the Lincoln/Douglas debate. She also wants the students to use a specific reasoning skill or skills. The resulting essay question might be:

> Douglas and Lincoln said many things in their debate. Identify their areas of agreement as well as their areas of disagreement. Then, select one of their areas of disagreement and analyze the arguments each has presented to determine which one has presented the best case. In your analysis, look at the logic of each argument as well as the accuracy of their information.

This essay question is actually designed to assess three elements in regard to the Lincoln/Douglas debate. Two of these involve thinking and reasoning and the third deals with declarative knowledge:

1. Students' ability to compare (see thinking and reasoning skill #1)

2. Students' ability to detect errors in logic (see thinking and reasoning skill #6)

3. Students' understanding of the accuracy of the information presented by Lincoln and Douglas

Lastly, Figure 4.4 shows that essay items can also be used to assess communication skills. It is clear that a student's essay response to the Lincoln/Douglas debate question could provide a teacher with information about that student's ability to clearly express ideas as well as other communication skills.

To summarize, essay questions can serve as an effective assessment tool for a variety of skill areas.

Performance Tasks

We have already discussed performance tasks in previous chapters. From our previous discussion and the information provided in Figure 4.4, it should now be clear that performance tasks can be used to effectively assess many types of knowledge and skill. A natural question to ask at this point is what is the difference between an essay question and a performance task? Actually, a good essay question is a performance task. To be even more specific, an essay question can be considered one type of performance task if it unites declarative knowledge with at least one of the reasoning processes. This means that the sample essay question on the Lincoln/Douglas debate would qualify as a performance task. If that question had only asked the students to retell the important aspects of the debate, however, it would only serve to assess the students' declarative knowledge. It would not be a performance task according to our definition, because knowledge would not be applied using one or more of the thinking and reasoning skills.

According to Figure 4.4, the only type of knowledge that cannot be effectively assessed using performance tasks is lifelong learning skills. We gave performance tasks a medium rating in this area because, while a teacher can gain some valuable insight regarding student competence by using performance tasks, more direct forms of observation are better suited to assessing lifelong learning skills.

Teacher Observation

Informal observation of students is among the most straightforward methods used to collect assessment data. This has been termed Akid watching@ by educators like reading expert Yetta Goodman (Goodman, 1978; Wilde, 1996). As the name implies, within this approach, the teacher observes and makes note of students' competence as they go about their daily business. This is the most "unobtrusive" way for teachers to gather assessment data because the students are not given any test or special assignment. The following example from Stiggins (1994) demonstrates how a teacher might observe a student=s social interaction skills relating to a school=s lifelong learning standard:

> A primary-grade teacher might watch a student interacting with classmates and draw inferences about that child's level of development in social interaction skills. If the levels of achievement are clearly defined in terms the observer can easily interpret, then the teacher, observing carefully, can derive information from watching that will aid in planning strategies to promote further social development. Thus, this is not an assessment where answers are counted right or wrong. Rather, like the essay test, we rely on teacher judgment to place the student's performance somewhere on a continuum of achievement levels ranging from very low to very high. (p. 160)

As depicted in Figure 4.4, teacher observation lends itself best to assessing procedural knowledge and lifelong learning skills because competence in these two areas manifests itself as observable behavior. For instance, a teacher can observe as a student demonstrates map reading skills C a procedure vital to the geography content area, or shows group leadership competency C a lifelong learning procedural skill.

Student Self-Assessment

Student self-assessment is perhaps the most useful form of assessment data. As suggested by the name, the assessment data within this approach come from the student herself. Wiggins (1993a) is so strongly in favor of student self-assessment that he has made it one of his nine postulates for a more thoughtful assessment system: "Postulate 4: An authentic education makes self-assessment central" (p. 53).

Hansen (1994) stresses that self-assessment is at the core of the development of higher order metacognitive skills. She also postulates that self-assessment aids in the generation of individual learning goals, which are central to the assessment process:

> Self-evaluation leads to the establishment of goals. That is what evaluation is for. We evaluate in order to find out what we have learned so we will know what to study next. People who self-evaluate constantly ask themselves, "Where am I going? Am I getting there? Am I getting somewhere? Am I enjoying the trip? Is this worthwhile? Do I approve of the way I'm spending my time?" (p. 37)

The student learning log is another useful tool for student self-assessment. When using this technique, the student records his perception of his progress relative to the standards and benchmarks set by the school or district. An example of a student log appears in Figure 4.6.

My evaluation of my understanding of the Lincoln/Douglas debate	My evaluation is based on the following evidence
I have a fairly good grasp of the /Douglas debate. I would rate myself as very ent in this topic.	My oral report on the debate included information that neither Lincoln nor Douglas specifically said. I had to take what they actually did say and combine it with information I knew about that particular situation to come up with some new ideas about the debate.

Figure 4.6. Sample Student Log.

As Figure 4.6 illustrates, the student provides his self-evaluation and the evidence to back that evaluation. In this example, he lists information regarding his declarative knowledge about the Lincoln/Douglas debate. This is followed by his self-evaluation and evidence to support the evaluation. This log might be used by the student later at an assessment conference with the teacher.

There are some parents, and a few educators as well, who doubt that student self-assessments are valid, because they suspect that students will tend to overrate their own understanding and skill. Those who have used student self-assessments extensively, however, have seen that these fears are unfounded. For example, Linda Darling-Hammond, Jacqueline Ancess, and Beverly Falk (1995) have reported a Aclear-headed capacity@ of their students to evaluate their own work (p. 155). Middle school teachers Lyn Countryman and Merrie Schroeder (1996) report that students' honesty and straightforwardness in self-assessment was remarkable to parents. One mother made the following comment after hearing her child=s self-assessment: "I feel our child was more honest with us than most teachers would be" (p. 68). Another parent remarked, "Students seem more open and honest about their performance. I didn't get the sugar-coated reports from advisors who tend to present negative aspects in a positive manner" (p. 68).

Assessment Conferences

Eventually, the classroom teacher must draw together the different types of assessment data that have been collected. We believe that a conference between student and teacher can greatly aid this process. This conference is designed to allow the teacher to share the assessment data she has gathered on the student, and for the student to show the teacher his own self-assessment. Assessment specialist Doris Sperling (1996) and curriculum theorist David Hawkins (1973) call

this student/teacher interaction collaborative assessment. Wiggins (1993a) notes that the very word Aassess@ indicates a collaborative method in its etymological root:

> The etymology of the word *assess* alerts us to this clinical C that is, client-centered C act. *Assess* is a form of the Latin verb *assidere*, to "sit with." In an assessment, one "sits with" the learner. It is something we do *with* and *for* the student, not something we do to the student. The person who "sits with you" is someone who "assigns value" C the "assessor" (hence the earliest and still current meaning of the word, which relates to tax assessors). But interestingly enough, there is an intriguing alternative meaning to that word, as we discover in *The Oxford English Dictionary*: This person who "sits beside" is one who "shares another's rank or dignity" and who is "skilled to advise on technical points." (p. 14)

In our opinion, Wiggins' comments encapsulate the essence of effective assessment in which teacher and student are collaborating to analyze the student=s strengths and weaknesses regarding particular learning outcomes.

The components of a student/teacher assessment conference are neither new nor complex. These conferences have been incorporated into the whole language movement for over twenty years (see Atwell, 1987; Calkins, 1986; Cazden, 1986; Hansen, 1987; Staton, 1980; Thaiss, 1986; Valencia, 1987; Young & Fulwiler, 1986). In brief, an assessment conference consists of the teacher first sharing her evaluation of a student's performance in regard to a particular standard or standards, as well as the evidence (e.g., quizzes, projects, and observations) that she used to reach her conclusions. In a similar fashion, the student then presents his self-evaluation and the supporting evidence. If the teacher utilizes a particular scale to rate performance on specific skills, then the student evaluates himself according to the same scale. Any discrepancies that exist between the teacher's and student's ratings on certain standards or benchmarks are subsequently discussed in detail so that the most accurate conclusion can be reached about the student's understanding and skill.

Step 3. Organize Your Grade Book Around Standards

If standards-based record-keeping and assessment are to take place, it is essential that the teacher organize his grade book around standards. The easiest way to organize a standards-based grade book is to allocate the columns in that book to standards rather than to assignments, tests, and activities. Consider Figure 4.7 to see how this might look.

Note that this particular book has room for six standards, which could be expanded to 12 standards if there is a fold-out page. Room is reserved at the top of the page under the heading Aassessment key.@ Here the teacher keeps track of individual assessments, activities, and homework assignments. There are seven items listed in our sample grade book in the assessment key:

A.	Homework:	September 7
B.	Quiz:	September 9
C.	Performance Task:	September 14
D.	Quiz:	September 16
E.	Homework:	September 21
F.	Performance Task:	September 23
G.	Unit Test:	September 25

Notice that in this marking period there were two graded assignments, two quizzes, two performance tasks, and a unit test. These assessments are linked to six standards, which cover the following content:

Standard #1:	percolation
Standard #2:	soil
Standard #3:	bar graphs
Standard #4:	hypothesis testing
Standard #5:	working with groups
Standard #6:	oral presentations

Also note that each student=s self-assessment is included within row *K*. The teacher entered this student=s self-assessment score into the grade book at the aforementioned student/teacher assessment conference.

Grade Book

Assessment Key:	1. Homework: Sept 7 2. Quiz: Sept 9 3. Perf. Task: Sept 14 4. Quiz: Sept 16	5. Homework: Sept 21 6. Perf. Task: Sept 23 7. Unit Test: Sept 25 8. _____	I. _____ J. _____ 11. Student Self-Assessment 12. Observations

Standards: Students:		#1 Understands percolation	#2 Understands soil information	#3 Designs and uses bar graph	#4 Generates and tests a hypothesis	#5 Contributes to groups	#6 Makes an oral presentation
Carmen Adams	A	4	4				
	B	3	4	4			
	C	4					
	D	4	4	3			
	E	4	3	4			
	F	3		4			4
	G	4	4	4			
	H						
	I						
	J						
	K	4	4	4	3	3	4
	L	4	4	4	4	4	4
		4 3	4 4	4 4	4 4	4 4	4
James Barton	A	2	2				
	B	2	3	2			
	C	3					
	D	2	2	2			
	E	2	3	2			
	F	1		3			1
	G	2	3	2			
	H						
	I						
	J						
	K	3	3	3	3	3	3

	L	2 1 2	3 2 3	2 2 3	2 1 1	2 2 2	2 2
Michael Caruso	A	3	3				
	B	3	3	3			
	C	4					
	D	2	2	4			
	E	3	3	4			
	F	3		4			4
	G	3	3	3			
	H						
	I						
	J						
	K	3	4	3	3	4	3
	L	3 3 4	3 3 3	4 4 4	3 3 3	4 4 3	4 4

Note: From R. J. Marzano and J. S. Kendall. (1996). *A Comprehensive Guide to Designing Standards-Based Districts, Schools, and Classrooms*. Copyright 8 1996 by McREL Institute. Reprinted with permission.

Figure 4.7. Sample Unit: Grade Book.

To understand how a teacher might assign scores to individual assignments, consider the entries for James Barton. In the box under standard 2 (information about soil) there are rows, each of which is preceded by a letter. These letters represent the assignments described in the assessment key at the top of the grade book. We see from Figure 4.7 that five assignments addressed the content about soil in standard #2. Those assignments were: the homework on September 7 (see A), the quiz on September 9 (see B), the quiz on September 16 (see D), the homework on September 21 (see E), and the unit test on September 25 (see F). In addition to the five assigned assessment activities, the teacher obtained and recorded James= assessment of himself on this standard (see K). Finally, the teacher made and recorded two informal observations of James= performance relative to this standard (see L).

It is also important to note that assessments commonly covered more than one standard. Assessment A, for example, is a quiz administered on September 7, which provided assessment information on both standards 1 and 2.

Lastly, note that in this record-keeping system some standards might have many more entries than others. In this marking period, every assessment covered content which related to standard 1, a clear indication that the teacher was focusing intently on this standard.

Organizing a grade book in the fashion depicted in Figure 4.7 causes a significant change in teacher thinking because it requires that one consider each assignment in terms of which standards it covers. For instance, if a homework assignment addresses three standards, then the teacher makes three notations in the grade book C one entry for each standard C rather than

one all-encompassing score. Teachers who have adopted this approach report that it moves them to plan their assessments early and in detail, rather than simply assigning chapter questions at the end of a reading passage or constructing a quiz consisting entirely of forced-choice items. Instead, the teacher must constantly ask which standards he means to address, what assessment data he will gather, and how he will gather it.

The use of numbers representing levels of individual performance rather than points totaling the number of correct responses is another radical aspect of this record-keeping procedure. In Figure 4.7, all assessment entries are scored on a scale of 1 to 4, indicating that the teacher is using performance levels akin to those shown in Figures 4.8 and 4.9.

4	**Advanced Performance:** Demonstrates a thorough understanding of the important information, i to exemplify that information in detail and articulate complex relationships and distinctions
3	**Proficient Performance:** Demonstrates an understanding of the important information; is able to exemplify that information in some detail
2	**Basic Performance:** Demonstrates an incomplete understanding of the important information, bu not have severe misconceptions
1	**Novice Performance:** Demonstrates an incomplete understand along with severe misconceptions

Figure 4.8. General Scale for Performance on a Declarative Benchmark.

4	**Advanced Performance:** Carries out the major processes/skills inherent in the procedure with relative ease and automaticity
3	**Proficient Performance:** Carries out the major processes/skills inherent in the procedure without significant error, but not necessarily at an automatic level
2	**Basic Performance:** Makes a number of errors when carrying out the processes and skills important to the procedure, but still accomplishes the basic purpose of the procedure
1	**Novice Performance:** Makes so many errors when carrying out the process and skills important to the procedure that it fails to accomplish its purpose

Figure 4.9. General Scale for Performance on a Procedural Benchmark.

It is important to note that Figure 4.8 is a rubric designed for declarative (i.e., informational) knowledge, while Figure 4.9 is intended to assess procedural knowledge. Look again at Figure 4.7 and assume that standard 2 about soil is primarily informational in nature. A quiz was administered on September 9 (see row B) which addressed this standard as well as standards 1 and 3. James= performance on that quiz indicated to the teacher that James Barton understood the important information appearing in the quiz about soil, and could explain that information in some detail. According to the rubric in Figure 4.8, James has demonstrated his ability to be at level 3 on a generic scale for declarative knowledge. Had James provided incomplete knowledge of soil but no major misconceptions, he would have scored at level 2. This is an example of how the generic rubric in Figure 4.8 can be adapted to cover specific declarative knowledge appearing in a quiz, homework assignment, or activity. The same can be done using the general rubric for procedural knowledge appearing in Figure 4.9. Consider the procedural knowledge inherent in standard 3 (designing and interpreting bar graphs). The September 9 quiz also assessed James' performance on this standard. In this case, the teacher judged that James had completed the basic procedure of reading or using bar graphs, but that he had made several errors in the process. This performance is concurrent with level 2 on the rubric.

The primary feature of this recording procedure is that the teacher makes a number of judgments during a grading period about each student's understanding and/or skills as they pertain to individual standards. This is accomplished by translating student performance on a test, activity, or homework assignment to a judgment of the degree of understanding or skill demonstrated by each student. When scoring a quiz, for instance, the teacher considers which items relate to each standard. She uses this data to make a judgment about each student's performance level relative to a specific standard, rather than simply adding up the number of correctly answered items that relate to a specific standard.

Some individuals, both in and outside the field of education, are suspicious of the role played by teacher judgment in this process because they assume that it incorporates a degree of subjectivity into grading. These skeptics fail to recognize that the traditional grading system grounded in the heuristic of Aadding up points@ is, by its very nature, subjective. Citing the work of fellow researchers (e.g., Ornstein, 1994), Guskey (1996b) asserts that the current method of assigning grades based on points Ais inherently subjective" (p. 17). Similarly, educator Carl Glickman (1993) notes that mainstream grading practices which rely on adding up points offer a false sense of objectivity because of the potentially complex manipulation of numbers used to Acalculate@ final grades.

A basic assumption of the grading process utilized here is that informed teacher judgment is a considerably more meaningful and accurate method of constructing grades. Wiggins (1993a) writes that:

> Judgment certainly does not involve the unthinking application of rules or algorithms C the stock in trade of all conventional tests. Dewey uses the words "knack, tact, cleverness, insight, and discernment" to remind us that judgment concerns "horse sense"; someone with good judgment is someone with the capacity to "estimate, appraise and evaluate." (Dewey adds, not coincidentally, "with tact and discernment.") The effective performer, like the good judge, never loses sight of either relative importance or the difference between the "spirit" and the "letter" of the law or rules that apply. Neither ability is testable by one-dimensional items, because to use judgment one must ask questions of foreground and background as well as perceive the limits of what one "knows." (pp. 219- 220)

Guskey (1996b) provides further evidence of the soundness and utility of teacher judgments. Citing the work of other researchers (e.g., Brookhart, 1993; O'Donnell & Woolfolk, 1991), Guskey explains:

> Because teachers know their students, understand various dimensions of students' work, and have clear notions of the progress made, their subjective perceptions may yield very accurate descriptions of what students have learned. (pp. 17-18)

At the end of a grading period, the teacher makes a judgment regarding each student=s performance on each standard. The teacher enters this score in the white box in the corner in each column of the grade book depicted in Figure 4.7. Consider James Barton=s overall score of 3 for standard 2 about soil. To compute this score, the teacher carefully considered how each assessment (i.e., each row) would contribute to the overall score. We believe that some entries should be more heavily weighted than others, and that this should be taken into account when calculating summary scores for a standard. Specifically, we are of the firm opinion that teachers should place great value on the student's self-assessment (entry *K* for each standard in the grade book). Many theorists and researchers (e.g., Conley, 1996; Fitzpatrick, Kulieke, Hillary, & Begitschke, 1996; Guskey, 1996b; Herman, 1996; Marzano, Pickering, & McTighe, 1993; McTighe & Ferrera, 1996; Mitchell, 1992; Mitchell & Neill, 1992; Spady, 1988, 1995; Wiggins, 1993a, 1993b, 1994) also suggest that the teachers place greater emphasis on more recent information than on other scores. Guskey (1996b) elaborates on the reasoning behind this:

> The key question is, "what information provides the most accurate depiction of students' learning at this time?" In nearly all cases, the answer is "the most current information." If students demonstrate that past assessment information no longer accurately reflects their learning, that information must be dropped and replaced by new information. (p. 21)

Thus, the overall score for a standard should represent the student=s knowledge and ability at the end of an instructional unit. Teachers should view scores recorded during the unit as pieces of information of varying degrees of importance.

Step 4. Assign Grades Based on Student's Performance on Standards

Most likely, a teacher will have to assign overall letter grades at some juncture in the semester or school year. Consider once more the report card in Figure 4.1. This report offers both a summary score for each standard and an overall subject grade. At some future date, report cards devoid of overall letter grades might find a place in the American culture. Currently, however, it is probably wise (in a political sense only) to assign overall letter grades to students even though individual standards scores offer ample information about student performance in specific standards.

Once a summary score has been assigned for each standard, the teacher can combine these standard level scores to compute each student=s overall letter grade. At this point, different weights can be applied to the individual standards. Figure 4.10 contains possible weights a teacher might assign to the standards in our sample unit.

Standard	Weight
plation	2
	1

Graphs	2
othesis Testing	1
king with Groups	2
Presentations	1

Figure 4.10. Weights Applied to Various Standards.

Notice that the teacher has weighted standards 1, 3, and 5, so that they have twice the influence on the overall grade as standards 2, 4, and 6. Such weights should be assigned to standards at the start of the grading period and the students made aware of them at that time. At the end of the grading period, the teacher would proceed to apply the weights to the summary standards scores as shown for the student Ashley Walker in Figure 4.11.

Student Name: Ashley Walker			
Standard	**Student Score**	**Weight**	**Quality Points**
1	3	2	6
2	3	1	3
3	3	2	6
4	1	1	1
5	3	2	6
6	2	1	2
	Totals	9	24

Figure 4.11. Computation of Total Quality Points for a Sample Student.

Ashley=s quality points for each standard are obtained by multiplying her overall score for a standard by the weight given to that standard. A student=s average standard score can be calculated using the following formula:

$$\frac{\text{Total Quality Points}}{\text{Total of Weights}}$$

As shown above, Ashley's quality points total 24. The total of the weights is 9. Determining Ashley's average score on the weighted standards is simply a matter of dividing her total quality points (24) by the total weight (9). Consequently, Ashley=s average standard score is 2.67, keeping in mind that some standards are weighted more heavily than others.

Next, the teacher would translate each student's average score to a letter grade. The following is a conversion scale which could be utilized to this end:

3.26 B 4.00	=	A
2.76 B 3.25	=	B
2.01 B 2.75	=	C
1.50 B 2.00	=	D
1.49 or below	=	F

By this scale, Ashley's average score of 2.67 would be translated into a letter grade as a C.

One might argue that the cutoff points between the grades seem arbitrary, which, in fact, they are. This is one of the major drawbacks to assigning overall letter grades. Guskey (1996b) stresses that this arbitrary quality of cutoff points is an inherent flaw of the overall grading system:

> The cutoff between grade categories is always arbitrary and difficult to justify. If the scores for a grade of B range from 80-89 for example, a student with a score of 89 receives the same grade as the student with a score of 80 even though there is a 9-point difference in their scores. But the student with a score of 79 C a l-point difference C receives a grade of C because the cutoff for a B grade is 80. (p. 17)

Guskey's comments also apply to the conversion system above. To illustrate, had Ashley received one more quality point on any of the six standards, her total score would have been 25 with an average of 2.78. This minor difference would have been enough for her to have been assigned a grade of B instead of a C.

It should be clear to the reader that we do not favor the use of overall grades to chart students' progress on standards. Measurement expert Richard Stiggins (1994) reminds us that a single symbol C in this case a letter grade C cannot reasonably report on all the complex learning that occurs in the classroom. Unfortunately, such grades are used in middle school and beyond in almost every American school district, so it is safe to assume that they will remain in use for quite some time. Therefore, if a teacher is in a school or district where overall letter grades are required for reporting students' performance on standards, then we suggest using the following guidelines:

1. Using well-informed judgment, assign scores for specific standards that present student levels of understanding and skill rather than individually scoring homework, quizzes, midterms, final exams, and so on, and then combining the scores.

2. For each course, provide a grading policy in writing in which you clearly outline how scores on standards will be weighted.

3. Clearly communicate to students and parents how standards are weighted and which standards influence the calculation of letter grades.

Although this approach is a compromise at best, the guidelines listed above will aid in making letter grades a more accurate reflection of students' performance on individual standards.

CHAPTER 5

CROSS-CUTTING ISSUES

Thus far, we have considered three general approaches to implementing standards: the use of external tests, the use of performance tasks and portfolios, and reporting out by individual standards. As mentioned previously, these approaches are not mutually exclusive. In fact, all three can be employed simultaneously. That is, a district or school could use a state test as a form of external assessment of students= performance on standards. In addition, the district or school could require students to complete performance tasks of their own design organized into portfolios. Finally, the district or school could also report student performance on standards in each course.

Regardless of the implementation model that is employed, there are a number of issues that a district must address. In this chapter we consider three of those issues.

The Issue of Levels

The issue of levels refers to the grade levels at which a district or school will hold students accountable for meeting specific standards. Theoretically, a district or school could be standards-based at every grade level. This would mean that students would not be allowed to pass from one grade to another without demonstrating competence in the standards and benchmarks specified at that level. Conversely, a district or school could be standards-based at high school graduation only. Within this approach, students would progress from grade level to grade level regardless of their performance on specific standards up until the 12th grade. However, at that point, demonstrated competence on specific standards would be a prerequisite to receiving a diploma. Finally, a district could be standards-based at the major transition points with the K-12 sequence of grades. Probably the most logical transition points are:

1. Between the primary and upper elementary grades

2. Between the upper elementary grades and middle school or junior high school

3. Between middle school or junior high and high school

4. At high school graduation

Here, students would be required to demonstrate competence in specific standards before they can pass from the primary level to the upper elementary level, from the upper elementary level to the middle school level, and so on.

One may infer that the approach with the least serious consequences is to be standards-based only at high school graduation, and the approach with the most severe consequences is to be

standards-based at each grade level. This latter position C being standards based at every grade level C seems extreme, particularly when one considers the research on grade-level retention.

Researchers Mary Lee Smith and Lorrie Shepard explain that there is a common sense belief that retaining students who have not demonstrated mastery information and skills at a particular grade level is actually advantageous to students:

> The assumption is that by catching up on prerequisite skills, students should be less at risk for failure when they go on to the next grade. Strict enforcement of academic standards at every grade is expected both to ensure the competence of high school graduates and lower the dropout rate because learning deficiencies would never be allowed to accumulate. (p. 84)

Unfortunately, this common-sense notion has been contradicted by virtually all of the research on retention (see Holmes, 1989; Grissom & Shepard, 1989; Shepard & Smith, 1989, 1990). Smith and Shepard note that the research on retention can be summarized in the following way:

- Students who are retained actually perform worse on average at the next grade level than those who have been promoted to the next grade.
- Dropouts are five times more likely to have repeated a grade than are high school graduates.
- Students perceive retention as a punishment.
- Retention generates a level of stress and a sense of failure that takes years to overcome.

It is our opinion that being standards-based at every grade level carries inordinate and unacceptable risks. In fact, we believe that the research against retention is so strong that a district or school should also be cautious even about being standards-based at major transition points.

The Option of Being Standards-Referenced

Being ***standards-referenced*** is an attractive option to being standards-based especially given the dangers inherent in retention. In a standards-based system, students must demonstrate that they have met the standards at one level before they are allowed to pass on to the next level. In a standards-referenced system, students= standings relative to specific standards are documented and reported; however, students are not held back if they do not meet the required performance levels for the standards.

Grant Wiggins (1993a, 1996) was perhaps the first modern-day reformer to recognize the utility of a standards-referenced approach. He reasoned that in addition to the inherent dangers of retaining students, it is unrealistic to expect all students to meet high standards in all content areas. For Wiggins, this type of nonthreatening Areferencing@ in and of itself may provide students with the motivation to reach levels of achievement to which they would otherwise not

aspire. Wiggins= assertion is based on the assumption that if students are presented with a goal (i.e., specific performance on standards) along with accurate information as to where they stand relative to the goal (i.e., their level of performance), they quite naturally may be motivated to improve their performance. This assumption is supported by much of the research on feedback (e.g., Glasser, 1981; Powers, 1973).

Wiggins= option can be used as a powerful implementation too. For example, if a district or school chose to be standards-based at the high school graduation level only, it could be standards-referenced at all other levels. Within such a system, students= progress on standards would be reported at each grade level. However, only at the level of high school graduation would students be required to meet specific standards. Similarly, if a district was standards-based at the four transition points described above, it could be standards-referenced at the other grade levels. This is depicted in Figure 5.1.

Mixing standards-based and standards-referenced approaches provides districts and schools with a wide range of options that retain the inherent power of holding students accountable for meeting certain standards, but alleviate the dangers inherent in retaining students at inappropriate levels.

Compensatory Versus Conjunctive Approaches

A final consideration a district or school should address is whether to use a *conjunctive* or a *compensatory* approach to standards. In a conjunctive approach, students must reach the minimum performance level on all standards (Plake, Hambelton, & Jaeger, 1995). To illustrate, consider the following mathematics standards:

1. Uses a variety of strategies in the problem-solving process

2. Understands and applies basic and advanced properties of the concepts of numbers

3. Uses basic and advanced procedures while performing the processes of computation

4. Understands and applies basic and advanced properties of the concepts of measurement

Grade 12	**SB**	Approach
11	**SR**	
10	**SR**	

9	**SR**	
8	**SB**	
7	**SR**	
6	**SR**	
5	**SB**	
4	**SR**	
3	**SR**	
2	**SB**	
1	**SR**	
K	**SR**	

SB = *Standards Based* (students are held accountable for meeting standards)
SR = *Standards Referenced* (students' progress is reported out at each grade level)

Figure 5.1. Options for Combining Standards-Based and Standards-Referenced Approaches.

5. Understands and applies basic and advanced properties of the concepts of geometry

6. Understands and applies basic and advanced concepts of statistics and data analysis

7. Understands and applies basic and advanced concepts of probability

8. Understands and applies basic and advanced properties of functions and algebra

If these standards were utilized in a conjunctive manner, a student's performance on each standard would be considered individually. For example, a student=s performance on mathematics standard 2 would be considered in isolation of her performance on the other seven standards. The student might do quite well on standards 2, 3, 4, and 8, yet do quite poorly on standards 1, 5, 6, and 7. Performance on one standard would have no bearing on performance on other standards.

In a compensatory approach, performance on one standard affects performance on others (Kifer, 1994). More specifically, performance on one standard can Acompensate@ for performance on another. To illustrate, assume that a student received the following scores (on a four-point scale) on the eight mathematics standards.

Standard 1: 1
Standard 2: 3
Standard 3: 3
Standard 4: 4
Standard 5: 2
Standard 6: 1
Standard 7: 1
Standard 8: 4

In a compensatory approach, the student=s strong performance on standards 2, 3, 4, and 8 would compensate for her weak performance on standards 1, 5, 6, and 7. Usually the compensation is accomplished by averaging the scores on specific standards within a domain. In the example above, the student=s average score on the eight mathematics standards would be 2.38. Other approaches include dropping the lowest scores from the average, weighting some standards higher than others in the calculation of the average, and considering the most common score (the mode) as the most representative score.

Conclusion

In this monograph, we have attempted to describe various models of standards implementation and some issues that characteristically must be considered when designing an implementation plan. It is our strong belief that the standards movement in this country will continue to grow. No longer will the question be asked AShould we implement standards?@ Rather, that question will be replaced by AHow will we implement standards?@ As this document has shown, there is no single best way of answering this question.

REFERENCES

Airasian, Peter W. (1994). Classroom assessment (2nd ed.). New York: McGraw Hill.

Anderson, J. R. (1982). Acquisition of cognitive skills. Psychological Review, 89, 369-406.

Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Anderson, J. R. (1990a). Cognitive psychology and its implications. New York: W. H. Freeman and Company.

Anderson, J. R. (1990b). The adaptive character of thought. Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J. R. (1993). Rules of the mind. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1995). Learning and memory: An integrated approach. New York: John Wiley & Sons.

Archbald, D. A., & Newmann, F. M. (1988). Beyond standardized testing: Assessing authentic achievement in the secondary school. Reston, VA: National Association of Secondary School Principals.

Atwell, N. C. (1987). In the middle. Portsmouth, NH: Heinemann.

Baker, E. L., Aschbacher, P. R., Niemi, D., & Sato, E. (1992). CRESST performance assessment models: Assessing content area explanations. Los Angeles, CA: National Center for Research On Evaluation, Standards, and Student Testing (CRESST), UCLA.

Baron, J. B. (1991). Strategies for the development of effective performance exercises. Applied Measurement in Education, 4, 305-318.

Baron, J. B., & Kallick, B. (1985). Assessing thinking: What are we looking for? And how can we find it? In A. Costa (Ed.), Developing minds: A resource book for teaching thinking. Alexandria, VA: Association for Supervision and Curriculum Development.

Berk, R. A. (1986a). Minimum competency testing: Status and potential. In B. S. Plate & J. C. Witt (Eds.), The future of testing (pp. 88-144). Hillsdale, NJ: Lawrence Erlbaum.

Berk, R. A. (Ed.). (1986b). Performance assessment: Methods and applications. Baltimore, MD: The Johns Hopkins University Press.

Berliner, D. (1984). The executive functions of teaching. The Instructor, 93(2), 28-40.

Berliner, D. C. (1979). Tempus Educare. In P. L. Peterson & H. J. Walberg (Eds.), Research on teaching. Berkeley, CA: McCutchan.

Bond, L., Friedman, L., & van der Ploeg, A. (1994). Surveying the landscape of state educational assessment programs. Washington, DC: Council for Educational Development and Research, and the National Education Association.

Brennan, R. L. (1983). Elements of generalizability theory. Iowa City, IA: American College Testing Program.

Brookhart, S. M. (1993). Teachers grading practices: meaning and values. Journal of Educational Measurement, 30(2), 123-142.

Calkins, L. M. (1986). The art of teaching writing. Portsmouth, NH: Heinemann.

Cannell, J. J. (1988). Nationally normed elementary achievement testing in America's public schools: How all 50 states are above the national average. Educational Measurement: Issues and Practices, 7(2), 5-9.

Carnevale, A. P., Gainer, L. J., & Meltzer, A. S. (1990). Workplace basics: The essential skills employers want. San Francisco: Jossey-Bass.

Cazden, C. B. (1986). Classroom discourse. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd ed., pp. 432-463). New York: Macmillan.

Center for Civic Education. (1994). National standards for civics and government. Calabasas, CA: Author.

Coleman, J. S. (1972). The evaluation of equality of educational opportunity. In F. Mosteller & D. P. Moynihan (Eds.), On equality of educational opportunity (pp. 140-161). New York: Vintage Books.

College Board, The. (1983). Academic preparation for college: What students need to know and be able to do. New York: College Entrance Examination Board.

Conley, D. T. (1996). Assessment. In R. E. Blum & J. A. Arter (Eds.), A handbook for student performance assessment in an era of restructuring (Sec. I-4, pp. 1-8). Alexandria, VA: Association for Supervision and Curriculum Development.

Consortium of National Arts Education Associations (1994). National standards for arts education: What every young American should know and be able to do in the arts. Reston, VA: Music Educators National Conference.

Costa, A. (1984). Mediating the metacognitive. Educational Leadership, 42, 57-62.

Countryman, L. L., & Schroeder, M. (1996). When students lead parent-teacher conferences. Educational Leadership, 53(7), 64-68.

Crafton, Linda K. (1996). Standards in practice: Grades K-2. Urbana, IL: National Council of Teachers of English.

Darling-Hammond, L., & Ancess, J. (1994). Graduation by portfolio at Central Park East Secondary School. New York: National Center for Restructuring Education, Columbia University.

Darling-Hammond, L., Ancess, J., & Falk, B. (1995). Authentic assessment in action: Studies of schools and student work. New York: Teachers College Press.

Dewey, J. (1916). Democracy and education. New York: Macmillan.

Dossey, J. A., Mullis, I. V. S., & Jones, C. O. (1993). Can students do mathematical problem solving? Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Doyle, W. (1992). Curriculum and pedagogy. In P. W. Jackson (Ed.), Handbook of research in curriculum (pp. 486-516). New York: Macmillan Publishing Company.

Durm, M. W. (1993). An A is not an A is not an A: A history of grading. The Educational Forum, 57(Spring), 294-297.

Durst, R. K., & Newell, G. E. (1989). The uses of function: James Britton's category system and research on writing. Review of Educational Research, 59(4), 375-394.

Farkas, F., Friedman, W., Boese, J., & Shaw, G. (1994). First things first: What Americans expect from public schools. New York: Public Agenda.

Feldt, L. S., & Brennan, R. L. (1993). Reliability. In R. L. Linn (Ed.) Educational measurement (3rd ed., pp. 105-146). Phoenix, AZ: Onyx Press.

Finn, C. E., Jr. (1990). The biggest reform of all. Phi Delta Kappan, 71(8), 584-592.

Fisher, C. W., Berliner, D. C., Filby, N., Marliave, R. S., Cahen, L. S., & Dishaw, M. M. (1980). Teaching behaviors, academic learning time and student achievement: An overview. In C. Denham & A. Lieberman (Eds.), Time to learn. Washington DC: National Institute of Education.

Fitts, P. M. (1964). Perceptual-motor skill learning. In A. W. Melton (Ed.), Categories of human learning. New York: Wiley.

Fitts, P. M., & Posner, M. I. (1967). Human performance. Belmont, CA: Brooks Cole.

Fitzpatrick, K. A., Kulieke, M., Hillary, J., & Begitschke, V. (1996). The instructional resource network: Supporting the alignment of curriculum instruction and assessment. In R. E. Blum & J. A. Arter (Eds.), A handbook for student performance assessment in an era of restructuring (Sec. iv-3, pp. 1-6). Alexandria, VA: Association for Supervision and Curriculum Development.

Fitzpatrick, R., & Morrison, E. J. (1991). Performance and product evaluation. In F. L. Finch (Ed.), Educational performance assessment (pp. 89-138). Chicago: Riverside Publishing Company.

Frederiksen, C. H. (1977). Semantic processing units in understanding text. In R. O. Freedle (Ed.), Discourse production and comprehension (Vol. 1, pp. 57-88). Norwood, NJ: Ablex.

Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. Educational Researcher, 18(9), 27-32.

Futrell, M. H. (1987, December 9). A message long overdue. Education Week, 7(14), 9.

Gandal, M. (1997). Making standards matter, 1997: An annual fifty-state report on efforts to raise academic standards. Washington, DC: American Federation of Teachers.

Geography Education Standards Project. (1994). Geography for life: National geography standards. Washington, DC: National Geographic Research and Exploration.

Glaser, R., & Linn, R. (1993). Forward. In L. Shepard, Setting performance standards for student achievement (pp. xiii-xiv). Stanford, CA: National Academy of Education, Stanford University.

Glasser, W. (1981). Stations of the mind. New York: Harper and Row.

Glickman, C. (1993). Reviewing America-s schools. San Francisco: Jossey-Bass.

Goodman, Y. M. (1978). Kid watching: An alternative to testing. National Elementary School Principal, 57, 41-45.

Grissom, J. B., & Shepard, L. A. (1989). Repeating and dropping out of school. In L. A. Shepard & M. L. Smith (Eds.), Flunking grades: Research and policies on retention (pp. 92-113). London: The Falmer Press.

Guskey, T. R. (1996b). Reporting on student learning: Lessons from the past C prescriptions for the future. In T. R. Guskey (Ed.), ASCD yearbook, 1996: Communicating student learning (pp. 13-24). Alexandria, VA: Association for Supervision and Curriculum Development.

Guskey, T. R. (Ed.). (1996a). ASCD yearbook, 1996: Communicating student learning. Alexandria, VA: Association for Supervision and Curriculum Development.

Hampel, R. L. (1986). The last little citadel. Boston: Houghton-Mifflin.

Hansen, J. (1987). When writers read. Portsmouth, NH: Heinemann.

Hansen, J. (1994). Literacy portfolios: Windows on potential. In S. W. Valencia, E. H. Hiebert, & P. P. Afflerrbach (Eds.), Authentic reading assessment: Practices and possibilities (pp. 26-44). Newark, DE: International Reading Association.

Hawkins, D. (1973). I, thou, it: The triangular relationship. In C. Silberman (Ed.), The open classroom reader (pp. 25-40). New York: Random House.

Herman, J. (1996). Technical quality matters. In R. E. Blum & J. A. Arter (Eds.), A handbook for student performance assessment in an era of restructuring (Sec. I-7, pp. 1-6). Alexandria, VA: Association for Supervision and Curriculum Development.

Hirsch, E. D. (1997). The schools we need: Why we don't have them. New York: Doubleday.

Kendall, J. S., & Marzano, R. J. (1997). Content knowledge: A compendium of standards and benchmarks for K-12 education (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Hirsch, E. D., Jr. (1987). Cultural literacy: What every American needs to know. Boston: Houghton Mifflin.

Holmes, C. T. (1989). Grade-level retention effects: A meta-analysis of research studies. In L. A. Shepard & M. L. Smith (Eds.), Flunking grades: Research and policies on retention (pp. 45-75). London: The Falmer Press.

If you have brains, you might decide to skip this test. (1997, March 28). The Wall Street Journal.

Joint Committee on National Health Education Standards. (1995). National health education standards: Achieving health literacy. Reston, VA: Association for the Advancement of Health Education.

Kifer, E. (1994). Development of the Kentucky Instructional Results Information System (KIRIS). In T. R. Guskey (Ed.), High stakes performance assessment: Perspective on Kentucky's educational reform (pp. 7-18). Thousand Oaks, CA: Corwin Press Inc.

Knight, P. (1992). How I use portfolios in mathematics. Educational Leadership, 49(8), 71-72.

Lane, S., Liu, M., Ankenmann, R. D., & Stone, C. A. (1996). Generalizability and validity of a mathematics performance assessment. Journal of Educational Measurement, 33(1), 71-92.

Linn, R. (1994). Performance assessment: Policy promises and technical measurement standards. Educational Researcher, 23(9), 4-14.

Linn, R. L., & Gronlund, N. E. (1995). Measurement and assessment in teaching (7th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Marzano, R. J. (1990). Standardized tests: Do they measure general cognitive abilities? NASSP Bulletin, 74(526), 93-101.

Marzano, R. J. (1992). A different kind of classroom: Teaching with Dimensions of Learning. Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J., & Costa, A. L. (1988). Question: Do standardized tests measure cognitive skills? Answer: No. Educational Leadership, 45(8), 66-73.

Marzano, R. J., & Kendall, J. S. (1992). Unpublished Data. Aurora, CO: Mid-continent Regional Educational Laboratory.

Marzano, R. J., & Kendall, J. S. (1996). A comprehensive guide to designing standards-based districts, schools, and classrooms. Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J., Pickering, D. J., Arredondo, D. E., Blackburn, G. J., Brandt, R. S., & Moffett, C. A. (1992). Dimensions of learning: Teacher's manual. Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J., Pickering, D. J., & McTighe, J. (1993). Assessing student outcomes. Alexandria, VA: Association for Supervision and Curriculum Development.

McDonald, J. P., Smith, S., Turner, D., Finney, M., & Barton, E. (1993). Graduation by exhibition: Assessing genuine achievement. Alexandria, VA: Association for Supervision and Curriculum Development.

McMillan, James H. (1997). Classroom assessment: Principles and practice for effective instruction. Boston: Allyn & Bacon.

McREL Institute. (1994). A summary report of studies of the effectiveness of standards-based education in the Aurora Public Schools. Aurora, CO: Author.

McTighe, J., & Ferrera, S. (1996). Performance-based assessment in the classroom: A planning framework. In R. E. Blum & J. A. Arter (Eds.), A handbook for student performance assessment in an era of restructuring (Sec. I-5, pp. 1-9). Alexandria, VA: Association for Supervision and Curriculum Development.

Meyer, C. A. (1992). What's the difference between authentic and performance assessment? Educational Leadership, 49(8), 39-40.

Mitchell, R. (1992). Testing for learning: How new approaches to evaluation can improve American schools. New York: The Free Press.

Mitchell, R., & Neill, M. (1992). Criteria for evaluation of student assessment. Washington, DC: National Forum On Assessment.

Mullis, I. V. S., Owen, E. H., & Phillips, G. W. (1990). America's challenge: Accelerating academic achievement (a summary of findings from 20 years of NAEP). Princeton, NJ: Educational Testing Service.

Myers, M. (1997). Personal communication.

National Assessment of Educational Progress. (1992, March 31). Framework for the 1994 National Assessment of Educational Progress Mathematics Assessment. Washington, DC: Author.

National Association for Sport and Physical Education. (1995). Moving into the future: National standards for physical education: A guide to content assessment. St. Louis: Mosby.

National Center for History in the Schools. (1996). National standards for history: Basic edition. Los Angeles: UCLA, Author.

National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform. Washington, DC: Government Printing Office.

National Council for the Social Studies. (1994). Expectations of excellence: Curriculum standards for social studies. Washington, DC: Author.

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

National Council of Teachers of English and the International Reading Association. (1996). Standards for the English language arts. Urbana, IL and Newark, DE: Authors.

National Education Goals Panel. (1991). The National Education Goals Report: Building a nation of learners. Washington, DC: Author.

National Governors Association. (1996, March). 1996 National Education Summit Policy Statement. Washington, DC: Author.

National Research Council. (1996). National science education standards. Washington, DC: National Academy Press.

National Science Board Commission on Precollege Education in Mathematics, Science and Technology. (1983). Educating Americans for the 21st century. Washington, DC: National Science Board Commission.

National Standards in Foreign Language Education Project. (1996). Standards in foreign language learning: Preparing for the 21st century. Lawrence, KS: Allen Press, Inc.

Neill, Monty, & others (1997). Testing our children: A report card on state assessment systems. Cambridge, MA: National Center for Fair and Open Testing (Fairtest).

Newell, A., & Simon, H. A. (1972). Human problem solving. Englewood Cliffs, NJ: Prentice Hall.

Newmann, F. M., Secado, W. G., & Wehlage, G. G. (1995). A guide to authentic instruction and assessment: Vision, standards and scoring. Madison, WI: Wisconsin Center for Educational Research, University of Wisconsin.

Norman, D. (1969). Memory and attention. New York: John Wiley and Sons.

O'Donnell, A., & Woolfolk, A. E. (1991). Elementary and secondary teachers= beliefs about testing and grading. Paper presented at the annual meeting of the American Psychological Association, San Francisco.

Olson, L. (1995a, April 12). Standards times 50. Struggling for standards: An Education Week Special Report, 14-22.

Olson, L. (1995b, June 14). Cards on the table. Education Week, 23-28.

Ornstein, A. C. (1994). Grading practices and policies: An overview and some suggestions. NASSP Bulletin, 78(561), 55-64.

Payne, D. A. (1974). The assessment of learning. Lexington, MA: Heath.

Perkins, D. N. (1992). Smart schools: From training memories to educating minds. New York: The Free Press.

Plake, B. S., Hambelton, R. K., & Jaeger, R. M. (1995). A new standard-setting method for performance assessments: The dominant profile judgment method and some field-test

results. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Powell, A. G., Farrar, E., & Cohen, D. K. (1985). The shopping mall high school. Boston: Houghton-Mifflin.

Powers, W. T. (1973). Behavior: The control of perception. Chicago: Aldine.

Project 2061, American Association for the Advancement of Science. (1993). Benchmarks for science literacy. New York: Oxford University Press.

Public Agenda (1998). Reality check: The status of standards reform. New York: Author.

Quellmalz, E. S. (1987). Developing reasoning skills. In J. B. Baron & R. J. Sternberg (Eds.), Teaching thinking skills: Theory and practice. New York: W. H. Freeman.

Reckase, M. D. (1995). Portfolio assessment: A theoretical estimate of score reliability. Educational Measurement: Issues and Practices, 14(1), 12-14, 31.

Resnick, L. B. (1987a). Learning in school and out. Educational Researcher, 16(9), 13-20.

Resnick, L. B. (1987b). Education and learning to think. Washington, DC: National Academy Press.

Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O=Connor (Eds.), Changing assessments: Alternative views of aptitude, achievement and instruction (pp. 37-76). Boston: Kluwer Academic Press.

Rowe, H. (1985). Problem solving and intelligence. Hillsdale, NJ: Lawrence Erlbaum.

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

Shavelson, R. J., & Baxter, G. P. (1992). What we've learned about assessing hands-on science. Educational Leadership, 49(8), 21-25.

Shavelson, R. J., & Webb, N. M. (1991). Generalizability theory: A primer. Newbury Park, CA: Sage Publishing.

Shavelson, R. J., Gao, X., & Baxter, G. R. (1993). Sampling variability of performance assessments (CSE Tech. Rep. No. 361). Santa Barbara, CA: UCLA, National Center for Research in Evaluation, Standards and Student Testing.

Shavelson, R. J., Webb, N. M., & Rowley, G. (1989). Generalizability theory. American Psychologist, 44, 922-932.

Shepard, L. (1989). Why we need better assessments. Educational Leadership, 46(7), 41-47.

Shepard, L. (1993). Setting performance standards for student achievement. Stanford, CA: National Academy of Education, Stanford University.

Shepard, L. A., & Smith, M. L. (1989). Flunking grades: Research and policies on retention. London: The Falmer Press.

Shepard, L. A., & Smith, M. L. (1990). Synthesis of research on grade retention. Educational Leadership, 47(8), 84-88.

Sierra-Perry, Martha. (1996). Standards in practice: Grades 3-5. Urbana, IL: National Council of Teachers of English.

Sizer, T. (1985). Horace=s compromise. Boston: Houghton-Mifflin.

Smagorinsky, Peter. (1996). Standards in practice: Grades 9-12. Urbana, IL: National Council of Teachers of English.

Spady, W. G. (1988). Organizing for results: The basis of authentic restructuring and reform. Educational Leadership, 46(2), 4-8.

Spady, W. G. (1995). Outcome-based education: From instructional reform to paradigm restructuring. In J. H. Block, S. T. Everson, & T. R. Guskey (Eds.), School improvement programs (pp. 367-398). New York: Scholastic.

Sperling, D. (1996). Collaborative assessment: Making high standards a reality for all students. In R. E. Blum & J. A. Arter (Eds.), A handbook for student performance assessment in an era of restructuring (Sec. iv-11, pp. 1-6). Alexandria, VA: Association for Supervision and Curriculum Development.

State of Florida, Department of State (1996). Science: Pre K-12 sunshine state standards and instructional practices. Tallahassee, FL: Author.

Staton, J. (1980). Writing and counseling: Using a dialogue journal. Language Arts, 57, 514-518.

Stevenson, H. W., & Stigler, J. W. (1992). The learning gap: Why our schools are failing and what we can learn from Japanese and Chinese education. New York: Touchstone.

Stiggins, R. J. (1994). Student-centered classroom assessment. New York: Merrill.

Stodolsky, S. S. (1989). Is teaching really by the book? In P. W. Jackson & S. Haroutunian-Gordon (Eds.), Eighty-ninth yearbook of the national society for the study of education, (Part I, pp. 159-184). Chicago: University of Chicago Press.

Thaiss, C. (1986). Language across the curriculum in the elementary grades. Urbana, IL: ERIC Clearinghouse On Reading and Communication Skills, and National Council of Teachers of English.

Uchida, D., Cetron, M., & McKenzie, F. (1996). Preparing students for the 21st century. Reston, VA: American Association of School Administrators.

Valencia, S. (1987, April). Novel formats for assessing prior knowledge and measures of reading comprehension. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

van Dijk, T. A. (1980). Macrostructures. Hillsdale, NJ: Lawrence Erlbaum.

Wiggins, G. (1989). Teaching to the (authentic) task. Educational Leadership, 46(7), 41-47.

Wiggins, G. (1991, February). Standards, not standardization: Evoking quality student work. Educational Leadership, 48(5), 18-25.

Wiggins, G. (1993a, November). Assessment, authenticity, context and validity. Phi Delta Kappan (pp. 200-214).

Wiggins, G. (1994). Toward better report cards. Educational Leadership, 52(2), 28-37.

Wiggins, G. (1996). Honesty and fairness: Toward better grading and reporting. In T. R. Guskey (Ed.), ASCD yearbook 1996: Communicating student learning (pp. 141-177). Alexandria, VA: Association for Supervision and Curriculum Development.

Wiggins, G. P. (1993b). Assessing student performances: Exploring the purpose and limits of testing. San Francisco: Jossey-Bass.

Wilde, S. (Ed.). (1996). Notes from a kid watcher: Selected writings of Yetta M. Goodman. Portsmouth, NH: Heinemann.

Wilhelm, Jeffrey D. (1996). Standards in practice: Grades 6-8. Urbana, IL: National Council of Teachers of English

Willis, S. (1996, March). Student exhibitions put higher-order skills to the test. Education Update, 38(2), 1-3.

Winograd, P., & Perkins, F. D. (1996). Authentic assessment in the classroom: Principles and practices. In R. E. Blum & J. A. Arter (Eds.), A handbook for student performance assessment in an era of restructuring (Sec. I-8, pp. 1-11). Alexandria, VA: Association for Supervision and Curriculum Development.

Yoon, B., Burstein, L., & Gold, K. (undated). Assessing the content validity of teachers' reports of content coverage and its relationship to student achievement (CSE Rep. No. 328). Los Angeles: Center for Research in Evaluating Standards and Student Testing, University of California, Los Angeles.

Young, A., & Fulwiler, T. (Eds.). (1986). Writing across the disciplines. Portsmouth, NH: Heinemann.