How to Assess Student Performance in Science

Going Beyond Multiple-Choice Tests

How to Assess Student Performance in Science:

Going Beyond Multiple-Choice Tests

Third Edition, 2005
Second Edition, 2000
First Edition, 1995

SERVE Center
at the University of North Carolina
at Greensboro

Associated with the
School of Education,
University of North
Carolina at Greensboro

PRODUCED BY

The SERVE Center at the University of North Carolina at Greensboro

WRITTEN BY

Susan M. Butler, Ph.D
Wendy McColskey, Ph.D.
Rita O'Sullivan, Ph.D.

EDITED BY

Donna Nalley, Ph.D., Director of Publications, SERVE

DESIGNED BY

Tracy Hamilton, Assistant Program Specialist, SERVE

TABLE OF CONTENTS

Acknowledgments

This publication, as the title implies, is about going beyond multiple-choice tests in order to assess student achievement. Why is this necessary? What has impelled the writing of this manual? To answer these questions, we first must take a look at how ideas about science teaching have changed. So, we begin this manual with a small history lesson in the development of science curricula.

Brief History of Science Curriculum Development

Welch (1979) characterized the social forces leading to science education reform of the 1960s as scientists' concern about outdated curricular materials, science manpower shortages, and the threat of Soviet technological supremacy. These forces set the stage for massive federal support for science curriculum development.

For approximately 20 years, the National Science Foundation supported extensive curriculum development and teacher inservice training programs in science education. Their curricula differed from old programs in its modernization of content, its emphasis on flexibility and variety in instructional tools, and the greater attention it gave to an overriding conceptual scheme, students' attitudes toward science, and the nature of scientific inquiry or hands-on student work.

In spite of all the support for curricular change over this period, there were also forces resistant to change, including the following:

- Many teachers were inadequately prepared in science and math, particularly at the elementary and middle school levels, and were insecure about making curricular changes.

- Concern in the 1970s focused more on special remedial classes, the basic skills, and mainstreaming than on science.

Welch (1979), in summarizing the achievements of the curriculum reform of the 60s and 70s, reported:

- Curricular alternatives were developed and disseminated (PSSC—Physical Science Study Committee; BSCC—Biological Sciences Curriculum Study; SCIS—Science Curriculum Improvement Study).

- Content was updated.

- New curricular materials emphasized science processes and hands-on work.

- Science manpower needs were met.

The reform of the 1990s and beyond differed from the earlier science curriculum reform in that it was a subset of a much larger educational reform movement fueled by a concern that our students would not be internationally competitive as adults. Changes were proposed across the curriculum, emphasizing higher-order thinking skills and problem-solving.

Science For All

The emphasis on science education in previous decades that resulted in the development of curriculum materials provided a framework on which the 1990s' efforts built. However, the 1990s differed from prior curricular reform movements in that they were geared toward scientific literacy for all students (National Research Council, 1999), not just better science education for future scientists. Such literacy is critical if the general public is to have a basis for making informed decisions about issues like nuclear power, personal health, the environment, reproduction (Loucks-Horsley, Brooks, Carlson, Kuerbis, Marsh, Padilla, Pratt, & Smith, 1990), and stem cell research.

Continuing this emphasis on science for all students, Project 2061, a reform effort of the American Association for the Advancement of Science issued a 1989 report called *Science for All Americans* (**www.project2061.org/publications/ sfaa/online/sfaatoc.htm**). This report suggested the knowledge, skills, and attitudes that students should have as a result of their K–12 science instruction. The "science for all" theme was also evident in the National Science Education Standards (NSES), produced and distributed by the National Research Council in 1995 and available online at **www.nas.edu**. In the NSES document, the National Research Council (1999, p. 2) states, "The intent of the *Standards* can be expressed in a single phrase: Science standards for all students…The *Standards* apply to all students, regardless of age, gender, cultural or ethnic background, disabilities, aspirations, or interest and motivation in science." This "science for all" orientation has most recently been reflected in the Elementary and Secondary Education Act of

2001, better known as the "No Child Left Behind Act." A re-emphasis on science testing as part of school accountability is also part and parcel of the "science for all" orientation nature of this Act.

Science Inquiry

"Science for all" is not the only theme emerging in science education. One can also track the development of an emphasis on science inquiry. The National Science Teachers Association (Texley & Wild, 1997, p. 62) notes that the National Science Education Standards marks a move "away from presenting information to encouraging student discovery." Tobin, Kahle, and Fraser supported this move away from content presentation to a more inquiry-based approach. In *Windows into Science Classrooms,* they noted (1990, p.151):

> If an instructional activity is to be consistent with the nature of science, it must engage students in attempting to generate answers to questions, rather than merely illustrating what is pronounced by assertion to be true in a textbook. When laboratory activities or demonstrations are used to illustrate the validity of what is known, the emphasis is placed disproportionately on what we think we know rather than on how we know it. In such situations, students are deprived of opportunities to think, predict, analyze, and discuss; that is, they are deprived of opportunities to *do* science (emphasis added).

The National Standards document also argues that students must *do* science (National Research Council, 1999, p. 2):

> "The Standards rest in the premise that science is an active process. Learning science is something that students do, not something that is done to them. 'Hands-on' activities, while essential, are not enough. Students must have 'minds-on' experiences as well."

This document goes on to note: "when engaging in inquiry, students describe objects and events, ask questions, construct explanations, test those explanations against current scientific knowledge, and communicate their ideas to others. They identify their assumptions, use critical and logical thinking, and consider alternative explanations. In this way, students actively develop their understand-

ing of science by combining scientific knowledge with reasoning and thinking skills."

In other words, "minds on" means that both students and their teachers need to pay attention to the quality and sophistication of student thinking. For example, teachers may need to examine the quality of students' efforts to draw conclusions from data and determine what next instructional steps are needed to improve this particular thinking skill.

Changing Assessment Practices

Because of the changing emphases in science education, traditional assessment practices must also undergo a metamorphosis. The impetus for students to do science fuels an impetus for teachers to find new methods of assessment; methods that allow them to track student progress toward the inquiry-based standards of science education that emphasize the quality of student thinking and student products. We are living in an era where the accumulation of facts is less important than the ability to manipulate or apply knowledge. Therefore, we can no longer rely solely on multiple-choice, fact-based testing. We must develop and use assessment methods appropriate to our higher expectations of students. This manual is intended to aid teachers in such development activities. It is written in response to the following statement, taken from *Inside the Black Box* by Black and Wiliam (1998, pp. 15–16):

> Teachers will not take up attractive sounding ideas, albeit based on extensive research, if these are presented as general principles, which leave entirely to them the task of translating them into everyday practice—their classroom lives are too busy and too fragile for this to be possible for all but an outstanding few. What they need is a variety of living examples of implementation, by teachers with whom they can identify and from whom they can both derive conviction and confidence that they can do better, and see concrete examples of what doing better means in practice.

This manual attempts to provide some "living" and "concrete" examples that will aid teachers in developing new assessment methods and encourages teachers to work together in doing so. The manual is particularly timely, in that assessment of science achievement is mandated in the No Child Left Behind Act`.

Current Views on Assessment

Introduction

Educational systems promote student growth in a variety of dimensions. Basic content knowledge can be effectively assessed with multiple-choice and completion tests. However educational reforms have become more concerned with higher-order cognitive dimensions (problem-solving, creativity), social dimensions (communication skills, ability to work in groups) and other dimensions (life-long learning). While they are objective and efficient, traditional assessment measures may not serve these kinds of goals as well as other types of measures. Before we can choose an accurate, efficient method of assessment, we must clearly understand the goals of science instruction. Do these goals encompass only the basic memorization of facts? If so, our traditional methods may be sufficient. If we wish to institute a science program that encourages dimensions that go beyond these basics, we will need to develop a repertoire of additional assessment methods. The organization of this manual is intended to aid teachers in developing expertise in identifying learning goals, choosing assessment methods, and communicating assessment results in such a way that student performance is enhanced.

Identifying Learning Goals

Let us begin by looking at the goals of science instruction. Only by clearly defining what we want students to know and be able to do can we then choose and plan effective assessments that accurately measure student achievement of these goals.

According to McTighe and Wiggins (2004), the true issue being debated by assessment reformers is not whether some assessment methods are superior to others, but rather what is worth assessing, given limited assessment time. The debate about assessment, then, is a "value" debate. What goals or outcomes do we value for students? Kohn (1999, p. 216) expresses this "value" idea as "Content: Things Worth Knowing" and suggests that "a good deal of what students are required to do in school is, to be blunt, not worth doing." By discussing the relative value we place on particular goals or outcomes, assessment experts are encouraging assessment reform. They direct curriculum developers and teachers to examine the curriculum itself and to ensure that goals of learning are clearly expressed, relevant to students, frequently challenging, and properly assessed.

> If the goal is for students to learn basic facts and skills, then paper-and-pencil tests and quizzes generally provide adequate and efficient measures. However, when the goal is deep understanding, we rely on more complex performances to determine whether our goal has been reached. (McTighe & Wiggins, 2004, p. 141)

CHAPTER ONE

It is important to remember that you are making choices about assessment right now. These choices may be constrained by what you have always done, what others think you should do, what you understand about assessment, or what you feel students expect you to do, but they are choices nonetheless. This manual is designed to provide you with the support you and other teachers at your school need to begin a process of defining the outcomes you value for students in science and developing assessment practices that encourage student progress toward desired ends. **CHAPTER 2** provides background knowledge and practical advice to help you set instructional goals.

This chapter also reminds us that how and what we test sends a clear message about what is valued. Traditionally, we have almost exclusively valued students' success at retaining and bringing forth a sample of the information they have internalized. When a teacher only emphasizes factual knowledge on tests, students conclude that remembering facts is the goal. When students are not given an opportunity to retest or improve their work, they may conclude that improvement is not valued. If higher-order thinking, problem-solving, and critical thinking are to be valued, then classroom assessments need to lend value to them. It is imperative for us to know our goals before choosing assessment methods.

Choosing Assessment Methods

Once instructional goals have been identified, assessment planning can begin. It is important to match the assessment to the learning goal to ensure that the assessment can accurately measure the goal. For example, let us suppose an instructional goal states, "Students will accurately, competently, and safely use scientific equipment." Is a multiple-choice test the best assessment method to choose for this goal? Wouldn't it be better to have students *demonstrate* the use of scientific equipment (Bunsen burners, microscopes, wave tables) if we wish to ascertain their competency for this task? Such demonstrations are often termed "student performances," and assessment methods used to judge them can be called "performance assessments." According to Wiggins (1989), such assessments require the following:

- Tests should involve real-life tasks, performances, or challenges that replicate the problems faced by a scientist, historian, or expert in a particular field; thus, they are complex tasks rather than drills, worksheets, or isolated questions.

- Students should understand up-front the criteria on which their work will be judged and be able to apply the criteria to their work.

- Students should be asked to demonstrate their control over the essential knowledge being taught by actually using the information in a way that reveals their level of understanding.

Others argue that performance assessments should

- Require students to perform tasks that include the highest skill levels of problem finding and solving to include role-playing, "real-life" simulations, investigation, major projects, and creative depictions. (Wiggins, 1992; Glatthorn, 1998)

- Use power verbs (such as *research*, *analyze*, *evaluate*, and *depict*) to reinforce that the student is demonstrating what he or she can do with information. (National Research Council, 1999)

- Where appropriate, allow students to be involved in creating the criteria against which their performance will be judged. (Stiggins, 2001)
- Include audiences in addition to the teacher to validate and judge student performances (e.g., scientists, other students). (Kohn, 1999)

Kohn also introduces the need for assessment to be reality-based; that is, based upon work that is done in the "real" (as opposed to the educational) world. Kohn expresses this as: "Thus, our question is not merely, *What's the task*? But, *How does the task connect to the world that the students actually inhabit*?" And he reminds us: "Children are people who have lives and interests outside of school, who walk into the classroom with their own perspectives, points of view, ways of making sense of things and formulating meaning. What we teach and how we teach must take account of these realities" (1999, p. 219).

How do we infuse our teaching and our assessments with "reality"? What is a "real-world" task? A few examples of generic kinds of tasks that have students *using* or *applying* information in ways that go beyond just recalling or recognizing correct information include the following:

- Leading a group to closure on an issue
- Collecting, analyzing, and interpreting data about the success of a program, product, or event
- Researching both sides of a controversy and reporting it objectively
- Developing criteria for rating the quality of a product, proposal, or recommendation

Such tasks are recognizable as part of many adult work environments and can be infused into the work demanded of students. For an academic, science-related example of an assessment involving a real-world task, see **FIGURE 1.1**.

CHAPTER 3 of this manual builds upon the idea of implementing performance-, authentic- and reality-based assessment by examining several different assessment methods that go beyond multiple-choice testing. These assessment methods include those found in **FIGURE 1.2**. Chapter 3 attempts to provide clear definitions of different assessment types and then suggests ways each type could be used in the science classroom.

Why do teachers need such a diverse toolbox of assessment methods that go beyond multiple-choice testing? To answer this question, let's first examine a typical, traditional classroom scenario.

> **SCENARIO:** The teacher teaches a unit on soil formation and then gives a unit test with multiple-choice, short-answer, and matching items to assess students' retention of the information. Students are told about the test one week in advance, and they bring no resource materials with them to the test. Students' tests are scored and returned and form the basis of the six weeks' grade.

Proponents of assessment reform argue that past assessment practices (as the ones depicted in the above scenario) are inadequate.

Glatthorn (1998, p. 8) characterizes such scenarios as "teaching to the test" and offers this classroom illustration:

> Students will have to take a short-answer objective test assessing their knowledge of the legislative process as employed in their state. A typical question asks students to define *bill* and *law*. The specific content of the test is confidential, with the test administered under conditions of high security. The teacher has identified the questions the test is likely to ask by reviewing previous editions of the test. The teacher prepares practice material on test-like items. Students spend most of their class time completing the practice exercises and checking their answers.

It is not that objective, fact-based tests are not important. As we stated at the beginning of this chapter, such tests are effective and efficient means of measuring basic knowledge. However, such fact-based exams should not be the only type of assessment method used by the teacher.

Fundamental problems with such fact-based, traditional assessment practices include:

- Narrowness of scope
- Limited expectations of students
- Overemphasis on memorizing isolated facts, rather than concentrating on connections and relationships
- Lack of student ownership in the learning process
- Lack of incentives for student improvement in their work

CHAPTER 3 is included in this manual to give teachers expanded alternatives to traditional assessment practices.

FIGURE 1.1
Sample Assessment Utilizing a Real-Life Task

ASSIGNMENT:

Research with your team the value and uses of whales across time and cultures. Analyze and evaluate the practical uses vs. environmental protection issues, and develop support for both. Choose a position and be prepared to justify and present your position to the class in a convincing manner.

ASSESSMENT METHODS:

1. Research quality will be assessed through teacher observation of teamwork and teacher review of a team journal of completed group work.

Teams are not allowed to proceed with developing their presentations until they can show they have adequately researched the topic.

2. Oral presentation skills will be assessed by peers and teachers using a rubric.

Source: Adapted from High Success Network training materials. Outcome-Based Education Summer Conference, Charlotte, NC, 1992; High Success Network, P.O. Box 1630, Eagle, CO 81631.

FIGURE 1.2
Assessment Methods

Observe students using
- Informal observations
- Structured observations

Soliciting information from students through
- Interviews
- Self-assessment questionnaires

Evaluate students' work using
- Open-ended questions
- Performance tasks
- Journals
- Exhibitions and culminating demonstrations (i.e., science fair projects)
- Portfolios

Communicating Assessment Results

After instructional goals are set and assessments are performed, teachers need to communicate assessment findings to students and to parents. **CHAPTER 4** provides an overview of assessment instruments and grading schemes that can provide timely and essential feedback to learners. The chapter begins with an overview of the types of rubrics (scoring guides) available to teachers and the effectiveness of each type. The chapter then briefly discusses grades and grading as a way to communicate student progress toward learning goals.

Using This Manual

As you read this publication, the authors hope you will:

- Consider the variety of possible student outcomes in science, and select those that are most important for students.

- Reflect on and choose appropriate ways to assess student performance for important outcomes.

- Develop appropriate criteria for judging student work, and consider the alternatives to the teacher as sole judge of student work (i.e., using peers, professionals from the community, and student self-assessment).

- Reflect on grading practices and how information from a variety of assessment methods might be incorporated into a composite picture of achievement.

- Consider ways to get yourself and your school started in analyzing current practices.

This publication is not intended as a text but as a self-study resource. We hope you will interact with it, respond to the questions posed, and use the manual as an opportunity to reflect on your assessment practices. We suggest that you work through the manual with at least one other teacher, if possible, because of the valuable sharing of ideas that will result.

Final Notes

A key point to remember as you go through this manual is that the way we assess our students speaks volumes about what we value in education. If throughout 12 years of school, students are assessed only on passive, non-creative work (worksheets, multiple-choice tests), how likely is it that they will become problem-solvers, creative producers, effective communicators, and self-directed learners as adults? By going beyond multiple-choice testing, we hope to foster these qualities in our students.

The first step in changing science education assessment is to have a clear understanding of your current practices. Please answer the following questions and discuss them with another teacher.

Self-Assessment Questionnaire

1. **List below, in your own terms, the four most important student outcomes that resulted from your science instruction last year. That is, what could students do well at the end of the year that they could not do well at the beginning of your instruction?**

2. **Which of the following kinds of work did you require of students?**

    ❑ Listen to lectures
    ❑ Take tests on text/lectures
    ❑ Take end-of-chapter tests
    ❑ Design experiments
    ❑ Read textbooks
    ❑ Talk with scientists
    ❑ Solve problems in a team setting
    ❑ Maintain journals of data collected
    ❑ Do hands-on investigations
    ❑ Make presentations to the class
    ❑ Other

3. **In your science classes, on a typical day, how often were most students engaged and challenged by their work?**

    ❑ All the time
    ❑ Very often (more than half the time)
    ❑ Often (about half the time)
    ❑ Somewhat often (less than half the time)
    ❑ Almost never

4. **Think about the assessment methods represented by the grades in your grade book. What might your grade book say to students about what you value in science education?**

DESIRED STUDENT OUTCOMES IN SCIENCE:

What Do We Want Students To Be Able To Do?

Educational goals provide the framework for assessing student progress. The goals a teacher has for his or her class have clear implications for assessment. Without a clear vision or articulation of what is to be accomplished in the time you have with your students, how do you know what to assess? Outlining your goals before beginning instruction is very important.

The National Science Education Standards publication (1999, p. ix) written by the National Research Council begins with this goal statement: "This nation has established as a goal that all students should achieve scientific literacy." This booklet goes on to describe such literacy as "Scientific literacy enables people to use scientific principles and processes in making personal decisions and to participate in discussions of scientific issues that affect society" (1999, p. ix). With this description, the National Research Council begins to break its overall goal (scientific literacy) into smaller component parts, as depicted in **FIGURE 2.1**.

FIGURE 2.1
Scientific Literacy

Clearly, the National Research Council is emphasizing that students need to learn to think like scientists (use scientific processes) as well as learn science concepts (knowledge of scientific principles). To achieve the goal of scientific literacy, the Council (1999, p. 104) has written content standards (statements that elucidate what students should know and be able to do) within eight different categories:

- Unifying concepts and processes in science
- Science as inquiry

- Physical science
- Life science
- Earth and space science
- Science and technology
- Science in personal and social perspectives
- History and nature of science

In this organizational scheme of content standards written within the 8 different categories, all of which support the twin goals of scientific literacy (understanding of science concepts and science processes), the National Research Council has refocused science instruction on new facets. This change is summarized in chart form (National Research Council, 1999, p. 113) and reproduced here in **FIGURE 2.2**. Note particularly the thinking processes and processes related to scientific inquiry that have added emphasis in the newer science curriculum.

FIGURE 2.2
Changes in Emphasis in Science Instruction

LESS EMPHASIS ON	MORE EMPHASIS ON
Knowing scientific facts and information	Understanding scientific concepts and developing *abilities of inquiry*
Studying subject matter disciplines (physical, life, earth sciences) for their own sake	Learning subject matter disciplines in the context of *inquiry*, technology, science in personal and social perspective, and history and nature of science
Separating science knowledge and science process	Integrating all aspects of science content
Covering many science topics	Studying a few fundamental science concepts
Implementing inquiry as a set of processes	Implementing *inquiry* as instructional strategies, *abilities*, and ideas to be learned

Source: National Research Council. (1999). National science education standards. *Washington, DC: National Academy Press, p. 113.*

The National Science Education Standards' stated goal (scientific literacy), the eight categories in which standards are clustered, and even the statements found in the "More Emphasis On" column of **FIGURE 2.2**, all provide only very general guidelines for teachers. To further elucidate what students should know and be able to do, the National Research Council also provides content standards within the eight stated categories. In the next section of this chapter, we will examine a sample content standard for grades 9–12 and attempt to interpret or *unpack* this standard to find the specific learning goals for students.

Unpacking the Content Standards

There is a wealth of science knowledge and scientific abilities/processes that could be taught to students. In fact, the overabundance of teaching possibilities can be overwhelming for teachers, who wonder where to begin and how deep to go. Teachers often feel that they must "cover" everything in the textbook. The National Science Education Standards provide one means of managing the task of identifying the essential science information for students. These stan-

dards distill the amount of information into a smaller subset of essential information. However, even these standards are not totally transparent; it will still take some expertise to understand exactly what they are trying to convey. In this publication, we will unpack the National Standards using Bloom's Taxonomy as our framework.

Bloom's Taxonomy (Bloom, 1956) grouped educational objectives into six distinct, hierarchical categories: Recall, Comprehension, Application, Analysis, Synthesis, and Evaluation. In practice, the Recall and Comprehension categories gradually came to be grouped together as "Knowledge." Throughout the years, many teachers have used this taxonomy to ensure that varied levels of thinking were encouraged in their classrooms. Here, we use the levels of Bloom's Taxonomy to unpack the meaning of the National Standards.

Let's begin with Content Standard A under Science as Inquiry in Grades 9–12. This standard states:

CONTENT STANDARD A:

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

- Abilities necessary to do scientific inquiry
- Understandings about scientific inquiry (National Research Council, 1999, p. 173).

This standard clearly relates to the overall goal of the National Science Education Standards—to promote scientific literacy. Note that it emphasizes both science knowledge and scientific processes. It appears closely tied to the "Ability to use scientific processes" box shown in **FIGURE 2.1**. However, this is still a very general statement. It does not provide sufficient specificity for teachers to understand precisely what students will need to know or be able to do.

In order to further elucidate the meaning of Content Standard A, the National Research Council provided a section titled, "Guide to the Content Standard." In this section, they include the following six

RESOURCES

These resources may be helpful in "unpacking" standards:

Ainsworth, L. (2003). *"Unwrapping": A simple process to make standards manageable*. Englewood, CO: Advanced Learning Press.

National Research Council. (1999). *National science education standards*. Washington, DC: National Academy Press.

Reeves, D.B. (2004). *101 questions and answers about standards, assessment, and accountability*. Englewood, CO: Advanced Learning Press.

Reeves, D.B. (2004). *101 more questions and answers about standards, assessment, and accountability*. Englewood, CO: Advanced Learning Press.

Reeves, D.B. (2004). *Making standards work*. Englewood, CO: Advanced Learning Press.

Texley, J. & Wild, A. (1997). *NSTA pathways to the science standards: Guidelines for moving the vision into practice*. Arlington, VA: National Science Teachers Association.

underlying abilities and concepts related to the "abilities necessary to do scientific inquiry" statement:

1. Identify questions and concepts that guide scientific investigations

2. Design and conduct scientific investigations

3. Use technology and mathematics to improve investigations and communications

4. Formulate and revise scientific explanations and models using logic and evidence

5. Recognize and analyze alternative explanations and models

6. Communicate and defend a scientific argument (National Research Council, 1999, pp. 175–176).

We will call each of the above a benchmark that helps explain the meaning of the standard. In this next section, we shall attempt to unpack two of these benchmarks even further, using Bloom's Taxonomy.

1. Identify questions and concepts that guide scientific investigations.

Which levels of Bloom's Taxonomy are implied in this statement? The verbs that begin each benchmark often provide clues to the levels of Bloom's. For example, "identify" in the first benchmark sometimes correlates with lower-level thinking: the *Knowledge* level. Students are asked to identify concepts that guide investigations, not apply them or analyze them. Other verbs that would signal the *Knowledge* category may include: *define, list, tell, label, match, select, choose, name, spell*, etc. Returning to the benchmark, we find that certainly, students will need basic knowledge of scientific investigations. They will need to know, for example, that investigations contain certain parts as problem-finding, hypothesizing, designing an experiment, controlling variables, reporting conclusions, etc. They will need to understand each of these separate processes, view examples of each, and distinguish between high-quality and low-quality processes. Students will also need background knowledge before they can begin to create their own investigations. They will need to find out what is already known, what scientific concepts may govern their investigation, and what safety concerns should be considered. Therefore, this benchmark implies that students must have opportunities to gain such *Knowledge*.

This is not simply a *Knowledge*-based benchmark. Students are not asked to learn "about" scientific investigations, but to actually perform one part of them. This means that students will also need to learn scientific processes, including ways to think like a scientist. This is evident in that students are asked to "identify questions." Here, we understand that students will need to formulate their own scientific questions (problem-finding) and write testable hypotheses for these questions. These activities move beyond the *Knowledge* level of Bloom's and into the *Application* and *Synthesis* areas. Here, students must gather what information they have been taught about scientific investigations and then use this knowledge to construct high-quality hypotheses.

Based upon our unpacking activity so far, what would we expect to see in the science classroom? Certainly, we would expect some introduction of vocabulary terms with accompanying exercises and perhaps some textbook readings on vocabulary terms to show how these terms fit into a scientific investigation. The teacher might also introduce some outside reading, describing a scientific investigation that led to the invention of a useful everyday object (Velcro, Post-It Notes, glue) or the story of a historic scientific investigation (Walter Reed and the cure for yellow fever; Fleming and the discovery of penicillin).

The teacher could highlight common investigational steps used in the discoveries or inventions and help students apply the steps identified to these investigations. A video clip from the movie T*he Medicine Man* might stimulate discussion of the importance of controlling variables. (The doctor in the movie seems to have found a cure for cancer using a tropical plant. He cannot reproduce the results, primarily because his original batch also contained ground-up insects that infested the tropical flower. The insects were the active ingredient in the cure, not the flower.) The teacher might stimulate student thinking by having them work in small groups to propose problems needing scientific investigations. From the class set of problems, the students could then work in their small groups to write testable hypotheses for these problems.

RESOURCES

De Kruif, P. (1926). *Microbe hunters*. Orlando, FL. Harcourt, Inc.

Pierce, J.R. & Writer, J.V. (2005). *Yellow jack: How yellow fever ravaged America and Walter Reed discovered its deadly secrets*. San Francisco, CA: Wiley.

Lee, R. (2002). *The eureka moment: 100 key scientific discoveries of the 20th century*. New York: Routledge.

Gratzer, W. (2002). *Eurekas and euphorias: The Oxford book of scientific anecdotes*. Oxford, England: Oxford University Press.

Schulman, T. (Writer), & McTiernan, J. (Director). (1992). *The medicine man*. {Motion picture}. United States: Hollywood Pictures.

In order to check the proficiency of students relative to this learning target, teachers might implement vocabulary quizzes, comprehension questions on readings, a short essay describing common steps found in several science investigations from the outside reading, and a rubric describing the qualities of a testable hypothesis that students could then use to assess their own hypotheses and those of peers.

2. Design and conduct scientific investigations

In this second benchmark, we find two new verbs: *design* and *conduct*. What levels of Bloom's Taxonomy do these verbs imply? In order to design and conduct an investigation, students will definitely have to *Apply* the *Knowledge* they

have learned about scientific investigations. In the design process, students must *Analyze* the problem in order to identify needed components and/or equipment, *Synthesize* information from multiple sources to help them choose the procedure, and *Evaluate* alternatives to choose the best method of investigation. In this benchmark, then, students will work at all levels of Bloom's Taxonomy. Here, again, the verbs at the beginning of the benchmark help signal the level of thinking required from the students. **FIGURE 2.3** may be useful in unpacking standards and benchmarks as it contains samples of these "signaling" verbs.

FIGURE 2.3
Verbs Signaling Cognitive Levels

COGNITIVE LEVEL IN BLOOM'S TAXONOMY	SIGNALING VERBS
Knowledge	identify, define, list, tell, label, match, select, choose, name, spell
Application	identify, make use of, plan, organize, develop, utilize, apply, try
Analysis	compare, dissect, inspect, categorize, contrast, simplify, distinguish, classify, examine, conclude
Synthesis	build, compile, invent, formulate, compose, construct, originate, change, adapt, solve, predict, make up, improve
Evaluation	criticize, judge, recommend, support, argue, justify, dispute, appraise, prioritize, assess, value, defend

What might this second benchmark look like in the science classroom? Previously, students may have identified problems and written hypotheses. From the class presentations of these problems and hypotheses, groups of students may choose one such problem, with its accompanying hypothesis and develop an investigation to prove or disprove this hypothesis. Alternatively, the teacher may propose a problem to the class and ask students to design an experiment to answer the question. A sample question might be: How can we determine the background level of radiation present in this classroom? The teacher could provide a graphic organizer that would require certain information (safety precautions, independent and dependent variables, equipment list, procedural steps, etc.).

In this manner, several groups may write proposed investigational procedures for the same problem. Each group could then present its proposal to the class. The class would *Analyze* all the proposals and decide which one would best answer the proposed scientific question. They would need to be prepared to justify their argument for one proposal over another (*Evaluation*). Finally, once a particular procedure was chosen, each group could actually conduct the experiment and report its results.

To check student proficiency relative to this benchmark, the teacher might use a "Scientific Investigation" rubric, an "Oral Presentation" rubric, and a short essay requiring students to justify their choice of the best procedure.

In the discussion of the two benchmarks, we have seen that Content Standard A encourages thinking at all levels of Bloom's Taxonomy. It is important that teachers take the time to dissect or unpack standards (whether national or state standards) to ascertain the levels of thinking required in each. Many times, a standard at first glance appears to be a Knowledge level standard but can have higher-order thinking skills embedded within it. It is also important for teachers to ask, "How will this look in my classroom?" as they read through standards. Such visualizations can aid teachers in planning high-quality lessons that will actually help students meet the standards.

APPLICATION

Take one of the remaining benchmarks listed below and unpack it for levels of thinking using Bloom's Taxonomy. Then, write a brief description of how this benchmark might be addressed in your classroom.

- Use technology and mathematics to improve investigations and communications.
- Formulate and revise scientific explanations and models using logic and evidence.
- Recognize and analyze alternative explanations and models.
- Communicate and defend a scientific argument.

Another Source for Student Learning Goals in Science

So far in this chapter, we have only examined one student goal of learning—that of scientific literacy promoted by the National Research Council. We have seen that this Council created standards to promote scientific literacy within eight different categories. Before we leave this discussion of "What Do We Want Student To Be Able To Do?" we should first examine some goals from another source. This new source is the National Assessment of Educational Progress (NAEP).

In the Framework for the 2005 NAEP Science Assessment, we find the matrix displayed in **FIGURE 2.4**.

FIGURE 2.4
NAEP Science Assessment Framework Matrix

KNOWING AND DOING	FIELD OF SCIENCE		
	EARTH	PHYSICAL	LIFE
Conceptual Understanding			
Scientific Investigation			
Practical Reasoning			

NATURE OF SCIENCE
Themes Systems, Models, Patterns of Change

Source: Science Framework for the 2005 National Assessment of Educational Progress. Retrieved 8.14.05 from **www.nagb.org/pubs/ s_framework_05/ch2.html***.*

Under the separate headings of these domains, the following student expectations are clustered:

Conceptual Understanding

1. Organize important science ideas and express them in their own words.

2. Demonstrate the acquisition of a meaningful knowledge base.

3. Successfully exchange ideas and information with other students.

4. Read, comprehend, discuss, and evaluate information in science articles.

5. Generate, research, and report on questions of interest.

Scientific Investigations

1. Demonstrate the use of science process skills (classifying, developing a research question, making predictions, collecting, analyzing, and interpreting data).

2. Demonstrate the use of laboratory skills.

3. Generate a hypothesis and design an experiment to test that hypothesis.

4. Determine if measurements are reliable and valid.

5. Make judgments about the adequacy of evidence supporting a hypothesis.

6. Develop alternative interpretations and look at data in more than one way.

Practical Reasoning

1. Work successfully through a complex problem with a group of other students.
2. Think abstractly and consider hypothetical experiences.
3. Consider several factors simultaneously.
4. Take a depersonalized view.

Nature of Science and Technology

1. Identify and summarize examples of how explanations of scientific phenomena have changed over time as new evidence emerged.
2. Demonstrate an understanding of the difference between correlation and causality.
3. Discuss the interaction of scientific knowledge and values as they relate to problems we face.
4. Summarize the review role of scientific organizations in avoiding bias and maintaining quality in published research.
5. Understand that scientific conclusions are based on logic and evidence, but no fixed set of steps makes up a scientific method.
6. Explore the advantages and disadvantages involved in the design and development of technologies.
7. Summarize examples of how scientific knowledge has been applied to the design of technologies.
8. Understand that models of objects and events in nature can be used to understand complex or abstract phenomena.
9. Understand that systems are often artificial constructs used by people to gain a better understanding of a complex idea and that a system construct entails identifying and defining its boundaries, identifying its component parts and the interrelations and interconnections among those parts, and identifying the inputs and outputs of the system.
10. Recognize patterns of similarity and difference, to perceive how these patterns change over time, to remember common types of patterns, and to transfer their understanding of a familiar pattern of change to a new and unfamiliar situation.

In the NAEP student expectations, we see the same trend (going from general statements to more specific ones) that was evident in the National Science Education Standards (NSES). We also find that several of the categories seem to overlap, as shown in **FIGURE 2.5**. Both sources emphasize learning science processes as well as science concepts. Thus, the two different sources appear to have similar ideas about what science students should know and be able to do as a result of activities within science classes.

How do your own student expectations compare to those of NSES and NAEP? At this point, please take a few minutes to reflect on what you feel are important student expectations for science instruction. You may wish to discuss your responses with other science teachers. The following questions may aid in this discussion:

- How would you rank order the NSES student expectations? The NAEP student expectations? Why would you choose this order?

- Do you feel that the NSES and NAEP expectations cover all the essential science content for your course? If not, what expectations would you add to the list? Are there any you would delete or revise?

In the space below, describe for your students and their parents the top 4 course outcomes you expect for the year:

1.

2.

3.

4.

FIGURE 2.5

Comparing Student Expectation Categories from National Science Education Standards and NAEP

NATIONAL SCIENCE EDUCATION STANDARDS	NAEP			
	CONCEPTUAL UNDERSTANDING	SCIENTIFIC INVESTIGATIONS	PRACTICAL REASONING	NATURE OF SCIENCE & TECHNOLOGY
Unifying Concepts and Processes in Science	X			X
Science as Inquiry		X	X	
Science and Technology				X
Science in Personal and Social Perspectives	X	X	X	X
History and Nature of Science				X

One Final Source of Student Learning Goals in Science

The most important source of student learning outcomes or goals that teachers should access is the state curriculum. Such state curricula are usually organized by grade levels in grades K–8 or by science disciplines (Biology, Chemistry, Earth Science, etc.) in grades 9–12. Like the National Science Education Standards, your own state curriculum may begin with broad goals that are then dissected into smaller and smaller component parts. For example, in Florida, the state science curriculum begins with the subject area (science), then breaks this into strands (The Nature of Matter, Energy, Force and Motion, Processes That Shape the Earth, Earth and Space, Processes of Life, and How Living Things Interact With Their Environment). Under each of these strands, a number of standards further explain what students should know and be able to do. The standards are then supported by benchmarks, which occur as the most specific level of the curricular hierarchy.

Such specificity can be very helpful to teachers who are trying to unpack written statements about what students should know and be able to do upon completion of their courses. To ensure that all essential content (science concepts, processes, and skills) encompassed in state curriculum is actually taught, teachers may find that designing a planning matrix similar to the one in **FIGURE 2.6** may be helpful. Here, the names of the major units constructed by the teacher are written across the top and standards from the state curriculum are written vertically on the left. The teacher can place checkmarks within the units where particular standards will be addressed. In this manner, the teacher can clearly map out a course of study that encompasses all student learning targets. Standards related to thinking skills or to understanding scientific processes (as "Use scientific processes to solve problems" and "Science, technology, and society are interdependent") can be explored in all units, whereas teaching of basic science concepts can be focused within particular units. Of course, the next steps for the teacher (as explained previously) are then to:

1. Visualize the actual activities that must occur in the classroom for students to achieve these targets.

2. Plan student assessments that will measure achievement of these targets.

Planning appropriate assessments for learning targets is the subject of our next chapter.

FIGURE 2.6
Planning Matrix for Learning Targets

STANDARDS	MAJOR UNITS					
	ENERGY SOURCES	EARTHQUAKES & VOLCANOES	SPACE/SOLAR SYSTEM	ENVIRONMENTAL ISSUES	WEATHER	LIVING THINGS
Use scientific processes to solve problems	X	X	X	X	X	X
All matter has observable, measurable properties			X			X
Basic principles of atomic theory			X			X
Energy may be changed in form	X					
Interactions of matter & energy	X					
Types of motion may be described, predicted, measured		X			X	
Types of forces and their effects on an object		X	X		X	
Processes in the lithosphere, hydrosphere, and atmosphere interact to shape Earth				X	X	
Need for protection of natural Earth systems				X		
Interaction and organization of Solar System with Earth			X			
Vastness of universe			X			

	MAJOR UNITS					
STANDARDS	**ENERGY SOURCES**	**EARTHQUAKES & VOLCANOES**	**SPACE/SOLAR SYSTEM**	**ENVIRONMENTAL ISSUES**	**WEATHER**	**LIVING THINGS**
Structure and function of living things						X
Process and importance of genetic diversity						X
Interdependent nature of living things				X		X
Consequences of using limited natural resources				X		
Natural events occur in patterns		X	X		X	
Science, technology, and society are interdependent	X	X	X	X	X	X

Source of standards: Florida Department of Education, Sunshine State Standards, Grades 6–8. Retrieved 8.15.05 from **www.firn.edu/doe/curric/prek12/pdf/science6.pdf**.

APPLICATION

Construct a matrix similar to the one found in **FIGURE 2.6** using standards from your state curriculum. Answer the following questions:

1. Are there standards that occur in EVERY unit you teach? Why or why not?

2. Must you teach a separate "nature of science" unit, or can standards related to this be incorporated into existing units?

3. How does constructing the matrix either increase or decrease your ability to see connections among science concepts? How might this affect your teaching? Your students' learning?

4. How comparable is your state curriculum to the NSES? To NAEP student expectations?

5. Do you feel that your state curriculum addresses all the essential scientific knowledge for your class/grade level? If not, what should you do?

Performance-Based Assessment: Observing Students

Assessment is an integral part of the teaching and learning process since the main goal of education is to produce or facilitate change in learners. How do we know if such a change is occurring? How do we know if students are becoming competent and knowledgeable? To obtain this information, we must select assessments appropriate to the desired student outcomes.

In **CHAPTER 1**, we listed several types of assessment methods available to teachers (see **FIGURE 1.2**), and in **CHAPTER 2**, we viewed classroom snapshots of how particular methods (essays, quizzes, oral presentations, demonstrations) might be used to judge student performance relative to particular benchmarks. In this chapter, we will elaborate on these methods that teachers can utilize to determine what students know or are able to do and particularly emphasize those that go beyond multiple-choice testing. In this chapter, we define *performance-based assessment*.

Performance-Based Assessment

If a teacher is interested in what students understand about types of rocks, she may choose to create multiple-choice questions to obtain this knowledge. On a multiple-choice test following the rock unit, these questions might appear:

1. The three classifications of rock are:
    a) conglomerate, metamorphic, and obsidian
    b) igneous, metamorphic, and sedimentary
    c) quartzite, igneous, and conglomerate
    d) gneiss, schist, and sandstone

2. On the Mohs scale, apatite is ranked as a 5. Which of the following rocks would scratch apatite, but not be scratched by it?
    a) talc with a ranking of 1 and calcite with a ranking of 3
    b) quartz with a ranking of 7
    c) diamond
    d) a and b
    e) a and c
    f) b and c

These are legitimate *Knowledge*-based questions that assess if students know that igneous, metamorphic, and sedimentary are the three types of rocks and that rocks with high numbers on the Mohs scale will scratch (but not be scratched by) rocks with low numbers. Therefore, if the learning goal is a *Knowledge*-based one, then these questions will certainly reveal whether students possess this knowledge.

Let us suppose that our state curriculum mirrors the National Science Education Standards, Content Standard D in Earth and Space Science for Grades 9–12. This standard states:

> As a result of their activities in grades 9–12, all students should develop an understanding of
> - Energy in the earth system
> - Geochemical cycles
> - Origin and evolution of the earth system
> - Origin and evolution of the universe

The key term in this standard is "understanding." What does it mean for students to *understand* the "origin and evolution of the earth system"? As we did in **CHAPTER 2**, we must first *unpack* this standard to determine the essential science information students will need to know. For example, this standard implies that Earth has changed since its origin—it is not the same today as it was originally. What then, have been the causal agents of these changes that students may need to know? The following concepts may occur to the experienced Earth Science teacher:

Uniformitarianism	Unconformity
Principle of Superposition	Earthquakes
Principle of Original Horizontality	Erosion
Folding and Faulted Layers	Volcanoes
Principle of Cross-Cutting Relationships	Deposition

We can define each of these terms for students, provide physical and virtual demonstrations (as using soft clay layers to model folding, mountain building and faulting, OR provide students with a link to a website that uses a multimedia presentation to demonstrate these), ask students to create 2-D and 3-D illustrations/models of the principles, provide textbook readings as well as readings from other source materials, and show students photographs of geologic sites while explaining which geologic forces/principles caused unique formations. How will we know, however, that students truly understand how the Earth changes? Are multiple-choice questions enough?

To demonstrate *understanding*, we want students to do more than just recognize or recall the right answer to a question. We want to stimulate their higher-order thinking skills, as application, analysis, synthesis, or evaluation. One way to do this is to use a performance-based assessment. Some examples of performance-based assessments related to the higher-order thinking skills for this standard might include:

> *Application* and *Analysis*: The teacher has already shown the students photographs of geologic sites and explained how these might have formed. Provide students with a new photograph (one the teacher did NOT explain) and ask them to write a plausible explanation of how this site formed, using at least three of the vocabulary terms.

> *Synthesis*: Ask students to create a five-step sequence of geologic events, using at least three of the vocabulary terms. Then, have them write a brief description explaining each step and provide an illustration of the landform at each step.

Evaluation: Ask students to support or refute this statement: *The Mississippi River Delta and Delicate Arch in Utah were formed by similar processes.*

All three of the above performance-based assessment examples provide students with an opportunity to demonstrate what they know, rather than just regurgitating a definition or recalling isolated bits of information. In **FIGURE 1.2** of **CHAPTER 1**, we listed three main categories of performance-based assessment methods:

1. Observing students using informal observations and structured observations

2. Soliciting information from students via interviews or self-assessment questionnaires

3. Evaluating student work using open-ended questions, performance tasks, journals, exhibitions, and portfolios

The three geologic change examples would all fall in the "evaluating student work" category, as all are open-ended questions (questions that require students to construct a response, rather than choose a response from a list of possible answers). In the next few chapters, we will examine each category of performance-based assessment in more detail and provide suggestions for implementing them in the science classroom. In this chapter, we focus on the first category —observing students.

Observing Students

Teachers constantly observe students. They see that Juan is arguing with his group, Serena looks confused about a new concept, Nikita is daydreaming, Chavez is working hard, etc. Such observations are *informal* in nature and may serve both assessment and classroom management functions. Teachers also make more formal observations of students, as when they use an observation instrument to collect data about student performance. These formal observations are classified as *structured* observations. Some goals or objectives can only be assessed by such structured observations. For example, it is difficult to imagine how a teacher would assess students' team problem-solving skills or success at independent lab work without observing them.

Informal Observations

With informal observations, teachers are actively observing the students, but no particular group or individual is the target of the observation. Similarly, this type of observation generally occurs spontaneously and may not have a predetermined focus. However, informal observations can be very useful assessment tools. Through such observations, teachers may, for example, become aware of students in their classes who are able to work independently and teachers can also identify those who require a great deal of assistance. More formal observations can then be planned to capture detailed information on these struggling students. Information from informal observations can greatly impact the classroom instruction, as the teacher uses such information to plan differentiated instruction for his/her diverse students. Such observational information can also provide the basis for reports to parents via phone calls or conferences.

Structured Observations

Structured observations, unlike informal ones, usually have a speci-fied focus and a specific target group (or individual). In order to col-lect information relevant to the focus of the observation, a teacher may use an observation instrument. Such an instrument is often in *table* or *matrix* form with students' names listed down one side and particular behaviors listed across the top. For example, suppose an elementary science teacher has recently set up five science activity centers where his students can individually engage in hands-on sci-ence. This teacher may wish to evaluate students' progress by seeing if the students stay on task and if they are able to work independently with the center materials. He develops a form similar to the one depicted in **FIGURE 3.1** to use in collecting this observational informa-tion about students. He lists the students being observed in the space provided. Then, he observes these students during a 10–15 minute individual hands-on science activity occurring at the five centers. For each student, he records information about their on-task behavior (a check in the box denotes on-task work while a blank box means the student was off-task) and notes if assistance is needed or solicited. If he is able to observe all five students within the 10–15 minutes and there is still time left over, he may perform another round (or several more rounds) of observation on these students.

FIGURE 3.1
Hands-On Science Activity Observation Form

Date of Observation

_____

 Time Observation Began _____ Time Observation Ended _____

Hands-on Science Activity Description:

STUDENT NAMES	OBSERVATION ROUND 1		OBSERVATION ROUND 2		OBSERVATION ROUND 3		OBSERVATION ROUND 4	
	ON TASK	ASSISTANCE NEEDED	ON TASK	ASSISTANCE NEEDED	ON TASK	ASSISTANCE NEEDED	ON TASK	ASSISTANCE NEEDED

Legend

✔ in On Task box means that student was working to complete the task when observed.

Code for Assistance Needed: **N** = None, **S**=Some, **M**= Much

The information collected on such observation forms could be used in a variety of ways. Structured observation data often allows teachers to profit from new information that may challenge some of the inferences they have made about students. For example, before collecting the Hands-On Science Activity data, the teacher might have assumed that Mai would have difficulty staying on task. However, after the first observation in September, he realized that Mai functioned very well when working independently. The teacher found, however, that Alice, Juanita, and George needed assistance to use the materials appropriately.

Observational data collected over time can be useful for showing changes in student performance. **FIGURE 3.2** displays the data collected about students' on-task behavior over three different observation periods. These data show a general pattern of improvement over time on independent lab work and also reveal which (and how many) students need improvement in this area. The teacher can share the data with students when discussing their performance in science class with them and setting goals to improve this performance.

FIGURE 3.2
Hands-On Science Activity
Observational Data Summary

Observation Dates: September 12, January 23, and May 5

STUDENT NAMES	NUMBER OF TIMES OBSERVED WORKING ON TASK		
	SEPTEMBER	JANUARY	MAY
Alice	1	2	3
Mai	4	5	4
Juanita	2	3	4
Michael	4	4	5
George	2	4	5

From this data, it is easy to see that Alice, Juanita, and George have made gains in on-task behavior, while Mai and Michael have maintained their ability to work independently.

Besides tables and matrices, another useful format for recording observation data is the taking of *anecdotal notes*. Anecdotal notes are simply narratives that describe observed behaviors. Such narratives are particularly appropriate to use when observing complex behaviors, such as group interactions, that do not lend themselves easily to a checklist format. For example, a science teacher may observe and describe a cooperative team problem-solving activity. The purpose of the structured observation would be to determine how students on the team contributed to the completion of the activity. An example of anecdotal notes taken by the teacher during this activity is displayed in **FIGURE 3.3**.

FIGURE 3.3
Anecdotal Notes on Group Problem-Solving Activity

Observer: Mrs. Lee

Time: 1:20-1:30 PM

Date: Sept. 12

Group Observed: Crystal, Jack, Ramon, and Anita

Purpose of the Observation:
To be able to describe to students how their behaviors contributed to or detracted from the group's efforts to solve the problem. One of the goals for the year is the development of group problem-solving skills. This assessment approach documents student functioning relative to this goal.

Notes:
Crystal reminded the group that they needed to choose a recorder. Ramon volunteered to be the recorder and write things down if they told him what to write. Jack said, "What are we supposed to do?" Anita looked at the worksheet and began reading aloud the directions for the activity. Jack started blowing in the air and talking about wind. Crystal told Jack to stop playing. He looked at his sheet for a moment and then started blowing again.

The first section on the worksheet asked the students to identify the different properties of wind. Crystal told Ramon to write down: "the way it blows." Anita offered, "how fast it goes." The next section asked the students to find a way to measure one of the properties they had identified. Crystal said that they should build a weather vane to show the direction the wind blows; Ramon and Anita agreed. Jack didn't say anything. He was busy drawing a sailboat. Crystal sent Jack off to the side of the room to get materials to build the weather vane. Jack returned with the materials and immediately started to put them together. Crystal went to the side of the room to get the things Jack forgot. Each of the children began building their own weather vanes. Jack wanted everyone in the group to see his when he blew on it. The other children began blowing on theirs. After a few minutes, Crystal decided that Jack's weather vane was the best.

These anecdotal notes constitute a written record of how the students worked together to solve the problem. The notes provide evidence about how each of the students contributed to the problem-solving activity and can help the teacher discern patterns of student behavior. From this brief narrative, it appears that Crystal is very task-oriented and is working as the group leader. Mrs. Lee (the teacher) can use this information to set goals for individual students or to structure future cooperative groups. Mrs. Lee will probably target groups experiencing difficulties in cooperative behavior for further observations, and the anecdotal notes from these can provide a basis for teacher comments and recommendations to the groups. Over time, a series of anecdotal notes may also help document how students changed the way they worked in teams.

APPLICATION

Reflect upon the student outcomes you wrote in **CHAPTER 2**, and then examine your state standards.

Identify those outcomes/standards that could be best assessed by a structured teacher observation. Examples may include outcomes/standards that would require students to act or behave in certain ways, as work cooperatively, organize materials, persist in work even after encountering obstacles or problems, etc.

Observations may also be useful when you suspect a student has a specific learning disability or when you are trying to identify the auditory, visual, or kinesthetic learners in your classroom.

1. Create an observation instrument for this observation.

2. Perform the observation.

3. Analyze your results.

4. Answer these questions:

   a) Did your results confirm previous inferences you held?

   b) Did the observation reveal any surprising results?

   c) What will you change about your classroom, based upon the results of this observation?

CHAPTER FOUR

PERFORMANCE-BASED ASSESSMENT: SOLICITING INFORMATION FROM STUDENTS

In the last chapter on observing students, several examples were provided of how teachers collect information about student performance. A second method of collecting information about students involves the analysis of replies that students give in *interviews* and on self-reporting *questionnaires*. Again, the types of assessment methods described here fall under the heading of performance-based assessment. In such assessment, students construct responses, rather than choosing responses from a given list. Performance assessments emphasize higher-order thinking skills (*Application*, *Analysis*, *Synthesis*, and *Evaluation*) that go beyond the *Knowledge* Level.

As mentioned above, when we solicit information from students to aid us in assessing their performances, we can use interviews or questionnaires. Interviews involve face-to-face verbal exchanges between the teacher and the student. In self-reporting questionnaires, students respond to written questions and statements. The focus of the interviews or questionnaires may be on a cognitive event (e.g., what students understand about a particular topic), how they feel (e.g., what they like or dislike about working in groups), or on personal behaviors (e.g., if they talk about science topics at home or read science books in their leisure time).

Interviews

Although individual interviews with students are time-consuming and difficult to manage in a classroom setting, there are several reasons why they are worth the effort:

1. *For those students who seem to have trouble with a particular concept or skill (as demonstrated on assessments), interviews may be a way of further assessing their functioning relative to instructional objectives. A series of probing questions can be developed that would be useful in deciding how to help students improve their performance.*

   Possible Use: Mrs. Juarez notices that Trung is an enthusiastic science student who frequently asks probing questions and volunteers correct answers to her in-class questions. However, Trung is doing poorly on written work—homework and tests. Mrs. Juarez is puzzled by this disconnect between Trung's verbal and written work. She schedules an interview with Trung.

2. *If a new unit is being developed, interviewing a sample of students of different abilities about their prior knowledge on the topic should allow the teacher to assess students' readiness to learn the new topic. Instruction could then be designed to target their entry level of knowledge.*

A Time This Would Have Helped: In Chemistry II, Mrs. Butler alludes to several concepts from Chemistry I class. Not everyone in Chemistry II had Mrs. Butler for Chemistry I last year. Finally, one brave student raises her hand and confesses, "I don't know what you mean. We didn't cover this last year."

3. *Interviews can send a message to students that a teacher cares about what they think, what they are interested in, and what they understand. Rapport is encouraged and student motivation may be increased.*

   A Time This Would Have Helped: A team of teachers in a middle school plans an interdisciplinary unit on Radiation. The science teacher decides to focus on scenarios from the Cold War emphasizing the dangers of radiation from atomic bombs. This teacher grew up in Florida during the Cuban Missile Crisis, so this topic is highly relevant to her. When she teaches the unit, however, she notices that the students are simply "going through the motions." They are not interested in the subject at all. What was highly interesting and motivational for her does not have the same effect on her students. Interviewing the students to find out what would interest them in this area could have increased student motivation to learn.

4. *Interviews allow students who have difficulty with written tests to express what they understand in a context that may be less threatening and anxiety producing. On the flip side, students who do well on written tests may have difficulty communicating their responses to verbal questions and may need practice in doing so.*

   Possible Use: Marlee flunks every written test in science, yet seems to know the science material. Mr. Chapman schedules an interview with Marlee before the next test in order to probe her knowledge of test items and to have her read and explain a sample passage from the text in order to explore any reading difficulties she is experiencing.

5. *Interviews provide teachers the opportunity to probe and ask follow-up questions in ways that challenge students to think beyond their current level of understanding and to organize their knowledge in more systematic ways. Thus, follow-up questions can be individualized such that students are pushed as far as their level of understanding permits.*

   Possible Use: Maria is the Jeopardy "queen" in Mrs. Sicco's science class. Every time this middle school class plays the review Jeopardy game, Maria wins. She knows ALL the facts. However, Mrs. Sicco notes that Maria does poorly on questions that ask her to synthesize knowledge and that Maria consistently receives poor marks on student-constructed concept maps. Mrs. Sicco schedules an interview with Maria to help her formulate connections among facts by helping her organize and categorize her knowledge on a topic.

6. *One common student outcome in science courses is that students will learn to communicate effectively. If science teachers promote this goal, interviews are clearly an*

assessment method of choice. That is, students should not only be assessed with written tests, but also should be asked to express what they know verbally.

Possible Use: David is one of Mr. Chang's students in 6th-grade science class. He maintains average to high scores on all written assignments, yet he rarely speaks in class. In fact, he is so quiet, he could easily become one of those students who "slip through the cracks." His scores and his behavior in class are not low enough or negative enough to warrant intervention or attention. He could pass through Mr. Chang's class without ever making a connection with Mr. Chang. An interview might reveal ways to get David more active verbally and reveal ways he could improve his performance in science. In addition, an interview could reinforce for David that practicing his verbal explanations is as important as practicing written ones.

Interviews, as the ones described above, can vary in their degree of structure. In **unstructured** interviews, the content and order of the questions vary with the student and are responsive to each student's answers. The example of Mrs. Juarez and Trung would exemplify such an unstructured interview. Mrs. Juarez is truly puzzled by the differences in Trung's written and class behaviors. She may simply begin the interview by pointing out the discrepancies and then key any further remarks on Trung's responses. Such unstructured interviews also occur any time the teacher and a student share personal dialogue, as when the teacher stops by a student desk when circulating. Such mini-interviews occur spontaneously, on a daily basis, and are used by teachers to assess students' competence relative to instructional examples.

In **semi-structured** interviews, there may be some themes identified to structure the interviews, but questions within those themes may be phrased differently for different students. For example, in the David/Mr. Chang scenario above, Mr. Chang may implement a semi-structured interview. The "theme" for the interview will be David's verbal behavior. In this interview, Mr. Chang will look for ways to increase this behavior; in interviews with other students, Mr. Chang may be looking to curtail such behavior!

In **structured** interviews, teachers ask students to respond to the same set of questions. The Radiation Unit example described above would have benefited from structured interviews of students. The students could have responded to the same set of questions about their interest in radiation. The middle school science teacher could then have planned a more motivating unit on radiation for these students.

An interview may, at times, substitute for a written test when the teacher wishes to determine what students know. If a teacher wants to give students an opportunity to be interviewed on their understanding of a topic rather than taking a quiz on this information, the set of questions should be similar for all students choosing this option. Therefore, this is another use of structured interviews.

Collecting Usable Data from Interviews

An interview, like an observation, is experience-based. It occurs quickly, and then it is simply over. In order to effectively use interviews, some means of capturing data is necessary. Trying to take anecdotal notes during an interview is difficult and may actually interfere with the purpose of the interview. If the teacher is using the interview to demonstrate caring, attention given to note-taking instead of attention given to the student will send the wrong message. Alternatively, interviews can be audiotaped. The teacher can then listen to the audiotape later to form inferences or conclusions or to write helpful anecdotal notes.

An interview instrument (very similar to an observation instrument) may also be developed prior to the interview. All structured interviews in which all students respond to the same set of questions will involve the use of such a data-capture instrument. Similarly, in the case of Marlee and Mr. Chapman above, Mr. Chapman could bring a list of science concepts to the interview to discuss with Marlee. Based upon Marlee's responses, Mr. Chapman could rate her knowledge of the concepts from poor to excellent. He could then help Marlee devise strategies to improve her performance on the "poorly" rated concepts.

In effect, Mr. Chapman is giving Marlee an oral exam, rather than a written one. Such oral tests may give those students who have poor literacy skills a chance to succeed. In addition, this assessment method provides the teacher with assurance that students understand the test question. Conversely, written exams make the assumption that students understand the questions asked. Another advantage of oral exams is also reported by some teachers. They report that students take oral tests more seriously because they feel such tests are more personal expressions of competence than a written test would be. Students may prepare more carefully if they know they must stand before a teacher and answer questions individually.

FIGURE 4.1 provides an example of an oral exam on the three phases of water. This exam could be considered a structured interview, in that the same set of questions is used with all students. A rating scale for answers is provided for each question, ensuring effective data-capture. Students respond to the question, the teacher records the students' answers, and then rates these answers by assigning point values.

FIGURE 4.1
Oral Exam on the Three Phases of Water

Student's Name: _____

Date: _____

SCORING KEY	POINTS AWARDED	QUESTIONS
1 point for each phase identified correctly (ice, water, steam)		1. What are the three phases of water?

0 = Incorrect
1 = Partially correct
2 = Satisfactory
		2. Describe each of the three phases:
		   a) Ice
		   b) Liquid
		   c) Steam

0 = Incorrect
1 = Partially correct
2 = Satisfactory
		3. What happens when water goes from one phase to the other:
		   a) Ice to liquid?
		   b) Liquid to Ice?
		   c) Liquid to Steam?
		   d) Steam to Liquid?

No rating		4. Is there anything you do not understand about water phases?

Total Points Awarded = _____ (Maximum is 17)

The results of the oral exam displayed in **FIGURE 4.1** could be used in a number of ways. Students who had less than 17 points could be assigned a peer coach who scored all 17 points on the exam. This peer coach could work on the questions with the students until he or she was ready to retake the exam. The second administration could result in a score entered into the grade book.

No matter what type of interview protocol is chosen (unstructured, semi-structured, or structured), it is important to obtain usable and useful data from the interview. This is particularly important in that interviews are very time-consuming. The teacher must ensure that the interview is actually worthy of the time commitment. The suggestions for interviews found in **FIGURE 4.2** may be helpful in this area.

FIGURE 4.2
Suggestions for Interviews

1. Use sampling techniques to choose participants for your interviews. In other words, if a small sample of your students can provide the information you require, don't try to interview all the students.

2. In one school year, try to ensure that every student in your class participates in at least one interview with you.

3. Keep the tone of the interviews positive and constructive. Try not to give verbal or facial expression cues that can be interpreted as meaning that an answer is silly or that the student has made an error.

4. Let students respond without interruptions, and give them time to think before they respond. (Remember Wait Time One and Wait Time Two. Wait Time One means waiting at least 5 seconds after you pose a question before calling on a responder. Wait Time Two reminds teachers to wait at least five seconds after the student has responded before speaking.)

5. Try to keep interviews short and focused on truly important questions.

APPLICATION

1. Review the six examples of interviews described earlier in this chapter. Choose one interview type that you would like to implement in your class (structured, semi-structured, or unstructured). Create a data-capture instrument for this interview and implement.

2. Reflect on the students in your class. Would an interview with any of your students be useful to you in assessing their performances? Make a list of students who might benefit from an interview.

3. Strategize ways to schedule student interviews (what will other students do while some are interviewed?).

Self-Assessment Questionnaires

Every assessment tool has advantages and disadvantages; some serve a particular purpose better than others. Student self-assessment questionnaires may be helpful in determining how students perceive their own knowledge, skills, or the quality of their work. Such questionnaires may also reveal concerns students have about their

academic progress, their prior level of experience with a topic or skill, their feelings about the class, and their interest in science as a career. Questionnaires are also excellent methods to use if teachers want to see the classroom through the eyes of a student. It is often useful to compare student perceptions to teacher perceptions in order to understand how classroom instruction and assessment procedures can be improved.

When used appropriately, self-assessment questionnaires actively involve students in reflecting on their own learning processes (promotion of meta-cognition) and emphasize the importance of students' awareness about what they know and what they need to know. Therefore, self-assessment questionnaires certainly fall within the parameters of performance-based assessment, in that they promote student cognition beyond the *Knowledge* level. These questionnaires are in a special category of performance-based assessment, however, in that they usually contain a mix of selected response and constructed response items. Most of the other performance-based assessment methods we have previously discussed in this manual relied on constructed response items only. Figure 4.3 displays one example of a self-assessment questionnaire that uses a mix of selected (see question 1) and constructed response items.

FIGURE 4.3
Science Skills Self-Assessment

Directions: *Read the questions and statements below and then answer each as best as you can. There are no right and wrong answers.*

1. How would you rate your interest in science right now?

    ❑ Very High   ❑ High   ❑ Medium   ❑ Low   ❑ Very Low

2. What did you like the most about science last year?

3. What did you like least?

4. Put a check by each instrument you have used. Beside each instrument, describe briefly what it does.

    ❑ Microscope _____

    ❑ Weight scale _____

    ❑ Thermometer _____

    ❑ Weather vane _____

    ❑ Ruler _____

    ❑ Barometer _____

    ❑ Compass _____

    ❑ Rain gauge _____

5. What do you like or dislike about working with a team of students?

    _____

    _____

The self-assessment questionnaire displayed in **FIGURE 4.3** is one that a teacher might give to students at the beginning of the year to better understand their science background and interests. In administering the questionnaire, the teacher might show the students each of the instruments listed in Question 4 so that students who knew how to use the instrument but had forgotten the name of the instrument could respond.

The teacher could use the assessment results in several ways. First, the teacher may want to summarize the frequency of responses to the interest question (**QUESTION 1**) as a baseline for comparison to responses on this same question at the end of the year. Summarizing the responses to the instrument question (**QUESTION 4**) in a frequency chart (instruments by number of students who had used each and could describe the function of each) could assist the teacher in judging how much remediation was needed. If cooperative learning skills were to be a focus for the year, the names of students who indicated dislikes about working in a team (**QUESTION 5**) could be listed, and anecdotal notes kept about any difficulties they had when teamwork was initiated.

Students can also be queried via self-assessment questionnaires on their understanding of science concepts. Yager and Kellerman (1992) note that a teacher might list the topics to be covered over a period of time (e.g., carbohydrates, concentration, starch, glucose, digestion). They suggest that students could be asked to rate each concept using the key found in **FIGURE 4.4**.

FIGURE 4.4
Key for Rating Understanding of Science Concepts

NUMBER	STATEMENT
1	I have never heard of it.
2	I have heard of it but do not understand it.
3	I think I understand it partially.
4	I know and understand it.
5	I can explain it to a friend.

Questionnaires on understanding of science topics do not necessarily have to be this formal or even written. When a new topic is presented in class, teachers can issue students three pieces of colored paper (green, yellow, and red). As an explanation of the new topic progresses, the teacher can stop periodically and ask students to hold up the appropriate piece of colored paper. Green means "Keep going. I'm with you." Yellow means "I'm a little confused. Please explain this again or in a different way." and Red means "Stop! You've lost me."

Such checks on student perceptions can inform the teacher of comprehension problems immediately—as they are happening. This prevents the teacher from simply assuming that all students understood the topic of the day. Self-assessment questionnaires of the above two types (written assessments in which students use a rating scale and in-class assessments using colored papers) are often perceived

by students to be less threatening than a pre-test or comprehension quiz. They can give students a sense of the different levels of knowing (meta-cognition) if used frequently in a class situation.

APPLICATION

Design one self-assessment questionnaire that would be useful to your class. This can be a questionnaire soliciting prior knowledge, present level of knowledge, or student feelings/beliefs. Complete the questionnaire yourself and then compare your answers to the summarized responses of your students. Answer the following questions:

- What surprised you about the student results?

- How can the information you garnered through this questionnaire help improve instruction or assessment procedures in your classroom?

- Did you discover any student misconceptions?

Performance-Based Assessment: Evaluating Student Work

In keeping with the title of this publication, we have focused primarily on assessment methods that go beyond the multiple-choice test. We have emphasized performance-based assessment in the past two chapters, and we continue this emphasis in **CHAPTER 5**. To this end, this chapter covers the following types of performance-based assessments: open-ended questions, performance tasks, logs, journals, portfolios, and exhibitions/projects. These assessments all involve the evaluation of student work. Such work is often tangible, as products are created (written answers to questions, log entries, journal entries, portfolio artifacts, science backboards, formal lab reports, etc.) Sometimes teachers must evaluate intangible student work, as oral presentations, student demonstrations, re-enactments, debates, etc. The purpose of this chapter is to offer suggestions for implementing both tangible and intangible assessments that involve teachers in evaluating student work.

Open-Ended Questions

Rather than having students select a response, open-ended questions ask students to produce a response. The length of the responses can vary considerably based upon the age of the student, the question asked, and the time provided to complete the question. Open-ended questions, like other performance-based assessment methods, require students to use higher-order thinking skills and therefore exercise more complex cognitive processes than simple multiple-choice questions. Gronlund and Linn (1990) found that open-ended questions particularly tapped into such high cognitive processes when students were asked to respond to the types of question starters found in **FIGURE 5.1**. A level of Bloom's Taxonomy is matched to each of these "starters" in **FIGURE 5.1**, showing how the question targets higher-order thinking skills and a sample science question is also displayed in this figure.

FIGURE 5.1
Open-Ended Questions Requiring High Cognitive Processes

STARTER FOR THE QUESTION	SAMPLE SCIENCE QUESTION	BLOOM'S LEVEL
EXPLAIN A CAUSE-EFFECT RELATIONSHIP	Why may too-frequent reliance on penicillin for the treatment of minor ailments eventually result in its diminished effectiveness against major invasion of body tissues by infectious bacteria?	Analysis
DESCRIBE AN APPLICATION OF A PRINCIPLE	Would you weigh more or less on the moon? On the planet Jupiter? Explain.	Application
FORMULATE A QUESTION, HYPOTHESIS, OR A CONCLUSION	What questions should a scientist ask in order to determine why more smokers than nonsmokers develop lung cancer?	Synthesis
DESCRIBE THE LIMITATIONS OF THE DATA	In this class, we conducted a survey concerning school uniforms. Are we ready to make a report to the school board on how our middle school feels about this issue? Why or why not?	Evaluation
EXPLAIN A METHOD OR PROCEDURE	One of the big ideas in physics is Newton's Third Law. State this law, explain its meaning, and give one real-life example (other than those used in the text or discussed in class) of this law in action.	Application
INTEGRATE LEARNING IN DIFFERENT AREAS	Using the human and wildlife population density maps below, make a recommendation about where to locate the new airport. Remember to preserve as many wildlife habitats as possible.	Synthesis
CREATE OR DESIGN SOMETHING (I.E., AN EXPERIMENT)	Devise an invention that makes an everyday task easier to accomplish. Use (and label) at least three simple machines on your design.	Synthesis
EVALUATE THE WORTH OF AN IDEA	The Florida Fish and Wildlife Commission is debating lifting the ban on alligator hunting in the state. Present the pros and cons of such a change in state policy.	Evaluation

As we have seen in **FIGURE 5.1**, open-ended questions can stimulate the use of higher-order thinking skills. Such complex open-ended questions can also help assess a variety of instructional goals, including conceptual understanding, application of knowledge, the use of science process skills, and divergent thinking skills. Examples of how open-ended questions can be used with each of these instructional goals are displayed in **FIGURE 5.2**.

FIGURE 5.2

INSTRUCTIONAL GOAL	QUESTION	EXPLANATION
CONCEPTUAL UNDERSTANDING	How would life and the conditions on earth be different if all bacteria and fungi became extinct? Explain the changes that might occur and give as much detail as possible (Grade 8). Source: Open response released item (1991–1992), Kentucky Instructional Results Information System. Kentucky Department of Education. Division of Curriculum, Assessment, and Accountability, Capital Plaza Tower, Frankfort, KY 40601.	The question asks students to access background knowledge, organize and express ideas in their own words, and generate a report on a question. All of these activities tap into the conceptual understanding the student has for the interdependence of life on Earth.
APPLICATION OF KNOWLEDGE	Using the weather map displayed on this page, make a forecast for the weather in North Carolina for the next day. Explain why you made the forecast (Grade 6). Source: Open response released item (1991–1992), Kentucky Instructional Results Information System. Kentucky Department of Education. Division of Curriculum, Assessment, and Accountability, Capital Plaza Tower, Frankfort, KY 40601.	Before answering this question, students have used weather maps and worked on weather predictions. They have learned the symbols associated with fronts, various types of precipitation, isobars, etc. and studied how each may affect a region's weather. They are now being asked to apply this knowledge to a new (previously unseen) weather map.
USE OF SCIENCE PROCESS SKILLS	You are a state scientist asked to develop an experiment to determine whether discharge from a factory is endangering Kentucky Lake (Grade 12). Identify several possible consequences of the discharge. Choose one of the consequences and design an experiment to investigate whether the consequence is actually occurring and if it is caused by the discharge. Describe how you would investigate, the kinds of data you would collect, and what you would do with your data. Source: Open response released item (1991–1992), Kentucky Instructional Results Information System. Kentucky Department of Education. Division of Curriculum, Assessment, and Accountability, Capital Plaza Tower, Frankfort, KY 40601.	Here, students will need to know the integral parts of a scientific investigation and how to apply this knowledge to an actual problem/question.
DIVERGENT THINKING SKILLS	Suppose there were no more disease in the world. List as many possibilities/consequences as you can for what might happen in the future as a result of this. Source: Assessment Ideas for Science in Six Domains (1992). Robert E. Yager and Lawrence R. Kellerman (Eds.). Science Education Center. Van Allen hall, University of Iowa, Iowa City, Iowa 52242.	Divergent thinking requires students to create multiple, original approaches to problems. Scientists use divergent thinking to generate research questions and hypotheses and to develop plans of action. Here, students are asked to consider the consequences of an action and use their background knowledge as well as divergent thinking to list the possible consequences.

1. Develop open-ended questions for the next unit of study you will teach. Develop one question for each of these levels of Bloom's Taxonomy: Application, Analysis, Synthesis, and Evaluation.

a. Consider the instruction that students will need in order to be ready to answer these questions. Incorporate this instruction into your unit plan.

b. Consider the practice in answering open-ended questions your students will need before such questions can become part of an assessment that counts as a grade. Incorporate this practice on answering open-ended questions into your unit plan.

2. Match the opened-ended questions for your next unit to particular learning goals (from national or state standards). Did you use some goals not listed in Figure 5.2? What were they? Share these goals with their accompanying open-ended questions with a colleague. Ask the colleague to comment on the match between the goal and the question. In other words, would this question actually help determine student progress toward the goal?

Student Responses to Open-Ended Questions

If open-ended questions are to be included on a test that will be graded, it is important for teachers to prepare students for this task. After many years of only encountering multiple-choice testing, some students may have difficulty with open-ended questions. Students will need in-class practice on writing answers to open-ended questions; they will need feedback on their practice performances; and they will need to understand the criteria that will be used to judge their responses.

At first, student responses to open-ended questions may be short, somewhat incoherent, and not well developed. It may be difficult to judge their understanding of the concept, simply because they do not possess sufficient communication skills to convey their thoughts. To aid students in developing such skills, the teacher can use several techniques. For example, the teacher can pick the best student responses to read aloud to the class and then ask the class to critique their own responses in terms of whether or not they met the standard exemplified in the example read aloud. She may ask the class to articulate the criteria they are using to judge their own work, based on what they heard in the example. A class compilation of criteria, along with explanations of each criterion would be helpful in defining the quality expected in the responses.

Students will need practice in incorporating these quality criteria into their own writing. Therefore, no grades should be taken until the second or even third administration of open-ended questions or until it is clear that students have had ample opportunities to understand the expectations.

Grading open-ended questions involves interpreting the quality of the response in terms of clearly articulated criteria. **FIGURE 5.3** displays several open-ended questions related to the apparent motion of the sun along with a rating scale to use in assessing the quality of student responses. Criteria for the responses include scientific accuracy of the explanation/description and the coherence of the response. Distinguishing between a score of 2 (accurate but not well written) and a 3 (accurate and well written) may help to impress upon students the importance of structuring their responses so that they are coherent to the reader.

FIGURE 5.3
Open-Ended Questions With a Rating Scale For Responses

Questions:

1. Why do we use the term "the sun's apparent motion"?

2. If we agree that the sun is not really moving across the sky, what is happening to make it look that way?

3. At 9:00 AM, a shadow is west of a tree; at 4:00 PM, the shadow is east of the tree. Explain why this happens.

4. Why do people in North Carolina see the sunrise before people in California?

Rating Scale:

0 – Incomprehensible/inaccurate explanation
1 – Provides partially accurate explanation
2 – Provides accurate explanation but not well written
3 – Provides very well-written and accurate explanation

Source: Rita Elliot, A.G. Cox Middle School, Pitt County Schools, Winterville, NC.

Before implementing open-ended questions in the classroom, the teacher would be advised to utilize the following suggestions:

- Be clear about the purpose of such questions. What instructional goals will they help you assess? For example, in **FIGURE 5.3**, the instructional goal targeted by the question may be "students will be able to explain phenomena relevant to the Earth/sun system."

- Answer the questions yourself before administering them to students. This will help you clarify your own expectations regarding an ideal student response.

- Develop a rating scale or point system to use with the questions. Share this rating scale with students before they begin to work. (More information about developing such grading schema is included in **CHAPTER 6**.)

- Read over a sampling of answers before you grade them. This will help you get an idea of the range of responses present for each question. It may be helpful to sort the responses

into piles based on the rating scale being used (all the ones together, all the twos together, etc.) before assigning a final scale value to the response.

The rating scale used in **FIGURE 5.3** used scientific accuracy and coherence as criteria for judging student responses. These criteria define what the teacher is expecting—what is being assessed. However, other assessments are possible. For example, student responses to open-ended questions can be analyzed to identify misconceptions or problems in understanding a concept. Rather than grading such questions, the teacher can choose to group the responses into categories of similar answers so that remedial instruction can respond to the kinds of errors being made.

Performance Tasks

Although many achievement objectives can be assessed with paper-and-pencil tests, there are other objectives that require students to actually demonstrate their competence. In some situations, given the purpose of the assessment (e.g. licensing people to drive cars), a performance test is necessary. It would be unthinkable (and dangerous!) to license people to drive on the strength of a written test on driving rules. Likewise in science instruction, there may be some skills (science investigation skills, skills in using science equipment) that are most appropriately assessed by having students perform tasks rather than take pencil-and-paper tests.

Such performance tasks can be used to assess a variety of instructional goals. Consider the Electric Circuits Performance Task described in **FIGURE 5.4**. This task has the ability to assess:

1.) Students' abilities to manipulate science equipment (in order to create working electrical circuits)

2.) Students' conceptual understandings of the two types of circuits (series and parallel)

3.) Students' abilities to self-assess and correct errors (checking to see if circuits work)

4.) Students' abilities to organize thoughts and express ideas coherently (writing answers to questions)

5.) Students' abilities to classify and analyze (relating how one circuit is different from another)

6.) Students' abilities to evaluate or choose alternatives (explaining why one particular difference is the most important)

Would a pencil-and-paper test have been able to accurately assess all these student abilities? It is difficult to see how such a test could address numbers 1 and 3 above.

FIGURE 5.4
Electric Circuits Performance Task

TASK:

Your job is to draw two circuits. One is a series circuit, and the other is a parallel circuit. Each circuit has one battery, wire, a switch, and two light bulbs. To prove that your drawings are correct, use the materials in your science kit to make each circuit the way you have drawn it. When you complete drawing and making the two circuits, you will answer these two questions:

1. What is the one important difference between a series and a parallel circuit?

2. Why do you think that difference is the most important difference?

PROCEDURE:

1. Review the assessment criteria for the Electric Circuits Performance Task (see Figure 5.5).

2. Draw the series circuit. Use arrows to show the path of the electricity in the circuit.

3. Make the series circuit you have drawn.

4. Draw the parallel circuit. Use arrows to show the path of the electricity in the circuit.

5. Make the parallel circuit you have drawn.

6. Answer the two questions.

Source: Hibbard, M.K. (2000). Performance-based learning and assessment in middle school science. Larchmont, NY: Eye on Education, p. 122. Source material has been abridged.

If science classes are to be about doing science, rather than just reading about science, then the use of performance tasks as the one shown in **FIGURE 5.4** represents a better match to the overall instructional goal.

The rating scheme for the Electric Circuits performance task is displayed in **FIGURE 5.5**. Depending on the purpose of the assessment, there are many different ways to judge how well students performed on the task.

FIGURE 5.5
Rating Scheme for the Electric Circuits Performance Task

PORTION OF THE TASK	TERRIFIC	OK	NEEDS WORK
DRAWING THE SERIES CIRCUIT	Drawn correctly so that it would work as a series circuit. Drawing is neat, organized, clear, and large.	Drawn correctly so that it would work. Drawing is not clear or neat enough.	Drawn so that it would not work as a series circuit.
CONSTRUCTING THE SERIES CIRCUIT	Circuit is made so it works as a series circuit. Circuit corresponds completely to drawing and uses the appropriate materials/quantities.	Circuit is made so that it works, but the construction only partially matches the drawing or only partially conforms to required materials list.	Circuit does not work as a series circuit.
DRAWING THE PARALLEL CIRCUIT	Drawn correctly so that it would work as a parallel circuit. Drawing is neat, organized, clear, and large.	Drawn correctly so that it would work. Drawing is not clear or neat enough.	Drawn so that it would not work as a parallel circuit.
CONSTRUCTING THE PARALLEL CIRCUIT	Circuit is made so it works as a parallel circuit. Circuit corresponds completely to drawing and uses the appropriate materials/quantities.	Circuit is made so that it works, but the construction only partially matches the drawing or only partially conforms to required materials list.	Circuit does not work as a parallel circuit.
WRITTEN ANSWER TO QUESTION ONE	Reason given clearly shows how the circuit wiring differs between the series and parallel circuit and refers to accurate drawings of the circuits to justify this reason.	Reason given clearly shows how the circuit wiring differs between the series and parallel circuit.	Reason given does not address differences in circuit wiring.
WRITTEN ANSWER TO QUESTION TWO	The answer contains a reference to the differences in the path of the electricity in the two different types of circuits and how this would affect the lighting of the bulbs.	The answer contains a reference to the differences in the path of the electricity in the two different types of circuits.	The answer does not contain a reference to the difference in the path of the electricity in the two different types of circuits.

Source: Hibbard, M.K. (2000). Performance-based learning and assessment in middle school science. *Larchmont, NY: Eye on Education, p. 123-124. Source material has been abridged and adapted.*

It is critical to be clear on the elements or features of a desired, strong performance. This rating scheme appears to emphasize the following elements of the task:

- Accuracy of drawings
- Neatness of drawings
- Accuracy of construction

- Correlation between drawing and construction
- Accuracy of differences between series and parallel circuits
- Justification of answers

These criteria are closely associated with, and accurately match the instructional goals addressed by the task. Therefore, this task has the potential to provide the teacher with valid assessment data relevant to the instructional goals.

Consider the following in implementing performance tasks in your class:

- Determine if a performance task is truly the best way to assess the learning target. For example, if basic knowledge is all that is required by the learning target, a multiple-choice question will be much more efficient in assessing this. Use performance tasks to measure instructional goals that call for a demonstration of abilities. For example, use performance tasks to demonstrate that students can actually "do science."

- Align the directions/procedures students will follow to the instructional goal/learning target. If you wish students to demonstrate lab safety skills, be sure that the procedures call for them to actually work in the lab (not just write about doing so).

- Align the rating scale to the instructional goal/learning target. For example, on the lab safety skills performance task, the rating scale may include such criteria as "wore safety glasses," "accurately followed directions," "used equipment appropriately," etc.

- Prepare students for performance tasks by allowing them to practice such tasks before grading them.

- Give students plenty of opportunities for feedback on their performances, by having students self-assess and peer-assess using the rating scale before you assess them with this same scale.

RESOURCES

The following books may provide you with some ideas for creating performance tasks:

Bosak, S.V. (2000). *Science is… A sourcebook of fascinating facts, projects, and activities.* Markham, Ontario, Canada: Scholastic Canada, Ltd.

Center for Performance Assessment. (2001). *Performance assessment series, Classroom tips and tools for busy teachers, Elementary school edition.* Englewood, CO: Advanced Learning Centers, Inc.

Center for Performance Assessment. (2001). P*erformance assessment series, Classroom tips and tools for busy teachers, Middle school edition.* Englewood, CO: Advanced Learning Centers, Inc.

Glatthorn, A.A. (1998). *Performance assessment and standards-based curricula: The achievement cycle.* Larchmont, NY: Eye on Education.

Hibbard, M. K. (2000). *Performance-based learning and assessment in middle school science.* Larchmont, NY: Eye on Education.

Rezba, R.J., Sprague, C., Fiel, R.L., & Funk, H.J. (1995). *Learning and assessing science process skills*, Third Edition. Dubuque, Iowa: Kendall/Hunt Publishing Company.

The following list of performance task examples may help stimulate your thinking about ways you might implement this type of assessment in your science classes:

- Use the equipment and materials provided to make the necessary measurements to calculate the density of each material.

- Create a working electrical circuit. Use this circuit to test the items in the bag. Report if each item is a conductor of electricity or a nonconductor.

- Ask one member of your group to step into the pan of lime chalk (the type used to mark lines on the football field) and then a) walk normally and b) run, leaving lime footprints on the asphalt of the parking lot. Measure this student's height. Determine a relationship between height and stride that might be useful to a detective at a crime scene.

- Using the stream table, demonstrate the creation of an ox-bow lake.

- From the genetic information provided to you, construct your creature, ensuring that this creature has the appropriate number of legs, eyes, body segments and appropriate colors of body and eyes, and appropriate shapes and sizes of antennae and tails. Pair with another student and "mate" your creatures. Construct models of all possible children from this pairing.

APPLICATION

Choose one of the performance task examples stated above (or make up one of your own). Then:

- Describe how this task aligns with a national or state standard (or standards).

- Explain why a performance task would be the best way to assess this standard (or standards).

- Develop procedures/student directions for the performance task that align with the standard (or standards).

- Create a list of criteria you would use to judge student achievement of this performance task. Explain how these criteria align with the standard (or standards).

- Describe the types of practice you will provide to students before grading the task.

- Describe how individual students will receive feedback on their progress.

Logs and Journals

Open-ended questions and performance tasks are ways to assess student learning at a particular point in the instructional process. Like these two assessment methods, logs can also be used periodically

to assess particular student actions or learning activities. Conversely, journals are dynamic (ongoing, continuous) assessment approaches that promote communication between the teacher and students, allow students to reflect on what they are learning, and foster students' active involvement in classroom activities.

Logs

A log provides documentary evidence of events and may also show the progression of such events. Students may be asked to keep scientific logs while running science experiments. The addition of growth factors to plants, as well as recorded heights of the plants at specified intervals are examples of data included in such logs. A detailed log can also help convince a teacher that, indeed, the student performed certain actions. In addition, it can reveal the exact nature of those actions. Because of its documentary properties, logs are frequently utilized to support student assertions or conclusions. They are commonly used within science fair experiments to document actions that students took in solving problems. The advantages logs bring to assessment include the following:

● They promote the achievement of instructional goals related to the nature of science. (Science is based on evidence; science findings are open to review.)

● They provide a track record, showing exactly what the student did and did not do.

● They help identify misconceptions and misunderstandings.

● They assist students in analyzing their own work. (If something doesn't work, students can track their own progress and then make necessary changes or corrections to ensure success.)

Journals

Journals are similar to logs, in that they provide a record of the progression of events. Generally, journals do not have the legalistic, evidentiary purpose of a log. While a journal documents events, it flavors those events with the opinions, feelings, and perceptions of the author. This "flavoring" of data with the consciousness of the author is what makes journals so useful to teachers. For example, a teacher may ask her science students to record "what you learned today." By reading the journals, the teacher can ascertain not only which concepts were conveyed to students, but also the level of understanding of the concepts achieved by her students. Such a journal entry would also encourage student meta-cognition as they assess their own levels of understanding.

Checking comprehension is only one use of journals. Journals can also be used to foster student reflection and critical thinking skills. Students can write journal entries as they begin to tackle a science problem, recording the question to be investigated and their own predictions. After the science investigation, students can revisit their predictions, explain why their predictions were or were not accurate, and reflect on the meanings or understandings they have derived from the investigation. Such reflections ask students to analyze their own thought processes and emphasize what changes in thinking have occurred. Such information on thinking changes is invaluable to the student and teacher.

Often, teachers provide prompts in order to encourage student writing in journals. The prompts listed in Figure 5.6 show how journal entries foster critical thinking at the higher levels of Bloom's Taxonomy.

FIGURE 5.6
Journal Prompts That Promote Critical Thinking Skills

JOURNAL PROMPT	LEVEL OF BLOOM'S TAXONOMY
1. Write three fun riddles for your friends. First, describe a solid, but don't tell your friend what it is. Give clues about the object. Start with its size and shape, how it feels, where you might find it, and its color, but remember not to name it. Then do the same for a liquid and a gas.	Application
2. Gravity is gone. Now everything is floating. How will this change the way you play during recess? Write a story about how you and your friends played without gravity.	Analysis
3. The sun is hiding, the crops are not growing, and everyone is going hungry. Write a fable that tells why the sun decided to hide and how he was convinced to come out again after seeing the effect he had on the germination, growth, and development of plants.	Synthesis
4. You have earned a lot of money, and you decide to use it to buy a pine forest so that those trees won't be cut down to make toothpicks. Your friend thinks you are silly to spend your money protecting the forest. Write a letter to your friend and explain why it is so important to protect this habitat. Remember to describe how many animals depend on the forest for food and shelter. Be sure to also say how humans benefit from the pine forest.	Evaluation

Source of journal prompts: Whited, A. M. (Ed.). (2005). Nonfiction writing prompts for science, Lower elementary. *Englewood, CO: Advanced Learning Press. #1, p. 29; #2, p. 39; #3, p. 57; #4, p. 53.*

Journals can also be used to assess attitudes toward science. Students can write their thoughts and feelings about class events. This use of journals as an expressive outlet for students is best seen as a two-way communication. That is, if the teacher does not respond to, probe, challenge, or ask for elaborations about the entries submitted, the full benefit of the journals will not be realized. Since students are being asked to share their own perceptions and opinions, there can be no wrong answers, and it is important that the teacher's responses reinforce this "no risk" environment. From such journals, teachers can gain valuable insights into the diverse interests, abilities, and attitudes

of the students in their science classes. Soliciting such information from students can also positively affect student motivation to learn.

The way journals are graded depends on the purpose of the journal and the age of the students. The act of keeping a journal can be considered as an objective in itself if the teacher believes that students need to structure, take charge of, or feel ownership in their own learning. The criterion for success on this objective might be the completion of the assigned journal entries or pages, not necessarily the quality of the entries. In this scenario, rather than *grading* the content of the journal, students are awarded points in a grading period if they have completed journal entries.

APPLICATION

Reflect on the units you teach throughout one school year. Write one journal prompt for each level of Bloom's Taxonomy that you could use with some or all of these units.

Portfolios

Portfolios, like logs and journals, contain collections of student work. A portfolio is defined as a purposeful, integrated collection of student work showing effort, progress, or a degree of proficiency. Physically, the portfolio is a container of evidence of a student's achievements, competencies, or skills. It is purposeful in that the collection is meant to tell a story about achievement or growth in a particular area. If multiple-choice and completion items are at one end of the assessment continuum representing very brief, quantitative, one-shot records of student achievement, then portfolios are at the other end, representing complex, qualitative, and progressive pictures of student accomplishments.

Why use portfolios? Portfolios may best be considered as tools to promote communication between the student and an outside audience (the teacher, parents, prospective employers, etc.) about student understandings, strengths, weaknesses, progress, and self-reflections. The use of portfolios, like any assessment method, starts with a consideration of these purposes. What are the objectives/standards that the portfolio will help the students achieve? Why are these objectives best assessed via a portfolio? What is the portfolio supposed to demonstrate? Several different examples of purposes for using portfolios in science classes are listed below:

RESOURCES

The following resource may help you in implementing portfolio assessment:

Barton, J. & Collins, A. (Eds.). (1997). *Portfolio assessment: A handbook for educators*. Menlo Park, CA: Addison-Wesley Publishing Company.

a. LEARNING TARGET: Ability to design an experiment

USE OF PORTFOLIO: To show progress in this ability over the year by including work on different assignments. Additionally, if the objective was to understand how students go about designing an experiment, the portfolio could contain all activities, drafts, and revisions leading up to the final design. Students could write reflections about their thinking at different stages in the design process.

b. LEARNING TARGET: Improving creative writing using science content knowledge

USE OF PORTFOLIO: To showcase the students' favorite/best pieces of creative writing. Such a portfolio could involve parents in helping students reflect on and choose their "best" pieces.

c. LEARNING TARGET: Read, summarize, and evaluate information in newspaper articles on science topics

USE OF PORTFOLIO: The portfolio may represent evidence of students' increasingly sophisticated efforts at critiquing these articles.

d. LEARNING TARGET: Show evidence of basic content knowledge

USE OF PORTFOLIO: Students could assemble all written tests into the portfolio and write a reflection piece after each test on how they could improve their performances.

e. LEARNING TARGET: Identify learning strengths/weaknesses

USE OF PORTFOLIO: Students could assemble their collections based upon their own strengths and weaknesses. Progress on weaknesses could be documented and reflected upon.

There is no particular right or wrong way to implement or use portfolios in the classroom. Rather, designing a portfolio represents a series of decisions. Some of the design questions to be answered after the instructional objective has been determined are listed below (Butler & McMunn, 2006, in press):

1. **Is the purpose of the portfolio to instruct, to support learning, or to assess?** Answering this question will help determine the types of artifacts the students will collect. For instance, a portfolio assembled solely for end of course grading purposes may contain only best-work pieces rather than a continuum of student work.

2. **What is the goal of using portfolios?** For example, is the goal to promote self-assessment? Student reflection on their own learning? Problem solving? Particular skills? Higher-order thinking? The goal for the portfolio will determine the design of the portfolio. For example, to promote self-esteem, a best works or memorabilia portfolio appears appropriate. However, if the goal is to improve students' proficiency at content-related skills, a skills portfolio would be best. A portfolio promoting student reflection would contain many subjective, journal-like artifacts, whereas a problem-solving portfolio would contain more objective work.

3. **What types of artifacts will be collected in the portfolio?** Will only written work be accepted, or will videotapes, posters, and computer disks also be acceptable? How many artifacts are necessary for documentation of a skill, goal, or purpose? The decisions made here will impact the size of the portfolio and its physical characteristics and may be influenced by the storage capacity of the classroom! If a file folder or binder is used, then perhaps only written work can be accepted. If an electronic portfolio is planned, all data may be stored on a disk or CD. The use of a single *quality* entry to prove a skill is recommended over the use of multiple entries for that one skill. (Surely if the student was successful once, he can be so again!) The number of artifacts also defines the type of

portfolio tool. If **all** student work is collected in the portfolio, the purpose is lost, and the assessment tool is just a notebook, not a portfolio.

4. **How will artifacts be selected for the portfolio?** Will the students select them, or will the teacher select them? How often will the portfolio be updated by adding artifacts? Must the students keep copies of all potential portfolio artifacts, or will the teacher maintain a file for this purpose? What are the criteria for selecting artifacts (how will the teacher or the students decide if a particular artifact documents a skill, purpose, or goal)? If the portfolio is intended to promote self-assessment, the students should choose the artifacts. Older students may keep their own working files, while younger ones may need help with this process since they have not yet developed organizational skills.

5. **How will students be oriented to the use of the portfolio?** The recommended method is to start slowly, giving students plenty of support and practice. A structured portfolio, in which expectations are explained to students, is recommended over a more open, unstructured design. Remember that change is difficult; be prepared for some student resistance to this new procedure. Perseverance and consistency are two key factors in the success of portfolio implementation.

6. **How will the portfolio be assessed?** A scoring guide, or rubric, is essential for this task, and this scoring guide should be shared with students before the assessment begins. However, if the work in the portfolio has already been assessed as individual pieces, should the overall portfolio also receive a grade?

7. **How will the information in the portfolio be housed?** Storage and handling of student information can be quite overwhelming, especially if the portfolio is one that travels with a student over an extended period. Many companies have developed software that helps manage the materials stored in a portfolio. Many of the student management systems used in districts also have a portfolio component housed within the management system for teacher and student use.

8. **What planning should be done before asking students to compile a portfolio?** Constructing the portfolio scoring guide before assigning the work will prevent student frustration, enhance the matching of the purpose to the artifacts, and ease the assessment task for the teacher. It is much simpler to assess an assignment if the plan for assessing it is written beforehand. Through careful planning, the teacher does not have to dread the moment he must confront a huge mound of papers, wondering what he will find in the contents of the portfolios his students have constructed.

Once these design questions are addressed, the portfolio can be planned and implemented. Like any of the other methods of evaluating student work, portfolios involve the development of criteria for judging good work. **FIGURE 5.7** lists criteria appropriate for a portfolio designed to address the instructional goal: Students will read, summarize, and evaluate information in newspaper articles on science topics.

FIGURE 5.7
Appropriate Criteria for a Newspaper Article Portfolio

Accurate summaries of at least eight newspaper articles

- All selected articles address science topics.
- All selected articles address DIFFERENT science topics.
- All original evaluations of articles are present.
- Each original (previously assessed) evaluation is followed by a student reflection on strengths/weaknesses of the evaluation (explains why the grade was justified).
- Ending reflection explains how weaknesses found in each original evaluation were addressed and summarizes changes that occurred over the course of the year in writing evaluations.

APPLICATION

Identify one instructional goal/learning target/national or state standard that could be addressed by implementing portfolio assessment. Work through the portfolio design questions to plan such an implementation. For number 6, it is sufficient to simply list important criteria, rather than develop the entire scoring guide.

Exhibitions/Projects

Exhibitions and projects provide opportunities for students to perform "real-life" tasks or wrestle with complex challenges. Such assessment types are usually of longer duration than performance tasks. Exhibitions and projects may run throughout a six-week grading period, or in the case of some culminating projects, extend to a year-long period. Exhibitions and projects have multiple steps (e.g. planning, researching, designing, implementing, etc.) and multiple criteria are needed to judge them. Students may be asked to structure an approach to a problem, investigate alternatives, produce a response, and justify approaches taken. More often than not, the tasks are assigned to teams of students, as that is how many "real-world" problems are tackled. Through such complex experiences, however, students develop into cooperative team members, problem-solvers, effective thinkers, quality producers, and self-directed learners.

Exhibitions and projects, like all performance-based assessments should be designed and selected to teach core curriculum content standards and should be scored using a rubric which was shared with students "up-front." Students can be given some choice as to the activities they will perform or the roles they will assume within the project. In addition, students should be required to meet interim deadlines for the project (which will aid the procrastinating student), to participate in planning the project (aid for the disorganized student), and to reflect on project activities (aid for the "surface" learner).

Of course, projects at different grade levels will vary in level of difficulty. The following examples may help in planning projects and exhibitions:

Elementary Level

- Students study the systems of the body and make life-size posters showing the location of major body organs.

- Students plan and design an appropriate backyard play area for a pet.

- Fourth-grade students run the school weather station, devising the weather instruments, using them to collect data, and making predictions about the weather, which are then reported during the morning announcements over the Public Address system at the school (National Research Council, 1999).

Middle School Level

- Students design and build model racecars to test the effect of tire sizes, gear ratios, and body design.

- Students choose a topic, plan, write, and produce a skit based on a scientific concept or principle.

- Teams of students compete to construct a framework that will support a full cup of water. The lightest framework using only allowed materials will win.

High School Level

- Science students reclaim an endangered estuary through clean up efforts and then turn the estuary into a "living classroom" for elementary students.

- Teams research one inherited human disorder and report to the whole class on mode of inheritance, symptoms, frequency of occurrence in the general population as well as specific populations, care needed for those suffering with the disorder, and effect on society (National Research Council, 1999).

- Students compete in science competitions in which they design and perform experiments to answer a research question.

- Students take on the roles of health care and support personnel in a hospital faced with a decision on whether to operate to separate conjoined twins.

- As a graduation requirement, individual students must perform research, write a research paper, and present their findings to an outside audience in order to complete their senior projects.

Projects and exhibitions, like the examples listed above, often serve two purposes, instruction and assessment. Just being involved in a project will require that students learn new knowledge and new skills. The grading criteria, shared with the students before they begin their work, can also be very instructional, in that such criteria spell out for students what quality work will entail. The teacher is using the project or exhibition, however, to find out what students know or are able to do, which is certainly an assessment function.

Projects and exhibitions are often ideally suited to the science classroom, as they require students to "do science," not simply read about it. Implementing such performance-based assessments in the classroom can be time consuming and challenging. The following suggestions, adapted from Davey and Rindone (1990), may assist teachers in planning:

1. Start with an issue, idea, scenario, or problem and test it by asking how important it is; how engaging it would be to students; how relevant it is to "real-life"; and what content areas could be learned within the content of the project/exhibition. Ask, "Does this align with my curriculum?" Considering the time investment, ensure that the project will provide rich data about mastery of a number of standards, if possible.

2. Begin to define the task more fully by asking what knowledge, competencies, skills, or dispositions students will have to use to complete the project or exhibition. This will focus attention on the project outcomes and on instructional objectives/learning targets. Revise and elaborate on the project until the learning targets align with the task that students are asked to perform.

3. Consider the context of the project/exhibition. What is the most appropriate medium for students to use (oral presentation, written product, computer simulation, a debate, a town meeting, etc.)? Should the task be done individually or in groups? Should experts from the community be accessed?

4. Consider the administration of the project/exhibition. What do students need to know before the work begins? What difficulties might be encountered? How will students receive assistance?

5. Consider how students' work on the task will be assessed. Will there be a checklist for work processes to guide students in the process of completing the task? What are the important features of a successful product? Who might assess student performance other than the teacher (peers, community professionals, other school staff, etc.)?

6. Talk over the proposed project/exhibition with colleagues and with students. Ask them to review the plan as well as the criteria that will be used to judge student work. Revise as needed, once the reviews are in.

APPLICATION

1. Choose a project that you have used in the past with your students. Reflect on the strengths and weaknesses of this project. Revise as needed.

2. Plan a new project for your science class. Use the six suggestions above to help you in this process.

One Last Look at Performance-Based Assessment

In the last three chapters, we have examined many types of performance-based assessments. All have been clustered, however, into three main categories:

- Observing students

- Soliciting information from students

- Evaluating student work

The performance-based assessments that fall into these categories all go beyond multiple-choice testing, and they have several qualities in common. They:

- Promote *doing* science.

- Stimulate higher-order thinking skills that ask students to do more than simply recall basic facts. Instead, these assessments ask students to apply, analyze, synthesize or evaluate.

- Present "real-life" challenges and promote the learning of "real-life" skills.

- Encourage self-assessment and self-reflection.

- Promote the development of self-directed learners, as many performance assessments provide students with the autonomy needed to evaluate their work.

- Provide more valuable insights into student thinking and student learning for teachers than do answers on multiple-choice tests.

- Allow instruction and assessment to overlap, as students learn to become quality producers while they are learning essential science content.

In order for performance-based assessments to be successful and effective, teachers must devote much "up-front" time to planning. These assessments must be carefully constructed to help students achieve particular standards, objectives, learning targets or expected outcomes of instruction. They must be aligned to the actual instruction that occurs on a day-to-day basis in the classroom. Finally, students must clearly understand the directions for the performance-based assessment, be provided time to practice such assessments, and be given grading schemes for the assessment before beginning work.

The next chapter in this publication, then, focuses on providing students with clear descriptions of teacher expectations, giving them timely and meaningful feedback on their performances, and reporting assessment results. To this end, **CHAPTER 6** covers **RUBRICS AND GRADING**.

RUBRICS AND GRADING

In this manual, we have reiterated several times that performance-based assessments are more time consuming and complex to create than simple, fact-based multiple-choice tests. However, we have encouraged teachers to implement these types of assessments because of the many benefits that result from such implementation. (See the list at the end of **CHAPTER 5**.) Such benefits far outweigh the difficulties involved in planning and implementing performance-based assessments. We must clearly articulate the difficulties so that teachers can a) be prepared for these and not ambushed or surprised by them, b) set realistic time schedules for planning and implementing performance-based assessments (allowing increased amounts of "up-front" planning time), and c) devise strategies to help allay difficulties.

In this chapter, we address one more difficulty: grading performance-based assessments. Like planning and implementing these types of assessments, grading provides many challenges that teachers using only multiple-choice tests will not encounter. Grading performance-based assessments will never be as easy as running the sheets through the Scan-tron machine, but there are methods and strategies teachers can use to make this process more manageable.

Reflect for a moment on an assignment you dreaded to grade. Then, read the scenario described in **FIGURE 6.1**.

FIGURE 6.1
The Science Fair Scenario

An interview with Micah, a middle school teacher, revealed the following scenario:

I once taught in a middle school where every 8th-grade student was required to complete a science fair project and enter it in the school science fair. A great deal of time in science class was, naturally, devoted to this project, particularly in the third grading period as the end of this period coincided with the date of the school science fair. On the day that the projects were due, parents and students descended on my classroom, bringing backboards and science equipment to display. By the end of the day, I had at least 160 backboards stacked in the back of my class. It was then that it really hit me—I had to grade these projects before the school science fair. I had about 3 days to do this. What would I do?

I began by setting up the backboards from three of my best students. By looking at their work, I started to make a list of criteria I could use to grade the projects. I assigned points to the criteria and then began to grade other backboards.

Unfortunately, in the Micah scenario, we find that insufficient planning time was devoted to designing the project. We recommend that teachers select the criteria to be used *before* assigning the project. In this manner, teachers can be assured that the grading criteria actually align with both the instruction and the instructional goals, and students can be informed of teacher expectations before they begin to work.

The technique itself is not a bad one: It is easier to create grading criteria by looking at student work. Teachers generally find it helpful to look at high-quality and low-quality student work in order to get a complete picture of the range of work possible. It would have been more helpful, however, if Micah had accessed previous student work (such as those projects from last year's science fair). That way the assessment design could occur before this year's students began to work, and the design could be shared with students up-front, not unfairly used to grade them after they completed their work.

Teachers always use a set of criteria to grade student work, even if they fail to articulate the criteria to themselves or to students. Grading schemes help make these hidden criteria visible to students. Once students clearly understand the expectations, they can more easily work toward achieving these expectations. Therefore, this chapter is devoted to making grading criteria visible and accessible. Several different methods are available for informing students of grading criteria before they begin work. These include point systems, checklists, and rubrics.

Point Systems

A point system assigns points for certain features of the student's response. Open-ended questions are often scored with this approach because points can reflect partial as well as full credit for a response.

For example, if third-grade students are given the appropriate equipment and asked to find out if stirring makes any difference in how fast sugar cubes and loose sugar dissolve (NAEP, 1986), the point system may be similar to the one shown in **FIGURE 6.2**. Here, the points are used to score students' oral responses to the question: Did stirring make a difference in how fast the two types of sugar dissolved?

FIGURE 6.2
Scoring the Sugar Question

POINTS AWARDED	DESCRIPTION OF RESPONSE
4	If the response states that both types of sugar dissolve faster when stirred, but loose sugar still dissolves faster than cubes
3	If the response indicates that stirring makes a difference but doesn't describe the fact that loose sugar dissolves faster than cubes
2	If the response describes the relative speed (loose dissolves faster) but not the effects of stirring OR if the response just describes what happens (e.g., stirring makes the cubes come apart)
1	Incorrect response
0	No response

Point systems are useful, then, in scoring responses where partial credit may be given. They are most helpful in scoring short answer open-ended questions, however, rather than in grading essay questions. Extended response essay questions entail more complex, detailed answers and therefore engender the need for a more complex grading system (see the Rubrics section below).

Checklists

Checklists, like point systems, are often used when complex responses are not expected from students. Checklists, however, are more likely than point scales to be used in judging student actions or behaviors. For example, a checklist can be used to indicate that a student has effectively completed the steps involved in a task or demonstration. **FIGURE 6.3** displays a checklist that could be used when evaluating student knowledge of the parts of the microscope and when evaluating students on the proper operation of the microscope.

FIGURE 6.3
Microscope Checklist

CORRECTLY IDENTIFIES:

❑ Stage

❑ Stage clips

❑ High power objective

❑ Eye piece

❑ Coarse adjustment knob

❑ Fine adjustment knob

PERFORMS THE FOLLOWING OPERATIONS CORRECTLY:

❑ Swings low power objective into place

❑ Places slide on stage and secures with clips

❑ Uses coarse adjustment to move low power objective as far down as possible

❑ Looks through eyepiece

❑ Uses coarse adjustment knob to raise low power objective until object is in focus

❑ Uses fine adjustment knob to bring object into focus

❑ Swings high power objective into place

❑ Uses fine adjustment knob to bring object into focus

As the example in **FIGURE 6.3** shows, checklists can be useful in evaluating simple student actions, particularly ones where there are a limited number of options. Examples include noting the presence or absence of certain actions (secures slide with clips), the sequence of actions (uses coarse adjustment knob to move low power objective down before attempting to focus with fine adjustment knob), or whether a student has given a correct or incorrect answer (correctly identifies microscope part).

Checklists are also effective in getting students to check their own work. For example, prior to taking up notebooks, a teacher may provide a checklist to students listing all the assignments that should be included. Students can use the checklist to evaluate the completeness of the notebook before handing it in.

Rubrics

So far, we have examined grading schemes that can be used in scoring relatively simple performance-based assessment, as short answer questions or simple student actions with limited options. Many performance-based assessments call for long or highly complex student responses. For these types of responses, a rubric is useful because a rubric can take into account many different criteria for judging student work. Rubrics can help students begin to understand that there are levels of quality to their work and to their thinking. Rubrics can aid students in learning that high-quality work is important. Too many students just turn in work to get it done, get the check for completion, and don't worry about crafting a high-quality response or product. Taking their work lightly can come back and hurt them down the road, in college or in their chosen professions. Rubrics, because they demonstrate the various levels of proficiency and because they define high-quality work, can help students to craft the desired high-quality products and responses.

There are two main types of rubrics: analytical and holistic rubrics. This section will define and give examples of each type before closing with a discussion of the advantages and disadvantages of each.

Analytical Rubrics

The most common format of an analytical rubric consists of a list of criteria down one side with proficiency levels listed and described across the page. **FIGURE 6.4** provides a simplistic example of this format for an analytical rubric used to score a physics problem-solving task. In this particular example, descriptions of the proficiency levels (Exceeds goals, Meets goals, Approaches goals, Goals not yet met) are not present.

FIGURE 6.4
Physics Problem-Solving Rating Scale

CRITERIA	EXCEEDS GOALS	MEETS GOALS	APPROACHES GOALS	GOALS NOT YET MET
1. Correctly states the problem and identifies the information needed to solve it and the steps needed to arrive at a solution.				
2. Produces reasonable estimates of data values that were not identified by the teacher but needed for the solution to the problem.				
3. Applies concepts and formulas related to motion (velocity, acceleration, average speed).				
4. Makes accurate conversions as needed to solve the problem.				
5. Communicates conclusions clearly, using examples as needed.				

Source: Adapted from Davey & Rindone (1990), "Anatomy of a Performance Task." Presented at the American Educational Research Association meeting, Boston, MA, from materials developed by the Bureau of Evaluation and Student Assessment, Connecticut State Department of Education.

As this rating scale shows, students have the opportunity to receive scores in several different dimensions (i.e., they will get a score for each criterion listed). Analytical rubrics can help both teachers and students diagnose strengths and weaknesses in a performance. This enables students to target the areas of their performances that need to be improved.

A rating scale like the one displayed in **FIGURE 6.4**, however, may make too many assumptions about clarity to be useful to students. In other words, since no descriptions of the proficiency levels are present, students may have difficulty in understanding teacher expectations. **FIGURE 6.5** displays a rubric that does provide descriptions at each level of proficiency.

FIGURE 6.5
Poster Displaying Science Principle Analytical Rubric

CRITERIA	EXCELLENT (4 POINTS)	ACCEPTABLE (3 POINTS)	RE-DO (2 POINTS)
SCIENTIFIC ACCURACY	Scientific principle is accurately and completely stated and supported by well-explained, accurate examples from real life.	Scientific principle is accurately stated and supported by accurate examples from real life. Examples are not as clear as they coud be.	Scientific principle is inaccurately stated *OR* no examples are present.
ORGANIZATION	Information is organized logically, in that a theme (or themes) is easily seen (e.g., chronological theme, classification theme, etc.)	Information is somewhat organized but organization is not consistent throughout.	No theme is present. No organization is apparent.
GRAPHICS	Poster contains at least 3 graphics. Graphics support the understanding of the principle by both providing real-life examples and giving schematics explaining how the principle functions.	Poster contains graphics, but graphics don't consistently show understanding of the principle (weak example or unclear schematics).	Poster contains no graphics or graphics that don't relate to the principle.

APPLICATION

Compare and contrast the scoring examples shown in FIGURES 6.4 and 6.5.

- Which approach would be more useful to students in planning and implementing their performances/products?

- Which would be more useful in encouraging student assessment?

- Which gives the clearer picture of teacher expectations?

Holistic Rubrics

Rather than assigning separate scores for each important aspect of a performance, holistic rubrics consider all the criteria simultaneously and result in a single summary rating or grade. This type of rubric may be more appropriate when the purpose is to provide students with an overall index of their performance on a task or product. Holistic rubrics are also used more often in culminating assessments, after students have already received feedback on their progress during the learning process. Figure 6.6 shows an example of a holistic rubric for the same performance task used in **FIGURE 6.5** (creating a poster displaying a scientific principle).

FIGURE 6.6
Poster Displaying Science Principle Holistic Rubric

OVERALL SCORE	DESCRIPTION OF PERFORMANCE
4	Scientific principle is accurately and completely stated. Student gives well-explained and accurate examples from real life. Poster is organized logically, in an easily discernable theme. Poster contains at least three graphics which illustrate both real-life examples of the principle and provide schematics to explain how the principle functions.
3	Scientific principle is accurately stated. Student gives several accurate examples from real life but they could be explained more clearly. Poster is organized, but organization or theme is inconsistent in places. Student can explain the organizational theme, however. Poster contains several graphics that both illustrate real-life examples and provide schematics.
2	Scientific principle is stated correctly, but may be somewhat incomplete. Student gives at least two accurate examples from real life but the explanations are somewhat unclear. Poster is partially organized, but no discernable theme is present. Poster contains graphics, but the graphics are weak examples.
1	Scientific principle is inaccurately stated or fewer than two accurate examples are given. No organizational theme is present. Poster contains no effective graphics.

APPLICATION

Compare the analytical rubric displayed in **FIGURE 6.5** with the holistic rubric shown in **FIGURE 6.6**.

As a teacher, consider:

- What are some advantages and disadvantages to creating and using analytical rubrics?
- What are some advantages and disadvantages to creating and using holistic rubrics?

Ask students to consider:

- What are some advantages and disadvantages in receiving and being assessed using analytical rubrics?
- What are some advantages and disadvantages in receiving and being assessed using holistic rubrics?

Advantages and Disadvantages of the Two Types of Rubrics

One distinct advantage of analytical rubrics is that this type of rubric gives more meaningful feedback to students. Since students receive scores for each criterion, it is easy for students to discern their strengths and weaknesses. For the teacher, however, analytical rubrics can be a little more time consuming to use, as each student must receive a score in every dimension on the rubric. Because of the potential for more defined feedback, however, analytical rubrics are very useful to use during the learning process. They provide structure and support for learning and help students clearly understand the teacher's expectations. They are very helpful in promoting student self-reflection and self-assessment, as students can check their own work against the descriptions of high-quality work provided on the rubric.

Holistic rubrics are usually easier for teachers to use in giving grades. However, because all the characteristics of the work are lumped together to create a single score, it is harder for students to understand their strengths and weaknesses.

Final Thoughts About Grading Schemes

As a teacher begins to implement performance-based assessments, it is important to examine alternative ways of grading assignments. Multiple-choice items can be scored very objectively. The student is offered a fixed number of options, and the option selected is compared to a scoring key (containing the "right" answers). Given the scoring key, any teacher would score the multiple-choice items in the same way. Performance-based assessments such as open-ended questions, journals, portfolios, performance tasks, and exhibitions and projects often have no one right answer. Therefore, a different way to score these items must be developed. A list of criteria is needed, along with descriptions of proficiency levels. Developing point systems, checklists, or rubrics help define quality work for the students and help the teacher score assignments from different students using the same criteria. Such grading schemes therefore enhance the objectivity of the teacher in judging student work.

The following guidelines may help you in deciding when to use a particular grading scheme and in creating high-quality grading schemes that can enhance student performance:

● Examine the task and choose the most appropriate format for the grading scheme. If a short answer is needed, a point system could be the best choice. For limited option tasks (the student either did a particular action or did not do a particular action), checklists may be best. For highly complex or extensive responses/behaviors, a rubric may be needed.

● Make a list of criteria to use in scoring student work by examining work from past years. Look at both high-quality and low-quality work to establish a range. Think about the weaknesses that occurred most often in past work and reflect on how your new grading scheme might help prevent this weakness from re-occurring.

- Provide scored examples of student work to your present students when introducing a new grading scheme. Looking at such work and seeing the scores it received can help students understand teacher expectations.

- Support students while they are learning by using the type of grading scheme that will give them the most meaningful feedback.

- For new tasks, try involving the students in helping you create the grading scheme.

- Always distribute the grading scheme to students before they begin to work.

- Encourage students to self-assess, using the grading scheme provided, before handing in work.

Compiling Grades and Communicating Student Achievement

Once all the multiple-choice and performance assessments are done, teachers compile individual grades into one overall grade for a marking period. The way that this compilation is done is often explained in a grading policy statement, usually distributed at the beginning of the school term. Overall grades then appear on report cards, which are sent home to parents.

For parents and students to understand how a student is progressing academically, they must be able to accurately interpret the overall grade shown on the report card. Brookhart (2004, p.7) states that "grades and other communication about student achievement should be based on solid, high-quality evidence. Teachers should be able to describe that evidence and explain how they arrived at any judgments about the quality of student work." The grading schemes discussed previously in this chapter will help teachers articulate the types of evidences they have collected and explain how student work was judged.

Another grading dilemma to consider is how *many* grades are needed to constitute the "solid, high-quality evidence" that Brookhart recommends. There is no magic number of grades; the teacher must be the judge of the amount of grades that is sufficient. However, using only one measure (e.g., a one-hour, paper-and-pencil exam) to determine a report card grade is clearly insufficient evidence. At the other extreme, assessing student performance daily would not provide students with the time needed to develop competences and skills preparatory to being assessed.

When implementing performance-based assessments, teachers may find difficulties in using point systems, checklists, and rubrics, as these often contain scores that must be converted to the standard grading scale used by the school. Students are more used to getting

RESOURCES

The following books may help teachers reflect upon their current grading practices:

Brookhart, S.M. (2004). *Grading.* Upper Saddle River, NJ: Pearson Merrill Prentice Hall.

Marzano, R.J. (2000). T*ransforming classroom grading.* Alexandria, VA: Association for Supervision and Curriculum Development.

O'Connor, K. (2002). *How to grade for learning: Linking grades to standards.* (2nd Ed). Arlington Heights, IL: Skylight Training and Publishing, Inc.

letter grade (A, B, C, D, F) or percentage scores (94, for example). Therefore, it is important to include a grade conversion chart with any grading scheme. Converted scores must be compiled in some way in order to formulate an overall report card grade. The procedure for formulating final grades is often spelled out in a grading policy statement. Such grading policies often include:

- How missing work will be counted

- The weight of particular assignments (as tests may count more than homework)

- The grading scale used by the school (percentage points needed to get an "A," etc.)

For an example of the use of a weighting scheme, see the weighting system shown in **FIGURE 6.7.** In this sample, demonstrating knowledge on tests represents 37% (100/270 points) of the total grade; science process skills, maintaining a journal, and completion of an extended group project each represent 19% (50/270 points); and creative writing represents 6% (20/270 points). The proportion of the total grade accounted for by individual assessments should communicate the relative importance of different desired outcomes (that is, more important outcomes carry more weight).

APPLICATION

Revise or create a grading policy statement for your classes. Consider increasing the amount of performance-based assessments you will implement this year.

How might this change your current grading policy?

How will weights of assignments change?

How can you use the grading policy to ensure that students actually *do* science?

FIGURE 6.7
Sample Grading Period Weighting System

Student A

DESCRIPTIONS OF ASSIGNMENTS	MAXIMUM POINTS AVAILABLE	POINTS EARNED	WEIGHT OF ASSIGNMENT	STUDENT A SCORE/MAXIMUM SCORE
Paper-and-pencil test on electricity	50	40	1	40/50
Performance test on electricity (making circuits)	25	20	2	40/50
Weekly lab assignments on science process skills (5 assignments x 10 pts each)	50	45	1	45/50
Two creative writing tasks (10 pts each)	20	20	1	20/20
Journal	50	50	1	50/50
Extended group project	50	45	1	45/50
Totals	**245**	**220**		**240/270**

The weighting system used in deriving report card grades should be related to the course objectives and explained to students so that they know the goals. In the example shown in **FIGURE 6.7,** students might be informed at the beginning of the grading period of the instructional objectives to be taught and the assessments to be used. The number of points needed for the different grade symbols (number of points to get an A; number of points to get a B, etc.) could also be communicated.

In such a weighting system, it is also important to stay flexible so as not to penalize students for poor quality assessments. For example, if students were told that a certain number of points constituted an A, but no students earned this many points due to a poorly constructed test, some adjustment to the point system would have to be made.

Student achievement status on important instructional objectives can be communicated in ways other than a single report card grade in science. Some teachers find that grades, although required by policy, are not particularly helpful in conferencing with students and parents about students' performance on specific learning goals. The actual grading schemes (e.g., checklists, rubrics) as well as anecdotal notes or observation instruments can be used in addition to grades or as alternative means of reporting to parents.

Getting Started

Traditional practices in assessment are based on beliefs about the purpose of education that are currently being publicly discussed and challenged. Performance-based assessments described in this manual meet some of the identified challenges related to developing students' higher order thinking skills. Assessment practices of teachers do not necessarily change once people become aware of the need for change. Change does not happen the day after an afternoon of inservice training. Generally, change is a slowly evolving process that occurs through experience, dialogue, and reflection.

Teachers need time to try new assessments, time to reflect on the success or failure of these new methods, and time to make revisions. Just as student learning is an individual process that is personally constructed, so is teacher learning about assessment practices. Changing assessment practices is not a simple, linear, lock-step process that all teachers follow in a prescribed manner. Rather, it is a process of becoming more purposeful about:

- Desired student outcomes in science

- The design of learning experiences in support of these outcomes

- The use of assessment methods that match well with desired outcomes

- The use of grading systems that reflect student achievement on these outcomes
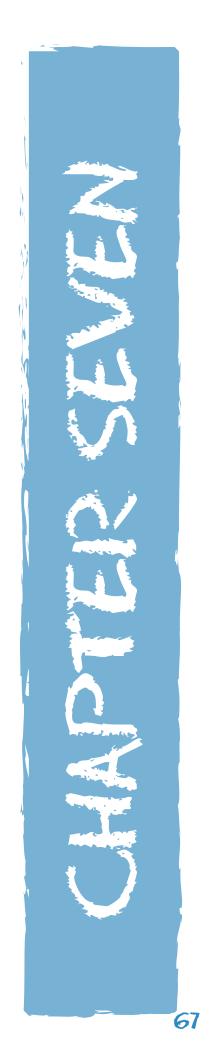
What are some contexts in which this more purposeful thinking about student assessment might be developed?

Some districts have initiated district-wide staff development efforts in assessment. The literature on professional learning suggests that a good staff development program is sustained over time. Teachers are more likely to change in a collegial setting with sustained administrative support (Loucks-Horsley, Brooks, Carlson, Kuerbis, Marsh, Padilla, Pratt, & Smith, 1990).

This kind of model might involve bringing together a volunteer group of lead science teachers from several schools who, with a facilitator:

- Spend a day on an overview of assessment (outcomes, methods, rubrics) as provided in this publication.

- Spend a day reflecting on science education goals and beginning to develop or adapt assessments to try out (e.g., observation forms, interview protocols, open-ended questions, performance tests, journal criteria, portfolio tasks, exhibition and projects).

- Come together as a group on a regular basis to share experiences, demonstrate the assessments developed and the student results obtained, continue to develop or find new assessments, and identify areas in which further assistance or information is needed.

The following year, the lead teachers could start a similar process for interested science teachers within their own schools.

CHAPTER SEVEN

Teachers, either individually or in informal groups could begin to reflect on their assessment practices. Incorporating performance-based assessment into the classroom may be easier if experiences, concerns, and frustrations are shared with colleagues. Sharing successful tasks and methods with other teachers also increases the number of assessments available.

There is no right place to start with assessment. There are many activities, depending on the prior experience, time constraints, interest, and resources of the teachers involved, which represent jumping-off points for changing or at least reflecting on assessment practices.

APPLICATION

Listed below are some examples of activities that might get conversations started about assessment practices:

- Articulate one very important desired student outcome (refer to CHAPTER 2). For example, a teacher might be interested in how well students can develop and test hypotheses in the content area under study. Review the assessment methods described in CHAPTERS 3–5 and choose an approach to assessing students' competence on this dimension that has not been tried before. Try the assessment approach to see what can be learned about student performance and about the assessment method chosen.

- Experiment with a format for a course syllabus that outlines for students the major goals you have for their performance and how their performances on these goals will be assessed and report card grades will be derived.

- Start a list of advantages and disadvantages of each of the assessment methods described in **CHAPTERS 3–5**. What do you feel you need to know from someone who has tried each method before you go any further?

- Develop a chart (as the one in FIGURE 6.7) showing how you combine assessment data in obtaining student report card grades. What kind of weighting system are you using?

- Analyze the tests you have used in the past. Try to improve the items used on these test after referring to the information provide in this manual about open-ended questions or performance tasks. Consider how you might improve or make more explicit the rubrics for the items.

- Start a folder of assessment samples from released state tests or item banks, other teachers, district tests, or published articles. Critique them for your purposes. Save samples of student work that demonstrate different levels of proficiency. Discuss these samples with other teachers to gain a shared view of mastery.

- Review the hands-on, experiential, or lab activities you use with your students. Identify the most essential ones, and experiment with rubrics that could be used to assess student performance on these tasks.

The process of incorporating and using a broader array of assessment methods can sharpen teachers' thinking about the meaning of student success in science. It can also result in improvements in the quality of instruction teachers design for students. Finally, if teachers are explicit and purposeful about their goals, students are more likely to evaluate the quality of their own work.

The benefits of experimenting with a variety of assessment methods lie as much in the conversations they engender between teachers and students and among teachers as they do in the information they provide on student competence. Students as well as teachers often become empowered as assessment becomes a dynamic, interactive conversation about progress through the use of interviews, journals, exhibitions, and portfolios. Through these assessment methods, teachers relate to students more as facilitators, coaches, or critics rather than as authority figures that dispense all information and knowledge.

Black, P. & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment.* London: School of Education, King's College. See also Black, P. & William, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80 (2), p. 139—148.*

Bloom, B.S. (1956). Taxonomy of educational objectives, the classification of educational goals, cognitive and affective domain. New York: David McKay Co. Brookhart, S.M. (2004). *Grading.* Upper Saddle River, NJ: Pearson Education, Inc.

Brookhart. S.M. (2004). *Grading.* Upper Saddle River, NJ: Pearson Merrill Prentice Hall.

Butler, S.M. & McMunn, N.D. (2006) *A teacher's guide to classroom assessment: Understanding and using assessment to improve student learning.* San Francisco, CA: Jossey-Bass, Inc.

Davey, B., & Rindone, D.A. (1990). Anatomy of a performance task. Presented at the American Educational Research Association Annual Meeting, Boston, Mass.

Florida Dept. Of Education, Sunshine State Standards, Grades 6–8. Retrieved Aug. 15, 2005 from **www.firn.edu/doe/curric/prek12/pdf/science6.pdf**

Glatthorn, A.A. (1998). *Performance assessment and standards-based curricula: The achievement cycle.* Larchmont, NY: Eye on Education.

Gronluncd, N.E., & Linn, R.L. (1990). *Measurement and evaluation in teaching* (5th ed.) New York: Macmillan.

Hibbard, M.K. (2000). *Performance-based learning and assessment in middle school science.* Larchmont, NY: Eye on Education.

High Success Network Training Materials provided at the Outcome-Based Education Summer Conference, Charlotte, NC. (1992). (Available from the High Success Network, P.O. Box 1630, Eagle, CO 81631).

Kohn, A. (1999). *Punished by rewards.* New York, NY: Houghton Mifflin Company.

Loucks-Horsley, S., Brooks, J., Carlson, M., Kuerbis, P., Marsh, D., Padilla, M., Pratt, H., & Smith, K. (1990). *Developing and supporting teachers for science education in the middle years.* Andover, MA: The National Center for Improving Science Education, The Network.

McTighe, J., & Wiggins, G. (2004). Understanding by design: Professional development workbook. Alexandria, VA: Association for Supervision and Curriculum Development.

NAEP. (1986). *A pilot study of higher-order thinking skills assessment techniques in science and mathematics.* Princeton, NJ: Educational Testing Service.

National Research Council. (1999). *National science education standards.* Washington,DC: National Academy Press.

Science Framework for the 2005 National Assessment of Educational Progress. Retrieved August 14, 2005 from **www.nagb.org/pubs/s_framework_05/ch2.html**

Stiggins, R. (2001). *Student-involved classroom assessment.* Upper Saddle River, NJ: Prentice-Hall, Inc.

Texley, J. & Wild, A. (1997). (Eds.). *NSTA pathways to the science standards: Guidelines for moving the vision into practice.* Arlington, VA: National Science Teachers Association.

Tobin, K., Kahle, J., & Fraser, B. (1990). *Windows into science classrooms.* Bristol, PA: Falmer Press.

Welch, W. (1979). Twenty years of science curriculum development: A look back. *In Review of Research in Education,* Berliner, D. (Ed.). Washington, DC: American Educational Research Association.

Whited, A. M. (Ed.). (2005). *Nonfiction writing prompts for science, Lower elementary.* Englewood, CO: Advanced Learning Press

Wiggins, G. (1989). A true test: toward more authentic and equitable assessment. *Phi Delta Kappan, 70(9),* 703-713.

Wiggins, G. (1992). On performance assessment: a conversation with Grant Wiggins. *Educational Leadership, 49(8),* 35-37.

Yager, R.E., & Kellerman, L.R. (1992). *Assessment ideas for science in six domains.* (Available from the Science Education Center, Van Allen Hall, University of Iowa, Iowa City, IA 52242).

REFERENCES

The SERVE Center at UNCG, under the leadership of Dr. Ludwig David van Broekhuizen, is an education organization with the mission to promote and support the continuous improvement of educational opportunities for all learners in the Southeast. The organization's commitment to continuous improvement is manifest in an applied research-to-practice model that drives all of its work. Building on research, professional wisdom, and craft knowledge, SERVE staff members develop tools, processes, and interventions designed to assist practitioners and policymakers with their work. SERVE's ultimate goal is to raise the level of student achievement in the region. Evaluation of the impact of these activities combined with input from stakeholders expands SERVE's knowledge base and informs future research.

This rigorous and practical approach to research and development is supported by an experienced staff strategically located throughout the region. This staff is highly skilled in providing needs assessment services, conducting applied research in schools, and developing processes, products, and programs that support educational improvement and increase student achievement. In the last three years, in addition to its basic research and development work with over 170 southeastern schools, SERVE staff provided technical assistance and training to more than 18,000 teachers and administrators across the region.

The SERVE Center is governed by a board of directors that includes the governors, chief state school officers, educators, legislators, and private sector leaders from Alabama, Florida, Georgia, Mississippi, North Carolina, and South Carolina.

SERVE's operational core is the Regional Educational Laboratory. Funded by the U.S. Department of Education's Institute of Education Sciences, the Regional Educational Laboratory for the Southeast is one of ten Laboratories providing research-based information and services to all 50 states and territories. These Laboratories form a nationwide education knowledge network, building a bank of information and resources shared and disseminated nationally and regionally to improve student achievement. SERVE's National Leadership Area, Expanded Learning Opportunities, focuses on improving student outcomes through the use of exemplary pre-K and extended-day programs.

SERVE Center
at the University of North Carolina
at Greensboro

P.O. Box 5367 • Greensboro, NC 27435
800-755-3277 • 336-315-7400 • Fax 336-315-7457

Ludwig David van Broekhuizen, Ph.D., Executive Director

WWW.SERVE.ORG