DOCUMENT RESUME

ED 363 622 TM 020 612

AUTHOR	McColskey, Wendy; O'Sullivan, Rita
TITLE	How To Assess Student Performance in Science: Going beyond Multiple-Choice Tests. A Resource Manual for Teachers.
INSTITUTION	Southeastern Regional Vision for Education (SERVE), Tallahassee, FL.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	Jun 93
CONTRACT	RP91002010
NOTE	72p.
AVAILABLE FROM	SERVE, 345 South Magnolia Drive, Suite D-23, Tallahassee, FL 32301-2950.
PUB TYPE	Guides - Non-Classroom Use (055)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Educational Assessment; Educational Change; Elementary Secondary Education; *Evaluation Methods; Grading; Interviews; Multiple Choice Tests; Observation; *Outcomes of Education; *Performance; Portfolios (Background Materials); Science Education; *Science Instruction; Science Teachers; Scoring; Self Evaluation (Individuals); *Student Evaluation; Teaching Guides
IDENTIFIERS	Open Ended Questions; *Performance Based Evaluation

ABSTRACT

The purpose of this manual is to encourage discussions among science teachers about desired student outcomes in science and assessments appropriate to these outcomes. Once teachers know clearly what they hope to accomplish with their students, choosing appropriate assessment measures follows. Assessment methods described center on observing students informally or formally; asking students questions through interviews and self-assessments; and evaluating student work through open-ended questions, performance tests, portfolios, and other assessments. Chapters provide information and examples for the following topics: (1) current views on assessment; (2) desired student outcomes in science; (3) student assessment methods; (4) rubrics and grading; and (5) getting started. Twelve figures illustrate the discussion. An appendix discusses reform in curriculum and instruction. (Contains 28 references.) (SLD)

SERVE

R&D Improving Practice in Education

How To Assess Student Performance in Science:

Going Beyond Multiple-Choice Tests

SERVE
SouthEastern Regional Vision for Education

How To Assess Student Performance in Science:

Going Beyond Multiple-Choice Tests

A Resource Manual for Teachers

by

Wendy McColskey and Rita O'Sullivan

June, 1993

SERVE

SouthEastern Regional Vision for Education

Affiliated with the School of Education,
University of North Carolina at Greensboro and
the Florida Department of Education

This document was produced with funding from the Office of Educational Research and Improvement, U.S. Department of Education, under contract no. RP91002010.

Table of Contents

About the SERVE Laboratory

SERVE, the SouthEastern Regional Vision for Education, is a coalition of educators, business leaders, governors and policymakers who are seeking comprehensive and lasting improvement in education in Alabama, Florida, Georgia, Mississippi, North Carolina, and South Carolina. The name of the Laboratory reflects a commitment to creating a shared vision of the future of education in the Southeast.

The mission of SERVE is to provide leadership, support, and research to assist state and local efforts in improving educational outcomes, especially for at-risk and rural students. Laboratory goals are to address critical issues in the region, work as a catalyst for positive change, serve as a broker of exemplary research and practice, and become an invaluable source of information for individuals working to promote systemic educational improvement.

Collaboration and networking are at the heart of SERVE's mission; the laboratory's structure is itself a model of collaboration. The laboratory has four offices in the region to better serve the needs of state and local education stakeholders. SERVE's Greensboro office manages a variety of research and development projects that meet regional needs for the development of new products, services and information about emerging issues. The development of this manual was funded through such an R & D effort. The laboratory's information office, affiliated with the Florida Department of Education, is located in Tallahassee. Field services offices are located in Atlanta, Greensboro, Tallahassee and on the campus of Delta State University in Cleveland, Mississippi.

To request publications or to join the SERVE mailing list and receive announcements about laboratory publications, contact the SERVE office in Tallahassee (address below).

SERVE
P.O. Box 5367
Greensboro, NC 27435
800-755-3277; 919-334-3211
FAX: 919-334-3268

SERVE
345 S. Magnolia Drive, Suite D-23
Tallahassee, FL 32301-2950
800-352-6001: 904-922-2300
FAX: 904-922-2286

SERVE
41 Marietta St., NW
Suite 1000
Atlanta, GA 30303
800-659-3204: 404-577-7737
FAX: 404-577-7812

SERVE
Delta State University
Box 3121
Cleveland, MS 38733
800-326-4548: 601-846-4384
FAX: 601-846-4016

Acknowledgments

This manual resulted from an R&D field experience which involved working with science teachers to develop performance assessments for a science curriculum reform project. The participation of these schools and teachers was invaluable to the development of this manual.

Conway Middle School, Horry School District, Conway, South Carolina

Josephine Dozier, Teacher
Lesia Hughes, Teacher
Helen McCormick, Teacher
Mary Miller, Teacher
Alberta Sampson, Teacher
Dancey Spivey, Teacher
Jazmin Vega-Cortes, Teacher
Gil Stefanides, Principal

Graham Middle School, Alamance County School District, Graham, North Carolina

Rosemarie Gavasto, Teacher
Nancy Smith, Teacher
Patsy Smith, Teacher
Sam Fowler, Principal

Greene County Middle School, Greene County School District, Snow Hill, North Carolina

Lorraine Brann, Teacher
Lori Braxton, Teacher
Susan Harrel, Teacher
Becky Holleman, Teacher
Candace Turk, Teacher
Brenda Shirley, Assistant Principal

Performance assessments developed by the teachers listed above were piloted in schools participating in the North Carolina Project for Reform in Science Education, a national science curriculum reform project initiated by the National Science Teachers Association and funded by the National Science Foundation. Thus, these teachers provided another sounding board for us as we experimented with student assessment methods in a "live" curriculum reform setting.

Participating schools and teachers from the North Carolina Project for Reform in Science Education included the following:

A.G. Cox Middle School, Pitt County School District, Greenville, North Carolina

Rita Elliot
Terry Hardin

Hanes Middle School, Winston-Salem/Forsyth County School District, Winston-Salem, North Carolina

Greg Bitting
Joneice Carroll
Barbara Hendrix
Gordon Knight
Barbara Massey

Noble Middle School, Virgo Middle School, New Hanover County School District, Wilmington, North Carolina

Beth Bruton
Ann Bradshaw
Mary Jessup
Sandy Klein

Northeast Middle School, Guilford County School District, Greensboro, North Carolina

Cindi Fields
Adele Hollerman
Gay Sartin
Sherri Smith
Katherine Watlington

Piedmont Open Middle School, Charlotte-Mecklenburg County School District, Charlotte, North Carolina

Amy Donnelly
Stacy Owens

Valley Springs Middle School, Buncombe County School District, Asheville, North Carolina

Martha Brookshire
Joel Kelley
Steven Kirstein

We also appreciate reviews of the publication provided by the following:

James Altschuld, National Center for Science Teaching and Learning, Columbus, Ohio
Keith Beamon, Director of Instruction, Johnston County Schools, North Carolina
Lorraine Brann, Lori Braxton, Teachers, Greene County Middle School, North Carolina
Angelo Collins, Associate Professor of Science Education, Florida State University
Rita Elliot, Teacher, A.G. Cox Middle School, Winterville, North Carolina
Paula Egleson, Graduate Associate, SERVE
Joe Follman, Editor, SERVE
Barbara Foster, Associate Director, SERVE
Terri Hardin, Teacher, A.G. Cox Middle School, Winterville, North Carolina
Glenda Johnson, Publications Design Specialist, SERVE
Susan Henry, Pupil Assessment Specialist, Charlotte-Mecklenburg Schools, North Carolina
Helen McCormick, Mary Miller, Teachers, Conway Middle School, Conway, South Carolina
Bruce Middleton, Teacher, Eastern Alamance High School, Mebane, North Carolina
Dianne Meiggs, Environmental Ed. Coordinator, Elizabeth City-Pasquotank Schools, North Carolina
Jerry Natkin, Evaluation Program Manager, SERVE
Larry Rainey, Project Director, Integrated Science, The University of Alabama
Joe Ryan, Center for Excellence in the Assessment of Student Learning, University of South Carolina
Dennis Silas, Teacher, Drew High School, Drew, Mississippi
Nancy Smith, Teacher, Graham Middle School, Graham, North Carolina
Bill Spooner, North Carolina Department of Public Instruction
Helen Weaver, North Carolina Project for Reform in Science Education
Wayne Welch, Professor of Educational Psychology, University of Minnesota

CHAPTER 1
CURRENT VIEWS ON ASSESSMENT

Introduction

Educational systems promote student growth in a variety of dimensions. Traditional dimensions have been the basics (reading, writing, basic math computations, and basic content knowledge). Basic content knowledge can be effectively assessed with multiple-choice and completion tests. However, with "educational reform" has come more concern 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 measures may not serve these kinds of goals as well.

The purpose of this manual is to encourage discussions among science teachers about desired student outcomes in science and assessments appropriate to those outcomes. Multiple choice tests have their place, but so do portfolios, observations, performance tests, student interviews, and exhibitions. Once teachers are clear on what they hope to accomplish with their students, choosing appropriate assessment methods follows.

According to one of the authentic assessment "experts:" "I think what's going on is something more radical than rethinking testing. What we're really doing is rethinking our purposes." (Wiggins, 1992, p. 37)

According to Wiggins (1992), the true issue being debated is not whether some assessment methods are superior to others, but rather what is worth assessing. The debate about assessment is a "value" debate. What goals or outcomes do we value for students and how do we best assess their progress toward these ends?

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.

Purpose of this manual

This publication is designed to help you do the following:

1. Consider the variety of possible student outcomes in science and select those that are most important for students.
2. Reflect on and choose appropriate ways to assess student performance on important outcomes. There are three major ways teachers have of assessing how students have changed relative to the goals of instruction: they can observe students, ask students questions and look at their work. The assessment methods discussed in this manual are shown at left.
3. 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)
4. Reflect on grading practices and how information from a variety of assessment methods might be incorporated into a composite picture of achievement.
5. Consider ways to get you and your school started in changing practices.

observe students using

- informal observations
- structured observations
- narratives

ask students questions using

- interviews
- self-assessment

evaluate their work using

- open-ended questions
- performance tests
- journals
- exhibitions/culminating demonstrations
- portfolios

This publication is not intended as a text but as a self-study resource. That is, it is not something to be read and then shelved. Rather, 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 would result.

Any suggestions you have for how we might improve the manual would also be appreciated. Call Wendy McColskey at the SERVE office in Greensboro, North Carolina, with any suggestions or questions (1-800-755-3277). SERVE is currently working with other regional labs in the country to collect samples of non-traditional student assessments being used by teachers in the classroom and by large-scale testing programs. Thus, the laboratory may be able to put you in touch with other schools that have developed assessments in areas of interest to you.

Self-Assessment

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?

   I. ____________________

   II. ____________________

   III. ____________________

   IV. ____________________

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

   _ Listen to lectures
   _ Take tests on text/lectures
   _ Read textbooks
   _ Take end-of-chapter tests
   _ Design experiments
   _ Talked with scientists
   _ Solved problems in a team setting
   _ Maintained journals of data collected
   _ Did hands-on investigations
   _ Made 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 gradebook. What might students infer about the purpose of science instruction from your choices about what is graded?

What is "good" assessment?

> **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 bring no resource materials with them to the test. Students' tests are scored and returned and form the basis of the six weeks' grade.

Traditionally, the goal of most subject area courses has been for students to be able to recognize or recall important facts, concepts or relationships that have been explained to them. Frequently used assessment methods (multiple-choice, matching, short answer) measure progress toward this goal.

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 retained. When a teacher emphasizes factual knowledge on a test, 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.

Proponents of assessment reform argue that past assessment practices are inadequate. Fundamental problems with previous testing practices include a) narrowness of scope; b) limited expectations of students; c) overemphasis on memorizing facts; d) lack of student ownership in the learning process; and e) lack of incentives for student improvement in their work. In other words, both the "what" and the "how" of student assessment need to be changed.

According to Wiggins (1989), "authentic assessment" means that

- 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.

The outcome-based education reform movement (Spady, 1988) is essentially an assessment reform movement. In designing programs, this movement suggests that we should define exit outcomes first, and then fit content and instructional methods to support the desired outcomes. Spady and Marshall (High Success Network, 1992) argue that 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, investigations, major projects and creative depictions
- 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
- allow for student creativity in their products or responses
- allow students to be involved in creating the criteria against which their performance will be judged
- include audiences in addition to the teacher to validate and judge student performances (e.g., scientists, other students)

The example on page 6 illustrates the above points.

Sample Assessment

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 checklist.

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.

In summary, these assessment movements argue strongly that the work we have asked students to do in the past (drills, worksheets, fact-level questions and answers, multiple-choice and short-answer tests) does not challenge and involve them, does not encourage creative, quality work, and does not provide them with experience in using and applying information and skills in a "real-world" way.

What is a "real-world" task? A few examples of generic kinds of tasks that have students *using* 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 about 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.

A key point to remember as you go through this manual is that assessing students involves gathering information about what they know, can do, or are like. If throughout twelve years of school, students are assessed only on passive, non-creative work (work-sheets, multiple-choice tests), how likely is it that they will become problem solvers, creative producers, effective communicators, and self-directed learners as adults?

Educational objectives (curriculum goals), the design of learning experiences (instruction), and student evaluation or assessment are considered the three legs of the educational process. Chapter II deals with identifying science education objectives. Chapters III and IV deal with student assessment methods. Additional information on science curriculum and instruction reform can be found in an appendix.

Chapter II

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? Assessments communicate to students what is important.

Key point: The assessments you use must reflect your purposes, and the purposes of your instruction must be made known to the students.

Consider these statements from a recent publication by the National Center for Improving Science Education:

> "If only new vocabulary is tested, there is an implicit message that science is mostly a matter of memorizing new terms."
>
> "If only factual knowledge is tested, the message may be that science is a static body of facts, principles, and procedures to be mastered and recalled on demand."
>
> "If tests call for the students to engage in active exploration and reflection, to pose new questions and solve new problems, the message can be that science is a mode of disciplined inquiry, applied specialized knowledge, investigative procedures, and rules of evidence for understanding both the natural world and the technologies through which humans have shaped that world to their ends." (Loucks-Horsley, 1990)

You might have different levels of goals or purposes that guide your work with students. For example, the Outcome-Based Education literature (High Success Network, 1992) describes outcomes for students that cut across all levels and courses. Possible outcomes are that students will become:

- Self-directed learners
- Quality producers and performers
- Goal-setters and pursuers
- Effective communicators
- Collaborative community contributors
- Creative and complex thinkers
- Innovative problem solvers

Such a list gives some structure and direction to assessment across all content and subject areas. Teachers can design work and experiences for students with these kinds of outcomes for the student in mind. However, teachers also have science goals or outcomes for students as shown in the following section. The science goals teachers choose for their students give more immediate direction for what they are trying to accomplish with their science instruction.

The National Assessment of Educational Progress develops and administers science tests to a national sample of students on a regular basis to provide a picture of student capabilities in science. These tests are based on a conceptualization of the important goals in science.

The following information is taken from the National Assessment Governing Board's *Science Framework for the 1994 National Assessment of Educational Progress* (NAEP Science Consensus Project, 1992). The 1994 NAEP assessment includes several categories of student outcomes:

1) **conceptual understanding**, with stress on the connections among disciplines and students' organization of factual knowledge
2) **scientific investigations**, meaning acquiring new information, planning appropriate investigations, using a variety of scientific tools, and communicating the results of the investigations
3) **practical reasoning**, meaning analyzing a problem, planning appropriate approaches, and evaluating results
4) **nature of science/technology**, meaning knowledge of the nature of science (i.e., that scientists invent explanations to fit observations but that these explanations may change with new evidence); and an understanding of how technologies are designed and developed

Articulating Outcomes

At this point, please take a few minutes to reflect on what you feel are important objectives for science instruction. It would be helpful to discuss your responses with one or more science teachers.

1. How would you rank order the NAEP goals of conceptual understanding, scientific investigation, practical reasoning, knowledge of the nature of science, and an understanding of how technologies develop for the level of science you teach?

2. Review the list of sample science outcomes shown on the next page. The list is intended only as a starting point for brainstorming. Reword, reorganize, and add to the list as needed. Then, think about the four most important science outcomes that you have in mind for students and enter those below under "List of Most Important Student Outcomes." You will likely revise these often before you are happy with them, but this is the information that is critical to discuss with other teachers, with students, and with parents so that everyone knows what you are aiming for with your instruction.

List of Most Important Student Outcomes in Science

In the space below, enter what you believe to be the most important outcomes for your students.

1. Outcome:

2. Outcome:

3. Outcome:

4. Outcome:

Examples of Science Education Outcomes for Students

Conceptual understanding

Students will be able to do the following:

1. organize and express in their own words important science ideas
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

Students will be able to do the following:

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

Students will be able to do the following:

1. work successfully through a complex problem with a group of other students

Nature of science/technology

Students will be able to do the following:

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. explore the advantages and disadvantages involved in the design and development of technologies
6. summarize examples of how scientific knowledge has been applied to the design of technologies

The sample outcomes listed above are statements about the intended results of science instruction. They contain action verbs (process objectives) that have students performing or behaving in some way (demonstrate, summarize, explore). A syllabus that lists topics to be covered by a course is an implicit statement that the objective of the course is to "know" some things about the areas listed. However, just knowing things is no longer an adequate goal for students. It is important to communicate to students what they are expected to be able to do with the content.

One tool that helps teachers sharpen their thinking about this mix of content and process objectives is a matrix such as the one shown on the following page. (It is similar in concept to a test specifications matrix--a tool used by test developers to ensure that test items are written in numbers proportionate to each cell's importance.) The behaviors or outcomes expected of students are listed on the side. The content areas considered essential for students to understand are listed across the top. A course objectives matrix is a way to identify and communicate critical course outcomes to students. A publication by the National Science Teachers Association (see appendix) lays out a suggested scope and sequence of science content in grades six through twelve that might be helpful in thinking through the essential content at your grade level.

In the matrix shown, the use of such verbs as "explain" and "apply" (#4 & #5) specify the level of content understanding to be gained. (Other examples of thinking verbs are synthesize, categorize, identify errors, analyze, summarize and compare and contrast.) Science process skills (#1, #2, & #3) are also listed as course objectives. The sixth objective ("work cooperatively to investigate problems") means that students will be expected to learn to problem-solve in a team setting. By maintaining a journal and portfolio (#7), students will hopefully become competent at reflecting on their learning and evaluating the quality of their work.

Before you go on to the next chapter, you may want to spend some time experimenting with a matrix that describes your science course objectives, perhaps building on the four "most important" science outcomes previously identified.

Sample Course Objectives Matrix (Sixth Grade)

Objectives	Essential Course Content Areas Studied				
Students will:	**Sun's Apparent Motion**	**Direction**	**Soil**	**Water Cycle**	**Biomes**
1. Use simple instruments and collect measurements					
2. Make observations and predictions					
3. Collect, record & interpret data					
4. Explain relationships, concepts, principles, theories, etc.					
5. Apply information to new situations.					
6. Work cooperatively to investigate problems.					
7. Maintain a journal and portfolio.					

Note: Matrix was adapted from those developed as a part of the North Carolina Project for Reform in Science Education, College of Education, East Carolina University, Greenville, NC.

Chapter III

STUDENT ASSESSMENT METHODS: WHAT ARE METHODS OF ASSESSING STUDENT PERFORMANCE THAT GO BEYOND MULTIPLE-CHOICE TESTS?

Although many people equate assessment with testing, in reality assessment involves far more than merely giving tests (Hopkins, Stanley, & Hopkins, 1990). That distinction is critical to understanding the array of methods presented in this chapter. To assess students is to collect information about their performance. It is an integral part of the teaching and learning process. The goal of education is to produce or facilitate changes in learners. How does one know if students are becoming competent and knowledgeable?

Numbers are often useful in assessment because they can be communicated and interpreted efficiently, but not all useful information about how and what students are learning is quantitative. In fact, many argue that there has been an over-dependence on quantitative test information and an underutilization of more subjective methods involving teacher observation and judgment. Grading, promoting, and placement into special education or remedial instruction are important educational decisions that should be based on a variety of assessment methods. "Let us not fall into the trap of asking whether we should use teacher judgments or test scores. Faced by complex problems of assessment and evaluation of student growth and the factors that influence it, we cannot reject any promising resource. Various sorts of information complement each other" (Hopkins, Stanley, & Hopkins, 1990, p. 8).

This section walks you through ways to gather information about students. For example, a teacher might be interested in what students understand about mapping and direction. Multiple choice items might ask:

1. What direction are you traveling when you drive from New York to California?

   a) east b) north c) west d) south

2. The sun sets in the ________.

   a) east b) north c) west d) south

A performance-based assessment, by contrast, would ask students to construct a map of their neighborhood or find their way from one part of the school to another using a compass. Performance-based assessments provide students with an opportunity to demonstrate what they know, rather than just recognizing or recalling the right answer.

Teachers assess students' performance by observing them at work, asking them questions, and/or reviewing their work. In this section, each method of assessment listed below will be described and examples provided.

Methods for Assessing Students—Performance-Based Assessments

1. **Observing Students Using**
   - informal observations
   - structured observations
   - narratives
2. **Asking Students Using**
   - interviews
   - self-assessments
3. **Looking at Students' Work Using**
   - open-ended questions
   - performance tests/tasks
   - journals
   - exhibitions/projects/culminating demonstrations
   - portfolios

When a teacher decides that a performance-based assessment is the most appropriate way to assess student progress relative to an instructional goal, a decision must also be made as to how the student response will be judged. The term, "rubric," refers to guidelines for judging responses. Rubrics and grading are discussed in Chapter IV.

OBSERVING STUDENTS

Informal Observations

Teachers are constantly observing students and making assessments about their performance that in turn influence future instruction. These are informal observations. With informal observations, teachers 1) are observing with no predetermined focus, and 2) the choices of whom to observe are not formalized. Through such observations, teachers might, for example, become aware of students in their classes who are able to work independently and those who require a great deal of guidance. Informal assessments are often the basis for information provided to parents (e.g., Kevin seems to enjoy learning; Collin works well independently).

Structured Observations

Occasionally, more formal observations are needed. Formal observations have 1) a specified focus and 2) sample behavior to be observed systematically. Some goals or objectives can only be assessed by observation. 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. Although informal observations occur daily, occasionally teachers may want to record information from their observations on a form. The form might be structured as a matrix of those being observed by the behaviors of interest.

For example, if a school has recently set up a science activity lab where students can individually engage in hands-on science activities, teachers may want to evaluate students' progress by seeing how the students increase their ability to stay on task and work independently with the materials in the lab. Teachers might develop a structured observation form for the lab as shown in Figure 1 on page 17.

Figure 1

Science Activity Observation Form

Teacher: ________________ Date Observed: ____________

Time observations began: ________________ Time observations were completed: ____________

Hands-on science activity (describe briefly): ____

____

Directions: List the students to be observed in the spaces provided. Plan to observe those students during a 10-15 minute, individual hands-on science activity. Observe each student and put a check in the box if he or she is on task. After you have observed each student on your list once, go back starting with student No. 1 and make a second round of observations, and continue until the activity ends. After the observation ends, check either "none," "some," or "much" under the level of assistance each student needed to complete the activity.

Student's Name	Working to complete assignment; on task	Assistance needed: None	Some	Much
1.				
2.				
3.				
4.				
5.				
6.				
7.				
8.				
9.				
10.				

The information collected by the Science Activity Observation Form could be used in a number of ways. Formal observations often allow teachers to profit from new information that may challenge some of their opinions about students. For example, before looking at the data a teacher might have guessed that Mai would have difficulty staying on task. After the September observations, the teacher realized that Mai functioned well independently; Alice, Juanita, and George needed assistance to use the materials appropriately.

Figure 2

Summary Science Activity Observation Form

Teacher: ROSE

Observations Dates: 1. Sept. 12 2. Jan. 23 3. May 5

Student's Name	Number of times observed working to complete assignment		
	Sept.	Jan.	May
1. Alice	1	2	3
2. Mai	4	5	4
3. Juanita	2	3	4
4. Michael	4	4	5
5. George	2	4	5
Total	13	18	21

Data collected over time can be useful for showing gains. Figure 2 shows a general pattern of improvement over time on independent lab work. This information may be used to demonstrate the benefits of the lab. It may also be used to show which and how many children need improvement. Finally, teachers might use it as a way to give students feedback about how they are improving.

Narratives

Progress on some instructional objectives can be tracked best through narrative records of observed behavior. A narrative is a written record. Such narratives are particularly appropriate for complex behaviors, such as group interactions, which cannot be described effectively with a checklist. For example, a teacher might observe and describe a cooperative team learning activity. The observation would begin with a purpose such as to see how students on a team contributed to completing an activity. An example of such a narrative appears in Figure 3.

The teacher would begin the assignment, explaining what the students needed to do. In this example, students have been asked to work in pre-assigned groups of four to explore how they might measure wind. Rather than facilitating the activity, the teacher has informed the class that she will only be observing.

Figure 3

Narrative Observation: Group Problem Solving

Observer: VIOLET **Time:** 1:20 (begin) to 1:30 (end) **Date:** Sept. 12

Purpose(s) 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.

Narrative:

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 of the activity sheet. 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 activity sheet 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; 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.

The narrative in Figure 3 is the teacher's written record of how the children worked together to solve the problem. Overall, it provides evidence about how each of the students contributed to the problem-solving activity and about the individual styles of the students. (e.g., Crystal appears to be a natural leader and very task oriented.) The teacher could use this information to set goals for individual students or structure group membership in the future. Future observations could be focused on those groups experiencing difficulties. The results could be provided to these groups with comments on how to improve their functioning. Over time, a series of these narratives might demonstrate how students changed the way they worked as a team.

ASKING STUDENTS

In the last section on observing students, several examples were provided of how teachers collect information about student behavior. A second method of collecting information about students involves the analysis of replies that students give in interviews and on self-report questionnaires.

Interviews involve face-to-face verbal exchanges between the teacher and student. In self-report questionnaires, students respond to written questions and statements. The focus of the interviews or self-assessment questionnaires may be on a cognitive event such as what students understand about a particular topic, how they feel (e.g., what do they like and dislike about working in groups?) or how they report they behave (e.g., do they talk about science topics at home or read science books in their free time?).

<u>Interviews</u>

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

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

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.

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

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 questions verbally and may need practice.

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.

6. One of the goals mentioned in Chapter II is that students will be able to communicate effectively as a result of their K-12 experience. If science teachers adopt 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.

Daily teacher/student dialogue which occurs during instruction can be seen as a series of unstructured interviews used by teachers to assess students' competence relative to instructional objectives.

Interviews can vary in their degree of structure. In unstructured interviews, the contents and order of the questions vary with the student and are responsive to each student's answers. 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. In structured interviews, teachers ask students to respond to the same set of questions.

Using information from interviews. The way that information from interviews is used depends on the context or purpose of the interview. Some examples follow.

1. If a new unit is being developed and the teacher is interviewing a small sample of students on their ability to explain and relate

Some suggestions for interviews are:

1. Use samples of the class for interviews when you can, rather than all students.
2. Keep the tone of the interviews positive and constructive. Try not to give verbal cues or facial expressions that can be interpreted as meaning that an answer is silly or that an error has been made.
3. Let students respond without interruptions and give them time to think before they respond.
4. Try to keep interviews short, and focus on important questions.

the concepts of adaptation and natural selection, tape recording the interviews might be helpful. The teacher could listen to the tape at a later time and look for misconceptions in student responses.

2. If the purpose of the interview is to assess students' progress on an objective having to do with accurately communicating scientific principles, a series of rating scales could be developed to describe poor, average, and good performance on a variety of dimensions (e.g., organization, coherence, and completeness of responses). Students could be asked to rate their own performance using these rating scales and then the teacher might share his or her ratings with the student.
3. Interview responses can also be graded as an oral test. Structured interviews may give those who have poor literacy skills a chance to succeed. In addition, this assessment method provides the teacher with assurance that students have understood the questions. (Written exams assume students understand the questions asked.)

Some teachers also report that students take oral tests more seriously because they are more personal expressions of competence than a written test. Students may prepare more carefully if they know they must stand before a teacher and be asked questions individually..

An example of a set of structured interview questions and a grading rubric are shown in Figure 4. The same set of questions is asked of all students. Student responses to the questions are recorded by the teacher and point values are entered on the form as the student responds to each question.

The results could be used in a number of ways. Students who had less than 17 points could be assigned a peer who had received 17 points, and the peer could work on the questions with the student until he or she was ready to retake the exam. The second administration could result in a score entered in the gradebook.

Figure 4

Oral Exam: The Three Phases of Water

Student's name:______________________________ Date: ____________________

Scoring key	Points	
1. 1 point for each phase identified correctly (ice, water, steam)	____	1. What are the three phases of water?
2a. through 3d. 0 = Incorrect 1 = Partially correct 2 = Satisfactory		2. Describe each of the three phases.
	____	a. Ice:
	____	b. Liquid:
	____	c. Steam:
		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:
	____	4. Is there anything you do not understand about water phases?
Enter Total Points (max. 17)	____	

Self-Assessment Questionnaires

Each assessment tool has its advantages and disadvantages and serves some purposes better than others. Student self-assessment questionnaires might be helpful in determining how students perceive their knowledge, skills, or the quality of their work, the concerns they have about their progress, their prior level of experience with a topic or skill, their feelings about the class, or their interest in science as a career.

When used appropriately, self-assessments actively involve students in reflecting on their learning process and emphasize the importance of students' awareness about what they know and what they need to know. Figure 5 presents a science self-assessment questionnaire 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 four (Figure 5), so that students who knew how to use the instrument, but had forgotten the name, 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 the end-of-year level of interest. Summarizing the responses to the instrument item (Question #4) in a frequency chart (instruments by number of students who had used each and could describe 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 notes kept about any difficulties they had as teamwork was initiated.

Figure 5

Science Skills Self-Assessment

Directions: Read the questions and statements below and answer as best 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 the 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?

Students can be asked to evaluate their understanding of concepts at any point in the instructional process. 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 following key:

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.

Such an approach to assessing students' understanding is less threatening than a pre-test and can give students a sense of the different levels of knowing, if used frequently in a class situation. Results of student ratings of each concept could be tabulated as a class activity.

LOOKING AT STUDENTS' WORK

An overview

Multiple-choice and completion items are not included in this manual as there are many sources that can be tapped for guidance on how to write these kinds of items (Gronlund & Linn, 1990; Hopkins, Stanley & Hopkins, 1990; Payne, 1992). Rather, our emphasis is on open-ended questions, performance tasks, journals, exhibitions, and portfolios. The development and administration of test questions, items, or tasks starts with finding the most appropriate assessment option, given the relevant instructional objectives. Each method has its place.

Keep in mind in constructing tests that

- course objectives should guide the decisions about what is assessed;
- the mix of content (what topics) and process (what kind or level of performance) must be considered;
- item development takes time and is not easy;
- multiple-choice, completion, open-ended, and performance items are appropriate for different purposes.

Completion items (e.g., The first man on the moon was ________) are direct, minimize the effect of guessing, and are good ways of assessing factual knowledge. However, they are limited in their ability to tap students in-depth understanding or reasoning about a problem.

Multiple-choice items can be used flexibly to tap into a variety of complex problem situations and can ask students questions using all levels of thinking (application, analysis, etc.). They can be scored reliably and administered efficiently to cover a wide range of material. However, they are difficult to write well and they do not require students to construct or frame a response.

Open-ended and performance task items fit well with the current emphasis on developing students into effective communicators, quality producers, and active, complex thinkers because they call for students to perform or construct a response.

Open-ended questions

Rather than having students select a response, open-ended questions ask students to produce a response. This approach is appropriate for assessing student performance on more complex cognitive outcomes. The length of the response could vary considerably given the age of the student, the question(s) asked, and the time provided to complete the question(s).

Some examples of generic starting points for open-ended questions that relate to higher-order, cognitive objectives are provided below (Gronlund & Linn, 1990):

- Explain a cause-effect relationship. Describe an application of a principle. Formulate a hypothesis or a conclusion. Describe the limitations of the data. Explain a method or procedure. Integrate learnings in different areas. Create or design something (i.e., an experiment). Evaluate the worth of an idea.

Some examples of open-ended questions that address students' skill at applying information and making judgments follow.

> - "Would you weigh more or less on the moon? On the sun? Explain." (Assesses the ability to apply a rule or principle in a specified situation.)
> - "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?" (Assesses understanding of cause and effect.)
> - "Is nature or nurture more influential in determining human behavior? Why?" (Assesses skill at supporting decisions for or against.)
> - "What questions should a scientist ask in order to determine why more smokers than non-smokers develop lung cancer?" (Assesses skill at formulating new questions.) (Payne, 1992, p. 174)

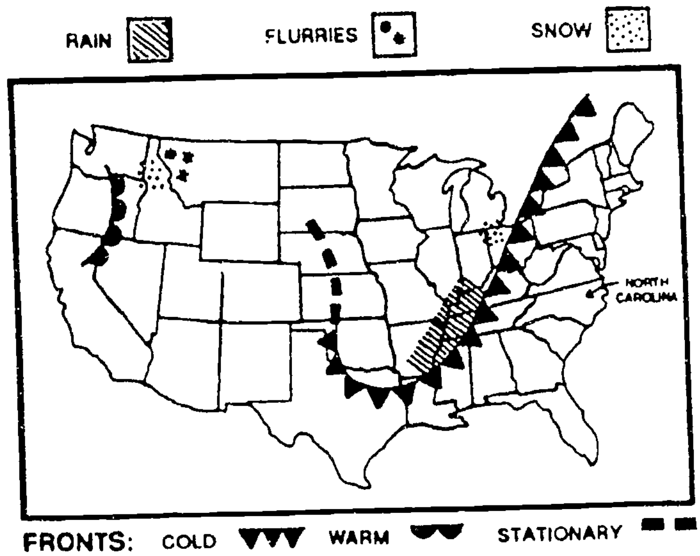
It should be noted that if an application has been taught explicitly and the answer is provided from memory by the student, rather than representing the presentation of a new situation, the objective assessed is factual recall, not application.

Open-ended questions can assess a variety of instructional goals such as conceptual understanding, application of knowledge via creative writing, the use of science process skills, and divergent thinking skills. Some examples are shown on page 28.

Conceptual understanding:

1) 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 a possible. (Grade 8)

2) Using the weather map shown below, make a forecast for the weather in North Carolina for the next day. Explain why you made the forecast. (Grade 4)



Source: Open-response released items (1991-92). Kentucky Instructional Results Information System. Kentucky Department of Education. Division of Curriculum, Assessment, and Accountability, Capital Plaza Tower, Frankfort, KY 40601.

Application and creative writing:

1) If the earth stopped rotating on its axis, how might our weather change? Be imaginative, speculative, and specific.
2) You are returning from your interstellar journey where you were in charge of soil analysis. It is your responsibility to write up a description of the biome(s) you found on your planet. Include information about the living and non-living components of the biome. Be sure to note food webs, food chains, altitude, rainfall, soil types, latitude and any other important information.

Source: Robin Freedman (October, 1992). Writing, student portfolios, and authentic assessment. In *The Watershed*, a quarterly publication by the California Assessment Program. California Department of Education, P.O. Box 944272, Sacramento, CA 94244-2720.

Science process skills:

1) Katie believes that students who do between 4 and 10 hours of homework per week make better grades than students who do not do homework or who do more than 10 hours of homework per week. To test this hypothesis, she is writing a survey that she will give to students at her school. (Grade 8)
- What questions should Katie include in her survey?
- Describe the scientific procedure Katie should use.
- Describe what Katie should do with the responses to her survey to find if her hypothesis is correct.

2) You are a state scientist. You are 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. 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.

3) The number of kernels that do not pop during microwave cooking varies with different brands of microwave popcorn. Design an investigation to determine which of three brands of microwave popcorn pops best. Be certain to describe each of the following: (Grade 12)
- everything you will do in your investigation;
- the kinds of data you will collect; and
- how you will analyze the data to determine which brand of popcorn pops best.

Source: Open-response released items (1991-92). Kentucky Instructional Results Information System. Kentucky Department of Education. Division of Curriculum, Assessment, and Accountability, Capital Plaza Tower, Frankfort, KY 40601.

Divergent thinking:

Creativity or divergent thinking is an important part of scientific thinking. For example, the generation of research questions and hypotheses and the development of plans of actions require an ability to posit multiple, original approaches. Tasks that engage students' creativity have no right answer. Open-ended questions can be developed that ask for multiple responses.

Students might be given sample situations that relate to the unit under study such as:

- "Suppose we lived in a world without insects"
- "Suppose there was no more disease in the world"

Students might then be asked to 1) write as many questions as they can that will help them understand the situation; 2) list as many possible causes as they can for the situation; and 3) list as many possibilities as they can for what might happen in the future as a result of what is taking place in the situation. The number and quality of the responses are indicators of creativity. The quality of each response might be assessed as I(irrelevant), P(pertinent), and U(unique).

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, IA 52242.

Making sense of students' responses. If open-ended questions are to be included on a test that will be graded, it is important for teachers to prepare students for the expectations held for them by communicating how their responses will be judged. After many years of multiple-choice testing, some students may have difficulty with open-ended questions. Their responses may be short, somewhat incoherent, and not well-developed. It may be difficult to judge their understanding of the concept because of weak communication skills.

At first, students may need more than one chance at expressing their understanding of essential concepts. Perhaps on the first administration of open-ended questions on a topic, the teacher could pick the best student responses and ask the class to critique other student responses in terms of whether or not they met this standard. No grades would be given until the second or third administration or until it was clear that students had ample opportunities to understand the quality of responses expected of them.

Grading open-ended questions involves interpreting the quality of the response in terms of some criterion. In the example shown in Figure 6, the criteria are the scientific accuracy of the explanation or description provided and the coherence of the response. Both criteria are included in a single scale. Distinguishing between a 2 (accurate but not well-written) and 3 (accurate and well-written) may help to impress upon students the importance of structuring their responses so that they are coherent to a reader.

Several suggestions for rating open-ended questions are offered.

1. Articulate the outcomes that will be assessed by open-ended questions. For example, the instructional objective assessed by the questions in Figure 6 might be "students will be able to explain phenomena relevant to the earth/sun system."
2. As part of choosing or developing questions to administer, answer the questions yourself to better clarify your expectations regarding an ideal student response. Determine in advance the elements you expect in a complete answer.
3. Develop a rating scale or point system to use with the questions. More information about rating scales and checklists is found in Chapter IV.
4. Read over a sampling of answers before grading them and get some idea of the range of responses to each question. It may be helpful to sort the responses to a question into piles based on the rating scale being used before assigning a final scale value to the response.

Another use of students' responses to open-ended questions is to analyze their responses for misconceptions or problems they are having in understanding a concept. Rather than grading responses, responses can be grouped into categories of similar kinds of answers so that future instruction can respond to the kinds of errors being made.

Figure 6

Open-Ended Questions about the Apparent Motion of the Sun

RUBRIC

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
Total possible points: 12

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 a.m., a shadow is west of a tree; at 4 p.m. it is east of the tree. Explain why this happens.
4. Why do people in North Carolina see the sunrise before people in California?

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

Performance tests/tasks

Although many achievement objectives can be assessed with paper-and-pencil tests, there are other objectives which are more appropriately assessed by having students actually demonstrate their competence. In some situations, given the purpose of the assessment (e.g., licensing people to drive cars), a performance test is a necessity. That is, it would be unthinkable 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 for example) which are most appropriately assessed by having the student perform tasks rather than take a paper-and-pencil test.

Although science teachers may make extensive use of hands-on and lab activities for instruction, they may not make effective use of performance tasks for assessment. If students are expected to be competent at measurement using different instruments or at using lab equipment such as microscopes, a performance test makes more sense than a multiple-choice test. For example, middle school students might be given the following materials:

Equipment:
100 ml graduated cylinder (clear plastic)
balance (and mass units if necessary)
tray or spill pan
calculator
liter of water
metric ruler

Materials:
glass marble
piece of wood dowel
small piece of plastic rod
aluminum rivet
a lead weight
5 pennies (use all 5 at one time)

Students would be instructed to use the equipment and materials to make the necessary measurements to calculate the density of each material. The questions shown in Figure 7 are samples taken from the performance task section of the Kentucky Instructional Results Information System.

Conceptual understanding can also be assessed with performance tasks. Paper-and-pencil tests may be used effectively to assess conceptual understanding of a topic (such as electricity), but students might be more engaged by and learn more from a performance assessment. For example, suppose a teacher has been working for the past month on electricity. Students have talked and read about electricity, performed experiments, and kept a journal record of what they have learned. The teacher decides that the end-of-unit test should be a graded performance task. An example of such a task (taken from the California statewide testing program field test) is shown in Figure 8. If science classes are to be about doing science rather than just reading about science, then the use of performance tasks represents a better match to instructional objectives.

If the teacher was grading this performance task, he or she might choose to use a rating scale such as shown in Figure 8. It is important to note that scoring guidelines such as this can not be developed in the abstract. It takes careful analysis of many student responses to the test to derive descriptive categories that accurately capture their performance levels.

Once the scoring guidelines are developed, rather than scoring each question separately, the teacher looks at the pattern of responses across the questions and assigns a number or letter grade based on the best match to the descriptive categories. A complete/incomplete scoring guideline would reduce the descriptive categories down to two; one which described acceptable performance across the questions and one for unacceptable levels of performance.

Depending on the purpose of the assessment, there are many different ways to judge how well students performed on the task. However, it is critical to be clear on the elements or features of a desired, strong response.

Figure 7

Performance Test – Density of Solids (Grade 8)

1. Using the equipment and materials provided, make the necessary measurements to calculate the density of each material. Be sure to record and label the unit of measurement used for each density. Use the chart below to record your information.

Density = $\frac{\text{Mass}}{\text{Volume}}$

Test Item	Mass	Volume	Density (and unit of measurement)
Glass Marble			
Wood Dowel			
Piece of Plastic Rod			
Aluminum Rivet			
Lead Weight			
5 Pennies (use all 5 at one time)			

2. Rank order your six materials from LEAST DENSE (1) to MOST DENSE (6) in the chart below.

	Material
1.	(Least Dense)
2.	
3.	
4.	
5.	
6.	(Most Dense)

3. Display the density of the materials in a graph which shows mass and volume.

4. Describe how you determined the volume of the lead weight and wooden dowel.

5. Based on what you have learned about the density of these six materials, which would you use to build a boat? Explain why.

Source: Kentucky Instructional Results Information System, Kentucky Department of Education, Division of Curriculum Assessment and Accountability, Capital Plaza Tower, Frankfort, KY 40601.

Figure 8

Performance Test – Electricity (Grade 6)

Test questions to students: You are a scientist working for a large computer company. Your assignment is to investigate electricity.

1. Make an electrical circuit using all the items on the table (battery, wire, light bulb, switch).

2. Quickly draw a simple picture of your circuit in the space below.

3. Did you build a complete circuit?
   Yes_____ No______

4. Explain how you know.

5. Open "BAG A." Use the clip and lead to make an electrical tester. Test each of the items in "BAG A" with your circuit. Place an X on the chart under the appropriate column to show what happened when each item was tested.

"BAG A" Items	Conducts Electricity	Does Not Conduct Electricity
Plastic Spoon		
Steel Washer		
String		
Penny		
Nail		
Rubber Band		

6. How are the items that do conduct electricity alike?

7. How are the items that do not conduct electricity alike?

8. Examine the item in "BAG B." Do you think it will conduct electricity? Why or why not?

Draft Scoring Rubric

The following draft rubric was developed to assist in scoring student responses to the Grade 6 Performance Field Test in science.

4 = Gives complete and acceptable answers to all questions; provides acceptable rational; includes a complete and accurate diagram of a circuit with supporting evidence; demonstrates understanding of the concept of electricity and conductivity; may use descriptive terms (conductor, flow, current, etc.).

3 = Gives fairly complete and acceptable answers to most questions; provides good answers, but rationale may be vague; includes a complete diagram of a circuit; shows understanding of the concept of electricity and conductivity; responds to questions #4 or #8 in an acceptable manner.

2 = Several incomplete or unsatisfactory answers; rationale is very limited; shows some understanding of the concept of electricity but not conductivity; diagram of a circuit may be missing or incomplete.

1 = Very little response (diagram only or few answers); partial answers to a small number of questions; no rationale; does not include a diagram of a circuit; contains at least one correct answer other than question #3.

Source: California Assessment Program. Science Performance Field Test, Grade 6, 1990. California State Department of Education, Sacramento, CA 95814.

Journals

Open-ended questions and performance tasks are ways to assess student learning at a particular point in the instructional process. Journals are dynamic assessment approaches that promote communication between the teacher and student, allow students to reflect on what they are learning, and foster students' active involvement in classroom activities.

Consider a middle school science classroom where there is no formal textbook used. Rather, science instruction consists of daily hands-on activities and teacher-student dialogue about a question (e.g., "How do soil types differ?"). Students might use journals to process information from hands-on investigations.

First, students might be asked to write a response to a question that addresses their current understanding about the question under study. Then, they might write their prediction about what might happen in a particular situation; enter data from observations; develop conclusions; and reflect on the main idea of the activity. These daily journal entries become a record of their class experiences. Teachers refer students back to journal pages that contain their work on certain topics in reviewing for tests and in making connections between topics. (The program described is the North Carolina Project for Reform in Science Education, a National Science Teachers Association Scope, Sequence, and Coordination project.)

Journals could also be used to assess attitudes toward science, if positive attitudes were an instructional goal. Students could write their thoughts or feelings about class events. Journal assignments might be structured for younger students, with more choice options added for older students. 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.

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 a teacher believes that students need to structure, take charge of, or feel ownership of 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 a completed notebook of journal entries. In the middle school program just described, teachers weight the journal as one-fourth of the grade for a grading period.

Exhibitions/Projects/Culminating Demonstrations

Complex, "real-life" tasks or challenges are a type of assessment commonly called "authentic assessment." In outcome-based education they are called "culminating demonstrations" to reinforce that they are assessments that bring a unit together. "Culminating demonstrations" are tasks that are worked on over a 6-week period, or in the case of graduation exhibitions, over a year-long period. They have multiple steps (plan, research, design, implement, etc.) and multiple criteria can be used to judge them. Such complex, "real-life" tasks are experiences that develop students into problem-solvers, effective thinkers, quality producers, and self-directed learners. 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. Students may be asked to play a "real-life" role as in Figure 9, where they are charged with acting as an investigative reporter.

Figure 9

Sample Authentic Assessment

You are an investigative reporter researching how a country's culture and history have contributed to views regarding certain endangered species. Your team's job is to decide what history tells you about the development of your country's views about the whale. (Each team has been assigned a different country.) You are to write a newspaper article reporting your information.

1. Identify all the possible historical influences you want to investigate.

2. In your team, identify both pro and con views and look at both practical and environmental protection issues.

3. Prepare your article in such a way that it reports both information and pro and con points of view.

4. Present your draft to the editors (another team) before you send it to the copy editor and print shop.

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

Authentic assessments are similar in concept to a senior-level college research project, master's thesis, or dissertation. A dissertation, for example, is an assessment in the sense that a committee must agree that the work is of acceptable quality for the student to be awarded a degree. Such a research project is also instructional in the sense that the student is becoming an expert in both content and research skills in the process of completing the task. The professor reviews and provides feedback (assessment information) about needed changes at many points along the way. Thus, the line between instruction and assessment becomes blurred.

Some suggestions for developing authentic tasks adapted from Davey and Rindone (1990) from the Connecticut State Department of Education are

1. Start with an issue, idea, scenario, or problem and test it by asking how important it is; how engaging would students find it; how relevant to "real life" it is; what content areas will be learned in the context of the project.
2. Begin to define the task more fully by asking what knowledge, competencies, or dispositions students will have to use to complete such a task (i.e., what instructional objectives are served?). Revise and elaborate on the task as needed until you are satisfied with both the purposes or objectives and the task structure itself.
3. Consider the context. What should be the medium for the product (oral, written, computer, debate)? Should the task include group activities? Should experts from the community be brought in?
4. Consider the administration of the task. What do students need to know before the task is given them? What difficulties might they encounter? How will assistance be provided?
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 (e.g., communicated conclusions in a clear and organized way, using visuals)? Who might assess student performance other than the teacher (e.g., community professionals, other students)?
6. Have colleagues and perhaps students review the task and the criteria for judging success. Revise as needed.

Portfolios

Portfolios refer to the process of assessing student progress by collecting examples of student classwork (Wolf et al., 1992). Physically, it is a container of evidence of a student's achievements, competencies, or skills. It is a purposeful collection in the sense 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? Although portfolios in art and writing are very common and familiar, math and science teachers are also beginning to employ portfolios as ways to collect, organize, reflect on, and display completed work (Hamm & Adams, 1991). Portfolios might best be considered as a tool to promote communication between the teacher and student about student understandings, strengths, weaknesses, progress, and self-reflections. Portfolios can bind teachers and students together in very personal and supportive ways.

The use of portfolios, like any assessment method, starts with a consideration of purposes. What objectives do you have for students that are best assessed by a portfolio and what is the portfolio supposed to demonstrate (Collins, 1992)? Some examples of purposes are shown below.

1. If the ability to design experiments were an objective for students, a portfolio might be used to show progress in this ability over the year by including work on different assignments. Or, if the portfolio were to be used to better understand how students go about designing an experiment, it could contain all activities, drafts and revisions leading up to the final design with reflections from the student about their thinking at different stages in developing the design.
2. If improving creative writing around science content knowledge were an important instructional goal, a portfolio might showcase a student's favorite pieces. Parents could help the students reflect on and choose their best pieces.

3. If a goal of instruction is that students will read, summarize, and evaluate information in newspaper articles on science topics, the portfolio might represent evidence of their increasingly sophisticated efforts at critiquing these articles.

4. Portfolios could be used as evidence of basic content knowledge. Students could be asked to keep all test papers in a portfolio and write a reflection piece after each test on how they could improve their responses.

5. Portfolios could be individualized such that students display work showing their particular strength or progress in a weaker area rather than having the portfolio be the same for all students.

There is no formula or single right way to do portfolios. Rather, designing a portfolio project represents a series of decisions. Some of the design questions to be answered after the instructional objective has been determined are listed below (Collins, 1992).

1. For what purpose will the portfolio be used (student self-reflection, a grade, a narrative report on student progress, parent conferences, promotion to the next grade)?
2. How often will the portfolio be reviewed and by whom?
3. Which portfolio pieces of work are required and which are selected by students?
4. Will work be produced alone or can it be a group portfolio?
5. Where will the portfolio be kept?
6. How much work should be included?

Like any of the other methods of looking at student work, portfolios involve the development of criteria for judging good work. Student progress relative to a certain instructional objective might be evaluated by developing criteria for individual pieces, for the portfolio as a whole, or for students' written reflections on work in the portfolio (Arter & Spandel, 1992). Criteria for a whole portfolio might be the quality or variety of pieces included, the quality and depth of self-reflection included, or the growth in performance as indicated by the products. Students' self-reflections about pieces might be evaluated on thoroughness support of statements made by describing specific aspects of the work, and how well ideas are synthesized (Arter & Spandel, 1992).

Chapter IV
RUBRICS AND GRADING

Rubrics

Multiple-choice items can be scored objectively. The student is offered a fixed number of options, and the option selected is compared to a scoring key. Given the scoring key, any teacher would score the items in the same manner. Performance-based methods (open-ended questions, performance tests, journals, exhibitions, and portfolios) depend to a greater extent on teacher judgment of a response.

The term "rubric," rather than scoring key, is used to refer to the guidelines laid out for judging student work on performance-based tasks. There are at least five ways to arrange the criteria against which the student's work will be judged.

1. **Point system.** 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 were given the appropriate measuring 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 (adapted from the NAEP example) might look like this:

4 points	if the response states that both types of sugar dissolve faster when stirred, but loose sugar still dissolves faster than cubes
3 points	if the response indicates that stirring made a difference but doesn't describe the relative difference (that loose sugar still dissolves faster)
2 points	if the response describes the relative speed (loose dissolves faster) but not the effects of stirring or if the response just describes what happens (stirring makes the sugar fall apart)
1 point	for incorrect responses
0 points	for no response

Typically, essay questions require longer and more complex responses from students. If an essay question is to be scored with a point system, the features of a successful response should be identified prior to grading the essay and given numerical values representing their value relative to each other. That is, some features of the answer might be worth more than others (e.g., perhaps each reason provided for a phenomenon is worth 2 points and the quality of the overall organization of the response is worth 1 point). The student's score is the sum of the point values for each feature identified by the teacher as present in his or her response.

2. **Checklists.** A checklist can be used to indicate that a student has effectively completed the steps involved in a task or demonstration. Checklists may be applied to written work (e.g., journals) or observable behavior.

Suppose students are asked to do an experiment to find out whether loose sugar dissolves faster or slower than sugar cubes (NAEP, 1986). Students are observed individually and a checklist is used for the following behaviors:

___ 1. loose sugar tested
___ 2. sugar cubes tested
___ 3. measurement of water and sugar were problematic
___ 4. measurements made effectively
___ had problems timing how fast sugar dissolved
___ 6. effectively timed how fast the sugar dissolved
___ 7. final answer consistent with evidence

The information might be used to diagnose students' strengths and weaknesses relative to different aspects of conducting the experiment.

Checklists are also effective in getting students to check their own work. For example, prior to turning in journals to the teacher, students could be given a checklist with all the assignments to be included. Or the journal could be reviewed and the checklist completed by another student.

3. **Analytic rating scales.** Rating scales describe performance along a continuum. Analytic rating scales are used to separately describe a product or performance on multiple dimensions. For example, in a writing task, the dimensions or criteria that might be rated are organization, mechanics, and creativity. Each important dimension of the task performance is rated on a two– (e.g., "acceptable," "not acceptable") or more (e.g., "inadequate," "partially satisfactory," "satisfactory," "exemplary") point scale.

For example, if students were asked to write a letter to someone from a different time period, such as ancient Egypt, on how measurement has changed over the years, the science teacher might rate students' work from 1 to 4 in terms of knowledge of measurement demonstrated. The English teacher might be asked to rate the same work using two dimensions: mechanics and organization, with a 1 to 4 rating given on each dimension. Thus, students would receive diagnostic feedback from two teachers (science and English) on three dimensions (knowledge of measurement, mechanics, organization) of their performance.

The strength of analytic rating scales is that they offer diagnostic information to the student about the strengths and weaknesses of their performance on a variety of dimensions so that they can better target the areas of their performance that need to be improved. The dimensions chosen and the descriptive categories used for the rating scales need to be chosen so that they communicate to students what is important to do well.

There are many different ways to label scale points. One approach to labeling scale points is to describe levels of goal attainment on the dimension identified (Davey & Rindone, 1990). For example, suppose a physics problem-solving task was presented to students. A rating sheet, as shown in Figure 10, might be used to provide feedback to students.

Figure 10

Sample Analytic Scale

Task Criteria	Ratings			
	Exceeds goal	Meets goal	Approaches goal	Goal not yet met
1. Correctly state the problem and identify the information needed to solve it and the steps needed to arrive at a solution	____	____	____	____
2. Produce reasonable estimates of data values not identified but needed for the solution	____	____	____	____
3. Apply concepts and formulas related to motion(velocity, acceleration, average speed)	____	____	____	____
4. Make accurate conversions as needed to solve the problem	____	____	____	____
5. Communicate 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.

4. **Focused holistic rating scales.** Rather than assigning separate scores for each important aspect of task performance, focused holistic ratings consider all the criteria simultaneously and result in a single summary rating or grade. This approach may be most appropriate when the purpose is to provide students with an overall index of their performance on a task or product.

For example, if high school students were asked to conduct a comprehensive interdisciplinary investigation on some practical problem presented to them, a focused holistic rating scale might be used. The student must demonstrate the following to received an "A":

- give clear responses that show understanding of the scientific concepts and ideas addressed;
- use scientific processes and tools to gather, record, and organize appropriate data in a logical fashion;
- write strong supporting conclusions based on evidence collected; and
- emphasize any additional data that is needed.

At the other extreme, a "F" response might be one in which the student does not demonstrate any understanding of the problem or concept; has data that are missing, inappropriate, or incomplete; and makes no attempt to state or complete conclusions.

It is important to keep in mind that the development of focused holistic scales can be time-consuming. An accurate description of different levels of performance on a task requires the careful analysis of many student responses.

The rubric shown for the performance test in Figure 8 is another example of a focused holistic approach to grading student responses. If the performance test on electricity represented an end-of-unit test, a focused holistic scale such as the one shown can be used to easily translate student responses into a grade.

5. **Holistic.** With holistic scoring, no specific rating criteria are identified. Instead, model responses are selected that represent numbers on the scale to be used. Student responses are compared to the model responses and are given a number corresponding to the model response they are most like.

In developing rubrics for student work, some teachers (High Success Network training materials, 1992) are finding the following helpful:

1. Provide examples to students of work that reflects the different points on a rubric. For example, if essay questions are to be evaluated on the degree to which conclusions are justified, provide examples of a "weak" justification as well as an "exemplary" justification.
2. Once students have developed an understanding of rubrics, involve them in brainstorming rubrics for work to be done so that they experience some ownership over the judging process.
3. For student work that is to be graded, be clear on the rubrics or criteria with students before they begin the task.
4. Try providing students with sample rubrics that they can use to assess their own work or that of peers.

Grading

The purposes of grading systems are twofold (Stiggins, 1991). First, they are a way of communicating to students, parents, and other decision-makers something about the student's achievement status. Second, they are intended as motivators (e.g., to impress upon students that the work is important and worth their attention).

Grading students' achievement status involves the teacher in a series of decisions. In *A Practical Guide for Developing Sound Grading Practices*, Stiggins (1991) describes some of the decisions that must be made if a single grade is to be given over an assessment period.

1. What criteria should be considered in determining a report card grade?

   A grade over a grading period is usually considered as a composite measure of student achievement on the objectives of the instruction, rather than a measure of student interest, attitude, or personality.

2. As a measure of achievement, what and how much grading data should be gathered?

   For the grade to be a valid measure of achievement in a subject or course, there must be a sufficient sampling of student performance on the critical subject or course objectives (or targets) to provide a fair and reliable assessment of the student's achievement status.

The teacher must be the judge, but at one extreme, using only one measure (e.g., a one-hour, paper-and-pencil exam) to determine a report card grade clearly is not a sufficient sample of student performance. At the other extreme, assessing student performance daily would not provide students with the time needed to develop competencies and skills preparatory to being assessed.

Chapter II asked that you identify four critical science outcomes held for students. These outcomes are targets for the instruction provided. Traditional and more performance-based methods are used to assess student progress toward these goals. All of the assessment methods described in Chapter III represent legitimate ways of assessing achievement status, but must be matched carefully with desired outcomes for students.

3. How are individual grades or scores combined into a single grade at the end of a grading period?

   Performance-based assessments can be readily incorporated into any grading system. Suppose you had the following five goals in mind for students. Students will be able to

   1. demonstrate knowledge of electricity,
   2. write creatively on a topic in a way that demonstrates knowledge of electricity,
   3. maintain a journal of daily investigations,
   4. work with a group on an extended project or problem-solving exercise having to do with electricity, and
   5. understand and use science process skills.

Given these five instructional objectives, student assessment data as shown in Figure 11 (p. 47) might be used to arrive at a report card grade. The first column, "Maximum Points," reflects the rubrics designed for each assessment.

In Figure 11, demonstrating knowledge represents 37 percent (100/270) of the total grade; science process skills, maintaining a journal, and completion of an extended group project each represent 19 percent (50/270); and creative writing represents 6 percent (20/270). The proportion of the total grade accounted for by individual assessments should communicate the relative importance of different desired outcomes (more important outcomes carry more weight).

Figure 11

Sample Grading Period Weighting System

Assessments of Five Instructional Objectives	Maximum Points	Student A: Points Earned	Weight	Student A Score/ Max. Score
1a. Paper and pencil test on electricity	50	40	1	40/50
1b. Performance test on electricity	25	20	2	40/50
2. Weekly lab assignments on science process skills (5 assignments X 10 pts. each)	50	45	1	45/50
3. Two creative writing tasks (10 pts each)	20	20	1	20/20
4. Journal	50	50	1	50/50
5. Extended group project	50	45	1	45/50
Total	245	220		240/270

The weighting system used in deriving report card grades should be related to course objectives and explained to students so that they know their goals. In the example above, students might be informed at the beginning of the grading period of the five instructional objectives and the assessments to be used. The number of points needed for the different grade symbols used could also be communicated.

In such a point 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 245 or more points constituted an A, but no students earned this many points due to a few poorly worded items on one 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 goals. Checklists, analytic scales, and narratives can be used in addition to grades or as alternative means of reporting.

In one science curriculum reform project, a teacher identified the instructional objectives for the six-week grading period and rated performance on the objectives on a five-point scale (Figure 12) in hopes of better communicating with students and parents about student progress. Some of the objectives listed in Figure 12 under "Science Processes" change during the year depending on the particular content or skills emphasized during a grading period. The form is being used in addition to a traditional report card.

Figure 12

Sample Science Progress Report (Sixth Grade)

Name ______________________________ ____________ Six Weeks

Rating Scale:

5 - Superior 4 - Very Good 3 - Average 2 - Needs Improvement 1 - Unsatisfactory

	Work Habits
	Used lab materials safely
	Returns signed papers
	Follows directions
	Makes good use of time
	Punctual

	Keeping Journals
	Keeps journal organized
	Completes assignments
	Expresses thoughts through journal entries

	Cooperative Skills
	Works well with group members
	Performs group role(s)
	Respects rights of others
	Accepts responsibility for behavior
	Shows self-control in talking and actions

	Science Processes
	Can orient a compass
	Can collect and organize data
	Can tell relationship between shadows and the sun's position
	Can use the sun to determine cardinal directions
	Participates in family science

Teacher Comments:

Family Comments:

Source: Terri Hardin, A.G. Cox Middle School, Pitt County Schools, Winterville, N.C.

Chapter V

Getting Started

Traditional practices in assessment are based on beliefs about the purpose of education that are being publicly discussed and challenged. Assessment practices 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, reflect on their success or failure, and make revisions. Just as student learning is an individual process that is self-directed and 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?

1. Some districts have initiated district-wide staff development efforts in assessment. The literature on staff development 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 et al., 1990).

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

- spend one 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 (i.e., observation forms, interview protocols, open-ended questions, performance tests, journal criteria, exhibition projects, portfolio tasks),
- 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.

2. 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 teacher(s) involved, that represent jumping-off points for changing or at least reflecting on assessment practices.

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

a. Articulate one very important desired student outcome (refer to Chapter II). 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 in Chapter III and choose an approach to assessing students' competence on this dimension that you have not tried before. Try the assessment approach and see what you learn about student performance and about the assessment method you choose.

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

c. Start a list of the advantages and disadvantages of each of the assessment methods described in Chapter III. What do you feel you need to know from someone who has tried each method before you go any further? Contact SERVE for more information.

d. Develop a chart (see Figure 11 on p. 47) showing how you combine assessment data in obtaining student report card grades. What kind of weighting system are you using?

e. Analyze the tests you have used in the past. Try to improve the items used referring to the information provided in Chapter III on open-ended questions and performance tests or consider how you might improve or make more explicit the rubrics for the items.

f. Start a folder of assessment samples from released state tests or item banks, other teachers, district tests, or published articles and critique them for your purposes.

g. Review the hands-on, experiential, or lab activities you use with your students. Identify those which are most essential 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 a facilitator, coach, or critic rather than as authority figures who dispense all information and knowledge.

References

Arter, J.A. & Spandel, V. (1992) Using portfolios of student work in instruction and assessment. *Educational Measurement: Issues and Practice, (Spring)*, 36-44. (Also available from Northwest Regional Educational Laboratory, 101 S.W. Main Street, Suite 500, Portland, OR 97204, is a "Portfolio bibliography.")

Blumberg, F., Epstein, M., McDonald, W., & Mullis, I. (1986). National Assessment of Educational Progress: A pilot study of higher-order thinking skills assessment techniques in science and mathematics. Final Report Part II. Pilot tested tasks. Princeton, NJ: National Assessment of Educational Progress.

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

Collins, A (1992). Portfolios: Questions for design. *Science Scope*, *15* (6), 25-27.

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

Freedman, R.L. (1992) Writing, student portfolios , and authentic assessment. *Science Portfolio: The Watershed*, 1 (1), 8-9. (Also see *Connections: Science by Writing.* Paradise, CA: Serin House Publishers).

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

Haertel, E. & Calfee, R. (1983). School achievement: Thinking about what to test. *Journal of Educational Measurement*, *20* (2), 119-132.

Hamm, M. & Adams, D. (1991) Portfolio. *The Science Teacher*, *58* (5), 18-21.

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.)

Hopkins, K.D., Stanley, J.C. & Hopkins, B. R. (1990) *Educational and psychological measurement and evaluation*. Englewood Cliffs, NJ: Prentice Hall.

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.

Loucks-Horsley, S., Kapitan, R., Carlson, M., Kuerbis, P., Clark, R., Nelle, G., Sachse, T., & Walton, E. (1990). *Elementary school science for the 90's*. Andover, MA: The National Center for Improving Science Education, The Network.

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

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

NAEP: Learning by Doing (1987). Princeton, NJ: Educational Testing Service.

Payne, D.A. (1992). *Measuring and evaluating educational outcomes.* New York: Macmillan.

Raizen, S., Baron, J.B., Champagne, A.B., Haertel, E., Mullis, I.V.S., & Oakes, J. (1990). *Assessment in science education: The middle years.* Andover, MA: The National Center for Improving Science Education, The Network.

Science Framework for 1994 National Assessment of Educational Progress: NAEP Science Consensus Project (1992). Washington, DC: National Assessment Governing Board.

Sizer, T.R. (1984). *Horace's compromise: The dilemma of the American high school.* Boston: Houghton, Mifflin. (For information on the Coalition of Essential Schools, contact Brown University, One Devol Square, Providence, RI 02912.)

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

Stiggins, R.J. (1991). *A practical guide for developing sound grading practices.* (Available from the Northwest Regional Educational Laboratory, 101 S.W. Main Street, Portland, OR 97204.)

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: AERA.

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.

Wolf, D.P., LeMahieu, P.G., & Eresh, J. (1992). Good measure: Assessment as a tool for educational reform. *Educational Leadership, 49* (8), 8-13.

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

Appendix

Reform in Curriculum and Instruction

Reform in Curriculum and Instruction

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 that were resistant to change:

- many teachers were inadequately prepared in science and math, particularly at the elementary and junior high 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 that

- curricular alternatives were developed and disseminated (PSSC, BSCC, SCIS),
- content was modernized,
- new curricular materials emphasized science processes and hands-on work, and
- science manpower needs were met.

The reform of the 1990s differs from the earlier science curriculum reform in that it is a subset of a much larger educational reform movement fueled by a concern that our students will not be internationally competitive as adults. Changes are being proposed across the curriculum, emphasizing higher-order thinking skills and problem-solving. In North Carolina, for example, all public schools have been directed to infuse critical thinking throughout the North Carolina Course of Study.

The emphasis on science education in previous decades that resulted in the development of curriculum materials provided a framework on which to build current efforts. However, the current efforts differ from prior curricular reform movements in that they are geared toward scientific literacy for all students, not just better science education for future scientists. This "science for all" goal is critical if people are to have a basis for making informed decisions about issues like nuclear power, personal health, the environment, and reproduction (Loucks-Horsley et al., 1990).

Several national science education reform efforts are described below:

1) With the support of the U.S. Department of Education, the National Academy of Sciences (through the National Research Council) has recently initiated a major effort to develop world-class standards for what students should know and be able to do in science, which are expected to be drafted by 1994. These standards, like those developed by the National Council of Teachers of Mathematics, will serve as guides for states, districts, and schools. Standards will be developed in the areas of curriculum, teaching, and assessment to present a vision of science education against which schools, districts, and states can compare themselves.

2) There have been at least two other significant efforts to develop some consensus on student outcomes for science. Project 2061 is a reform effort of the American Association for the Advancement of Science (AAAS). The project issued a report in 1989, called *Science for All Americans*, that suggested the knowledge, skills, and attitudes that students should have as a result of their K-12 science instruction. The project describes the outcomes expected of 17-year-olds. Teams of educators are working on developing curriculum models and more specific student outcome statements based on the AAAS publication. For more information, write to

   Project 2061,
   1333 N Street, NW,
   Washington, D.C., 20005.

The National Science Teachers' Association has a reform effort that proposes dismantling the layer-cake organization of science courses (biology, then chemistry, then physics) in favor of courses that integrate material from the different science disciplines so that all disciplines are taught every year. In an effort to show how science content might be better organized to promote scientific literacy for all students, the NSTA published *The Content Core: A Guide for Curriculum Designers. The Core* suggests the topics that should be covered in grades 6-12 and at what level they might be covered. For more information on SS&C or a copy of The Content Core, write to

> SS&C c/o National Science Teachers' Association,
> 1724 Connecticut Avenue, NW,
> Washington, D.C. 20009.

3) Many states already have standard course-of-study guides listing desired outcomes for students at different levels of science instruction. In the past, many of these state curriculum guides have focused on students' obtaining knowledge of discrete bits of information such as

   - knowing about human body systems and their functions and
   - knowing about the structure of atoms and molecules.

However, states are currently developing new curriculum frameworks that emphasize skills needed in the process of doing science in addition to content knowledge. For example, Alabama, Florida, Georgia, Mississippi, North Carolina, and South Carolina all have plans to publish new or revised science curriculum frameworks in 1993 or 1994. SERVE has recently received a grant to develop a Regional Mathematics and Science Consortium. Through this grant, SERVE will keep the region informed of exemplary science and math curriculum frameworks through a collaborative effort among all regional educational laboratories to collect, analyze, and synthesize state frameworks for math and science.

How should science be taught?

If science is not about the memorization of facts, but a problem-solving process through which we work to understand problems, how can science education be changed to reflect the problem-solving process?

Ir Windows into Science Classrooms, Tobin, Kahle, and Fraser (1990) note the following:

> "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.

For a teacher to instruct in the processes of science rather than about the established facts of science, a fundamental shift in activities and priorities is required. The teacher must move from

1) conducting an exercise to illustrate what is asserted to be the correct answer by the textbook, to
2) assigning problem-solving exercises during which students are asked to consider specific questions by testing a particular hypothesis or alternative hypotheses" (p. 151).

According to these authors, science classes are often characterized by rote learning and recall of memorized information. They suggest that in order to move toward the development of problem-solving skills in their students, teachers must incorporate rigor, relevance, and representative structure as elements of their teaching.

Rigor: A rigorous science classroom should have instructional objectives built around higher-order processes (problem-solve, predict, observe, analyze) and not just the ability to memorize facts. The amount of work completed and the difficulty of the work posed to students are issues in developing rigorous science instruction. Work for students should be designed to give them experience in the processes of science (recalling and imagining, classifying and generalizing, comparing and evaluating, analyzing and synthesizing, deducing and inferring). Finally, if higher-order cognitive objectives are targeted and students are given work that goes beyond rote memorization, then methods of student evaluation and grading must be consistent and go beyond direct recall of information.

Relevance: Regarding the information that is presented in classrooms, Tobin, Kahle and Fraser (1990) suggest it is critical to think about why information is taught and, in particular, its relevance to students' lives. For example, should a unit on vertebrates concentrate on having students learn the names of the bones in the body, or understand the biology and epidemiology of AIDS?

Representative structure: Some have characterized this issue as the selection of fewer topics taught well. Tobin, Kahle, and Fraser (1990) argue that the goal of instruction should be to help students come to terms with the major organizing ideas of a discipline and why the idea or theme (such as evolution) occupies the place it does within the discipline.

Currently, numerous curriculum reform projects are trying to articulate and implement new visions for science teaching. Two of these that might be contacted for information are:

Dr. Larry Rainey
Project Director, Integrated Science
The University of Alabama
206A Temple Tutwiler
Tuscaloosa, AL 35487-0167

Erma Anderson
Project Manager, Scope, Sequence and Coordination
National Science Teachers Association
1724 Connecticut Ave. NW
Washington, DC 20009

Feedback Form

How to Assess Student Performance in Science

Meeting the needs of educational practitioners is very important to us at SERVE. Your feedback on this publication will permit us to better assist you, and your recommendations will be incorporated into future editions. Please help us by providing a brief response to the following:

1. Is this publication a useful resource for supporting teachers in reflecting on science assessment practices? ❑ Yes ❑ No Why or why not?

2. Which part did you find most useful and why?

3. Which part did you find least useful and why?

4. How do you plan to use this document, or how have you used it?

5. How could SERVE further assist you, your school or your district?

6. Would you be willing to participate in a Focus Group on alternative science assessments?
   ❑ Yes ❑ No

Name: ______________________ **Position:** ______________________

Affiliation: ______________________

Address: ______________________

City: ____________ **State:** ______ **Zip:** ________ **Telephone (___)** ________

Please mail or fax completed form to: SERVE
P.O. Box 5367
Greensboro, NC 27435
Phone: 800-755-3277
FAX: 919-334-3268

Thank you for completing this form.

SERVE PRODUCTS AND SERVICES ORDER FORM

SERVE
SouthEastern Regional Vision for Education

Hot Topics

Developed with input from educators throughout the Southeast, these research-based guidebooks offer information, resources, descriptions of exemplary programs, and contacts for additional information. See p. 64 for prices.

Quantity

—— *Reducing School Violence*
- ❑ Regional Edition (110 pages)
- ❑ Florida Edition (112 pages)

—— *Appreciating Differences: Teaching and Learning in a Culturally Diverse Classroom* (116 pages)

—— *Interagency Collaboration: Improving the Delivery of Services to Children and Families* (118 pages)

—— *Using Technology to Improve Teaching and Learning* (90 pages)

—— *Schools for the 21st Century: New Roles for Teachers and Principals* (94 pages)

—— *Comprehensive School Improvement* (95 pages)

—— *Problem-Centered Learning in Mathematics and Science* (60 pages)

SERVE Reports

—— *Southern Crossroads: A Demographic Look at the Southeast*, by Harold Hodgkinson (90 pages, $7–see p. 64 for large-order discounts.)

—— *A Public-Private Partnership: South Pointe Elementary School* (31 pages, $5)

—— *Supporting Family Involvement in Early Childhood Education: A Guide for Business* (50 pages, $5)

EDTALK

Co-produced with the Council for Educational Development and Research, this series focuses on educational improvement. See p. 64 for prices.

Quantity

____*EDTALK: What We Know About Mathematics Teaching and Learning* (69 pages)

____*EDTALK: Rural Schools on the Road to Reform* (70 pages)

____*EDTALK: What We Know About Science Teaching and Learning* (70 pages)

Sharing Success

The SERVE *Sharing Success* series describes outstanding school and district programs in the Southeast that focus on the national goals for education. Single copies of the following are available at no cost:

❑ *Sharing Success: Mathematics and Science Education* (72 pages)

❑ *Sharing Success: Promising Programs in Preschool-to-School Transition* (62 pages)

Telecommunications

SERVE produces a variety of VHS tapes that give stakeholders in education the opportunity to see school improvement and reform initiatives in action.

Quantity

Southern Crossroads–Noted demographer Harold Hodgkinson examines demographic trends in the Southeast and discusses the challenges and opportunities they present (30 min., $29.95). *Special: Order *Southern Crossroads* publication and videotape for only $35. ________

A Generation of Hope–An examination of seven key characteristics of effective school programs to reduce and prevent student drug use (30 min., $29.95). ________

Journey Toward Change–offers ideas, strategies, and inspiration to school improvement teams from educators who have participated in successful school improvement initiatives. A complementary film to *Hot Topics: Comprehensive School Improvement* (25 min., $14.95). *Special: Order report and tape for $20. ________

Improving Mathematics and Science Education

- **General Audiences**–A close look at exemplary mathematics, science, and technology programs in several schools in the Southeast (30 min., $29.95) ________
- **Policymakers**–A panel discussion with the region's chief state school officers, business leaders, and others on promoting change; features exemplary school programs in mathematics, science, and technology (60 min., $29.95) ________
- **Teachers/Practitioners**–A panel discussion with award-winning teachers on how to implement innovative programs and practices; features exemplary school programs in mathematics, science, and technology (90 min., $29.95) ________

ORDERING INFORMATION

Prices for *Hot Topics, EDTALK, and Southern Crossroads* are as follows:

$7 each for 1-49 copies $6 each for 50-99 copies $5 each for 100 copies or more

Name:______________________________ **Title:**____________________
(Please print)

Address: (☐ home ☐ work)

____

____

City:______________________________ **State:**________ **Zip**__________

Phone:(____)____________________ **Fax:**(____)____________________

TO ORDER: Make check or purchase order to NEFEC/SERVE (Federal ID# 59-6000-821). Remove or copy this order form and send it with your check or purchase order to **NEFEC/SERVE,** Route 1, Box 8500, 3841 Reid Street, Palatka, FL 32177. Payment must accompany order.

Free SERVE PRODUCTS AND SERVICES ORDER FORM

SERVE SouthEastern Regional Vision for Education

The SouthEastern Regional Vision for Education (SERVE) offers a number of products and services designed to promote educational improvement. Single copies of the following items are available at no cost. Please check the box for the item(s) you wish to receive and mail to SERVE, 345 South Magnolia Drive, Suite D-23, Tallahassee, FL 32301-2950.

Name: ______________________ **Title:** ______________________
(Please print)

Address: ☐ home ☐ work (check one)

____

City: ______________________ **State:** ________ **Zip:** ____________

Phone: (_____) ______________________ **Fax:** (_____) ______________________

PUBLICATIONS

- ☐ *SERVE Brochure*—overview of laboratory activities
- ☐ *Sharing Success: Mathematics, Science, and Computer Education*—descriptions of 54 outstanding school and district programs in the southeast (72 pages)
- ☐ *Sharing Success: Promising Programs in Preschool-to-School Transition*—detailed descriptions of 19 schools and/or districts in the Southeast with outstanding early childhood transition programs (62 pages)
- ☐ *What Teachers Have to Say About Creating Innovations in Education: Proceedings from the Sharing Success Forum* (24 pages)
- ☐ *Minority Teacher Recruitment*—a policy brief (4 pages)
- ☐ *The Need for Improved Mathematics and Science Education*—a policy brief (8 pages)
- ☐ *What's Happening in Educational Assessment?*—a report for policymakers on the latest trends in assessment (13 pages)
- ☐ *A Resource Booklet on Transitions* (35 pages)
- ☐ *Reducing Baby Bottle Tooth Decay in Native American Communities (6 pages)*
- ☐ *Retention in Early Childhood Education* (4 pages)
- ☐ *Assessment in Early Childhood Education* (4 pages)
- ☐ *Early Childhood Regional Symposium Proceedings, 1991* (20 pages)
- ☐ *Spotlight on Success: Early Childhood Regional Symposium Proceedings, 1992* (35 pages)

COMPUTER/ELECTRONIC INFORMATION SERVICES

Database Information Services Clearinghouse (DISC)
DISC is SERVE's computerized research service for identifying periodicals and other publications on topics related to education. Offered free to educators in the Southeast, DISC provides access to the Educational Resources Information Center (ERIC) and over 300 other computer databases of information on education. Educators who request literature searches from DISC will recieve research packets that include abstracts of relevant articles and other publications, a copy of one or more relevant journal articles, and a copy of an ERIC document.

For further information about this service, call (800)352-3747 or write to SERVE at the address at the top of this page.

SERVE-Line
SERVE-Line is a nationwide computerized communication system that educators can use to access and exchange information. With SERVE-Line (and a computer and a modem), educators can send messages to each other through an electronic mail system; share opinions, ask questions, and make announcements pertaining to education; copy public domain educational software; and learn about educational events, the latest educational research, and other information of interest. The majority of SERVE-Line's menu items are available for use at no cost.

To connect with SERVE-Line via a computer and modem, dial (800)487-7605. For more information, call (404)577-7737 or (800)377-5011 or write to SERVE, 41 Marietta Street, NW, Suite 1000, Atlanta, GA 30303.

SERVE
R&D

SERVE OFFICES

P.O. Box 5367
Greensboro, NC 27435
800-755-3277
919-334-3211

345 South Magnolia Drive
Suite D-23
Tallahassee, FL 32301-2950
800-352-6001
904-922-2300

41 Marietta Street NW
Suite 1000
Atlanta, GA 30303
800-659-3204
404-577-7737

Delta State University
P.O. Box 3183
Cleveland, MS 38733
800-326-4548
601-846-4384