DOCUMENT RESUME

ED 347 045 SE 051 821

AUTHOR	Raizen, Senta A.; And Others
TITLE	Assessment in Science Education: The Middle Years.
INSTITUTION	Biological Sciences Curriculum Study, Colorado Springs.; Network of Innovative Schools, Inc., Andover, Mass.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	90
CONTRACT	R168B80001
NOTE	131p.; For the first set of reports, see ED 314 235-238.
AVAILABLE FROM	The NETWORK, Inc., 300 Brickstone Square, Suite 900, Andover, MA 01810 ($15.00).
PUB TYPE	Information Analyses (070) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	*Academic Achievement; *Elementary School Science; *Evaluation; Innovation; Intermediate Grades; Junior High Schools; *Middle Schools; *Science Education; *Secondary School Science; Student Attitudes; Teaching Methods

ABSTRACT

The mission of the National Center for Improving Science Education, a partnership between the NETWORK, Inc., and the Biological Sciences Curriculum Study (BSCS), is to promote changes in science curricula, science teaching, and assessment of student learning in science. The center analyzes and makes recommendations for policy and practice at the national, state, and local levels. As part of this task, the center synthesizes and translates the findings, recommendations, and viewpoints expressed in research studies and develops practical resources for policymakers and practitioners. This document is part of a second set of reports that focus on science and mathematics education for young adolescents. Included are chapters entitled: "(1) "Assessment: The Middle Years"; (2) "The Opportunity"; (3) "Goals for Science Education and the Assessment Challenge"; (4) "The Context of Science Education in the Middle Years"; (5) "Assessment in Middle-Level Science: Improving Current Practice"; (6) "Innovative Assessments: New Directions"; (7) "Assessments and Policy"; and (8) "Recommendations." Appended are the references, a listing of assessment panelists, and an index. (KR)

Assessment in Science Education: The Middle Years

Senta A. Raizen
Joan B. Baron
Audrey B. Champagne
Edward Haertel
Ina V.S. Mullis
Jeannie Oakes

The National Center for Improving Science Education

A Partnership Between

The NETWORK, Inc.
Andover, Massachusetts
Washington, D.C.

and

BSCS
The Biological Sciences Curriculum Study
Colorado Springs, Colorado

SE051531

Assessment in Science Education: The Middle Years

Senta A. Raizen
Joan B. Baron
Audrey B. Champagne
Edward Haertel
Ina V.S. Mullis
Jeannie Oakes

The National Center for Improving Science Education

A Partnership Between

The NETWORK, Inc.
Andover, Massachusetts
Washington, D.C.

and

BSCS
The Biological Sciences Curriculum Study
Colorado Springs, Colorado

National Center for Improving Science Education

The mission of the National Center for Improving Science Education is to promote changes in state and local policies and practices in science curriculum, science teaching, and assessment of student learning in science. To do so, the Center synthesizes and translates the findings, recommendations, and perspectives embodied in recent and forthcoming studies and reports, and develops practical resources for policymakers and practitioners. Bridging the gap between research, practice, and policy, the Center's work is intended to promote cooperation and collaboration among organizations, institutions, and individuals committed to the improvement of science education.

The Center, a partnership between The NETWORK, Inc., of Andover, Massachusetts and Washington, D.C., and the BSCS (Biological Sciences Curriculum Study) of Colorado Springs, Colorado, is funded by the U.S. Department of Education's Office of Educational Research and Improvement. For further information on the Center's work, please contact The NETWORK, Inc.; 300 Brickstone Square, Suite 900; Andover, Massachusetts 01810.

To order copies of the Center's reports for the elementary and middle years, or the Center's integrative reports, please contact the Publications Department, The NETWORK, Inc.; 300 Brickstone Square, Suite 900; Andover, Massachusetts 01810. Bulk order discounts are available.

This report is based on work sponsored by the Office of Educational Reserch and Improvement (OERI), U.S. Department of Education, under grant number R168B80001. The content of this report does not necessarily reflect the views of the OERI, the Department, or any other agency of the U.S. Government.

Foreword

The mission of the National Center for Improving Science Education is to promote changes in science curricula, science teaching, and assessment of student learning in science. The Center analyzes and makes recommendations for policy and practice at the national, state, and local levels. As part of this task, the Center synthesizes and translates the findings, recommendations, and viewpoints expressed in recent and forthcoming studies and develops practical resources for policy makers and practitioners. The Center's work bridges the gap between research, practice, and policy, and it promotes cooperation and collaboration among organizations, institutions, and individuals committed to improving science education. This report is one in a series. The first set of five reports, released between mid-1989 and mid-1990, focused on science education in the elementary years:

- Science and Technology Education for the Elementary Years: Frameworks for Curriculum and Instruction
- Assessment in Elementary School Science Education
- Developing and Supporting Teachers for Elementary School Science Education
- Getting Started in Science: A Blueprint for Elementary School Science Education
- Elementary School Science for the 90s: A Guide to Action

The first three reports focus on curriculum and instruction, assessment, and teacher development and support. The fourth report is a summary of the findings and recommendations documented in the first three. The Action Guide is a practical tool that science supervisors can use to carry out the Center's recommendations. This document, Assessment in Science Education: The Middle Years, is part of a second set of reports that focus on science and mathematics education for young adolescents. The other reports in this second series include:

- Science and Technology Education for the Middle Years: Frameworks for Curriculum and Instruction
- Developing and Supporting Teachers for Science Education in the Middle Year.
- Building Scientific Literacy: A Blueprint for Science in the Middle Years
- Science for the Middle Years: A Guide to Action

The synthesis and recommendations in this report were formulated with the help of the panel whose members are listed on page 113. We gratefully acknowledge the help of the many people who have supplied materials and made recommendations and suggestions for the text of the report. While the list would be too long to acknowledge each contributor, we wish to give special thanks to Sally Crissman of Shady Hill School, who provided several of the assessment anecdotes for chapter 5. We also thank Elizabeth Stage of the California Science Project, University of California and William Cooley, University of Pittsburgh—their reviews did much to help improve this report. Thanks are also due to the support of the Center's monitor at the U.S. Department of Education, Wanda Chambers.

The Center, a partnership between The NETWORK, Inc. of Andover, Massachusetts, and the Biological Sciences Curriculum Study (BSCS) of Colorado Springs, is funded by the U.S. Department of Education Office of Educational Research and Improvement. For copies of this report or further information on the Center's work, please contact The NETWORK, Inc.; 300 Brickstone Square, Suite 900; Andover, Massachusetts 01810.

Contents

Chapter I

Assessment: The Middle Years

In this report, the Center addresses the assessment of early adolescents' science learning. The Center defines early adolescence as ages ten through fourteen. As we point out in chapter 4, school arrangements for this age group vary tremendously. Students in this age group might be attending an elementary school, a middle school, or a junior high school; each of which can span a variety of grades—even a K through twelfth grade school. One teacher or a team of teachers working together might provide instruction; or, individual teachers responsible for a specific subject—as in high school—might provide instruction. Often, administrative needs and traditions govern school organization and instruction within a district.

In this report, we address science education and assessment for *all* early adolescents, no matter what kind of school they attend. Because each of the terms *middle school, junior high school*, and *middle grades* carries organizational and instructional connotations, we use the more neutral *middle level* (or *middle years*) and *early* (or *young*) *adolescents* when we discuss science education for students in the ten-through fourteen-year-old age group. When discussing specific types of schools, we indicate which grades or age groups are appropriate.

Obviously, anyone engaged in an effort to help schools do a better job in science education for early adolescents must focus on two issues: (1) improving the science curriculum and science instruction, and (2) improving the quality of teaching and the competence of science teachers. But why worry about assessment? Six reasons are readily apparent, and, although the reasons for the classroom teacher will differ somewhat from those of the policymaker, together these reasons provide a strong case for improving current assessment practices. The teacher should use assessment for the following reasons:

- Assessments help guide instruction and make it more effective. Assessment should be used to establish what students bring to the classroom and what they are learning as instruction and classroom activities proceed.
- Assessments impress on the students, school staff, and parents the importance of science education and the expectations for science learning at the middle level.
- Assessments document accurately and comprehensively each student's progress at the end of an extended period of instruction—a semester or school year, or when a student moves on to a new classroom.

The policymaker should use assessments for three purposes:

- To monitor the outcomes of science instruction, and, in particular, the students' achievements and competencies in science.
- To provide, when combined with other information, the base for formulating approaches that might improve science education.
- To provide guidance on how resources invested in science education might be augmented or used differently.

Through these means, assessment can—and does—exert a powerful influence on science education, an influence that has grown as mandates for assessment have grown. Whether this influence is for good or ill, however, depends on how tests and other forms of assessment are constructed and their results used. The goals of science education, curricula and instructional techniques that reflect these goals, and the tests and other means of assessment used to establish what the students have learned and can do in science. Otherwise, assessments will distort the goals, the curriculum, and what the teacher chooses to do in the classroom. This is as true for assessments controlled and conducted by teachers for their own purposes as it is for externally mandated assessments intended for policy uses. Moreover, it does little good to improve teachers' assessment practices without making consonant improvements in large-scale assessments, so that both will reflect the kind of science education that advances the intellectual development and interests of young adolescents.

In the Center's report on assessment in elementary school science education (Raizen et al., 1989), we focused mainly on how assessment can serve instruction, that is, how teachers might enlarge and improve their assessment strategies by monitoring not only their students' progress in science, but the effectiveness of their own science instruction. But externally mandated, large-scale assessments conducted for policy purposes also must encourage and be consonant with good science teaching. Therefore, in the previous assessment report, we attended to this type of assessment as well and explored the inherent difficulties in tests administered to large numbers of students.

In the current report, we maintain our emphasis on improving teachers' assessment practices and have limited our treatment of large-scale assessments, because the tests and assessments teachers carry out in the classroom have more direct consequences for individual students—for their learning and their future engagement with science—than do district, state, or national assessments. Also, teachers have available an array of assessment strategies that can deeply probe the students' progress and link with more relevance to instructional practices in particular classrooms. These useful strategies are only just beginning to be introduced into large-scale assessments. Nevertheless, the lack of correspondence between many large-scale assessments and good science education for early adolescents continues to be a troubling concern.

In this report we first review the capabilities and interests of early adolescents and consider the nature of an education, especially in science and technology, that can build on those capabilities and interests. In chapter 2, we discuss what is known about the cognitive and social development of ten-to fourteen-year-old students. Our coverage includes, but is not limited to:

- the student's increasing potential for engaging in the kind of thinking that characterizes science,
- the student's developing reasoning skills, and
- the growing student interest in evaluating themselves and others,
- the students' continuing need for concrete experiences, even as they are helped to develop formalized abstract thinking patterns.

In chapter 3, the Center presents goal statements it has developed in the companion curriculum and instruction report (Bybee et al., 1990). The goals reflect the growing capabilities of early adolescents to deal with science content and methods, and they reflect widely made recommendations for the education of students in this age group. The goals address not only science content, but also the relationships between science and technology and the relationships between science, technology, society, and individuals—a particularly motivating subject for this age group. Also discussed in chapter 3 are the assessment challenges these goals pose and the knowledge, skills, and dispositions the students should acquire.

The dilemmas of assessment at the middle level are quite like those encountered in the elementary grades—especially those concerning the need to assess thinking skills as opposed to assessing knowledge of subject matter. Assessment at the middle level encounters significant complications, however. Early adolescents exhibit thinking skills that are more complex than those exhibited by elementary students, and their knowledge base is larger. Also, if one educational goal is to develop critical-thinking and problem-solving capabilities for a variety of situations, assessment of transfer must be addressed. Furthermore, if learning-to-learn and self-assessment skills are educational goals for the adolescent years, strategies for their assessment must be developed and included in assessment batteries.

Because assessment must be set in the schooling context, we next review, in chapter 4, the nature of science programs and their broad middle-level school environment. We first consider current recommendations for the education of early adolescents in general and science education in particular. We then contrast and compare those recommendations with actual practice in today's middle-level schools.

Chapters 5 and 6 are the core of this report. In these chapters, we detail our conception of assessment and instruction in the service of science learning. In chapter 5, we point out the opportunities presented by the growing cognitive abilities of early adolescents. These opportunities include assessments that inform teachers of their students' progress in science and help teachers guide the course of instruction. We recognize the continuing, although limited, utility of well-designed traditional tests, but we also attempt to broaden the definition of what counts for assessment. Several examples illustrate how scientific inquiry itself can serve to provide assessment opportunities and how teachers can weave assessment into their science teaching. In chapter 5, we stress approaches possible now in any good middle-level science classroom. In chapter 6, we consider future directions in assessment and discuss several approaches still in experimental or trial stages. Several of these approaches come from state initiatives for assessing science or mathematics learning; thus, chapter 6 builds a bridge to chapter 7, in which we take up assessment for policy purposes. A key point of chapter 7 is that decision makers who wish to improve science education need information not only about the full range of learning outcomes, but also about the context in which these outcomes are being achieved—types of students, characteristics of programs and instruction, and types of teachers and teaching conditions.

In summary, what does good assessment of middle-level science education look like?

- Assessments should match exemplary instruction. Assessment exercises should be indistinguishable from good instructional tasks and will often be interwoven with them.
- Exercises should include hands-on performance tasks that allow students to demonstrate their proficiency in laboratory activities and scientific thinking.
- Assessments should probe the student's depth of understanding as well as knowledge of subject matter
- The emphasis should be on both the approach and the product: how an answer was obtained, how the student carried out a hands-on activity or conducted an investigation, and the student's final result.
- Some assessments should be built around a student's research or design project, free from the time constraints usually imposed by tests and assessments. Opportunities should be provided for self-assessment and course correction as the students proceed through the project, so that the teacher can check whether the student's proficiency in these important management skills has grown. Such projects also would allow judgments on competence in writing, presentation of data, the use of mathematics, and—if appropriate and available—use of the computer.
- The notion of "product" or "performance" must be enlarged to include not just written reports about experiments and answers to test questions, but also speeches, models, drawings, group presentations, and displays.
- There should be opportunities for group work designed around tasks too complex for students to undertake individually. In addition to providing information on the student's science learning and performance, this would allow the teacher to judge how effectively the individual communicates and contributes to the group, that is, the student's ability to collaborate effectively.

- If policymakers wish to use assessments for making improvements in science education, then they must take care to collect information not only on student outcomes, but also on the schooling context and science programs that produce the student outcomes. In chapter 8, we conclude this report with recommendations for actions that will bring about assessments consonant with the goal of providing effective science education for all young adolescents.

Chapter II

The Opportunity

During early adolescence humans develop the capacity to think in a way qualitatively different from the thinking typical of students in the early elementary grades. A large proportion of youths, however, do not realize this capacity. Studies done in the United Kingdom with students in comprehensive secondary schools show that only 20 to 30 percent had developed formal operational thinking skills—the thinking skills commonly used in science—although the percentage ranged from 60 to 85 percent in elite secondary schools (Shayer and Adey, 1981). Renner et al. (1976) came to similar conclusions about twelfth-grade students in the United States. Failure to develop these higher order thinking skills places limits on an individual's contributions to society and potential for personal development. Formal education consists of structured experiences and opportunities to reflect on these experiences. Formal education is critical to the realization of the capacity for reasoning and higher order cognition. Also, science, as an important component of formal education at the middle level, can directly support the development of formal operational thinking. But to do so, science education must be designed to take advantage of the early adolescent's cognitive and social development.

The Adolescent and Adolescence: Perspectives

The magnitude of the physiological, cognitive, and social changes that take place in early adolescence, the years from ten through fourteen, is second only to the magnitude of changes that take place in the first eighteen months after birth. The rate of physical growth accelerates, the secondary sexual characteristics develop, and the physiology of the brain changes.

Two of the socio-psychological factors that are distinctly different in these two periods of the human life cycle are the degree of awareness on the part of the young person that physical and psychological changes are taking place, and the distribution of control between the youth and responsible adults. Unlike the eighteen-month-old child, who must be constantly reminded by adult caretakers that it is only so big, adolescents are acutely aware of their rapid growth and the appearance of secondary sexual characteristics. Furthermore, adolescents are embarrassed that others— adults or peers—also have noticed these changes. In addition to the more noticeable physical changes, early adolescents are also developing an awareness of their thinking processes and the powerful new strategies available to them.

The locus of control is also very different in these two periods of development. While the infant exerts independence as it enters early childhood, the adults in the infant's life have the edge in size, reasoning capacity, and control over resources. The adolescent, in contrast, is becoming physically competitive with adults. The adolescent is developing a capacity to reason as an adult, and biological forces compel the adolescent to be independent, although they nevertheless realize that adults are still very much in control.

Majority culture in the United States views adolescence as a traumatic, unpleasant period in life through which young adults must be shepherded as quickly as possible. Adults tend to assume that adolescents find this period in their lives as painful as those around them do, although evidence from research and interviews with adolescents contradicts this assumption (Committee on Adolescence, Group for the Advancement of Psychiatry, 1968; Offer et al., 1981). Whether the "trauma" of adolescence is inevitable and universal or an artifact of particular cultures is an empirical question for which no firm evidence exists. Our stance on this matter—based more on philosophy than science—is that the extent of the "trauma" can be reduced considerably if society provides more support for youth in this period. What is unclear is the kind of support that is best. Adults fear for the safety of adolescents who tend to look outside the home for values and models on which to pattern their behavior. The typical adolescent engages in high-risk behaviors, some as dangerous as using illegal drugs and alcohol, experimenting with sex, or operating motor vehicles irresponsibly. Adolescents often act as if they believed that they were impervious to the dangers of everyday life. Consequently, parents and educators alike contrive ways to protect them. Two strategies for coping with this "dangerous" time are

- Keep the early adolescent busily engaged in desirable activities—studying, sports, art and music lessons—so that neither time nor energy remains for participating in undesirable activities.
- Create an environment—a playpen, if you will—in which the adolescent is prevented from engaging in potentially injurious behaviors. The second strategy, unfortunately all too typical of American education, provides a safe environment that does not promote the development of the cognitive capacities of early adolescents.

The following description of the intellectual development possible during this period is based largely on the work of Jean Piaget, whose observations continue to influence the practice of science education. While the theoretical interpretations and practical implications of Piaget's research have been the subject of considerable debate, his descriptions of the reasoning characteristics of infants, children, and young adults illuminate most current thinking about the learning and teaching of science (Flavell, 1963; Case, 1985). Present-day cognitive researchers who are building on Piaget's work on the evolution of thinking tend to emphasize the continuity rather than the difference between the reasoning stages that Piaget has described (Carey, 1985). Rather than postulating that individuals engage in fundamentally different ways of thinking at different stages of development, these researchers hold that the less-skilled thinkers, regardless of age, know a great deal less about the domain with which they are dealing, and it is this lack of knowledge and understanding of concepts in a particular domain that keeps them from engaging in the more complex reasoning process. The function of formal education, then, is to deepen the students' knowledge and understanding so they can develop the higher order thinking skills described by Piaget. In the next section we summarize Piaget's descriptions of higher order thinking as they are relevant to science education, particularly the development of formal operational thinking.

Formal Operational Thinking

In early adolescence, students begin to display a qualitatively different kind of thinking about the natural world and the individual's place in it than that generally displayed by younger children. Early adolescents acquire more knowledge and a more sophisticated organization of that knowledge, and their intellectual development proceeds to the point at which scientific thinking can be observed. According to Piaget, formal operational thinking represents the highest form of human thought and is characterized by the individual's ability to:

- engage in hypothetical-deductive reasoning,
- engage in propositional reasoning,
- use combinatorial analysis and proportional reasoning,
- reflect on one's own thought processes, and
- consider issues and situations from different perspectives.

Hypothetical-deductive reasoning. The ability to conjecture alternatives to reality and to test systematically the alternatives against available data indicates an individual's ability to use hypothetical-deductive reasoning. This ability entails controlling variables and reasoning from a set of premises. The competencies might be teachable (Linn and Levine, 1976), although they appear to some extent to depend on the formulation of a given problem. Hypothetical-deductive reasoning enables individuals to have thoughts that go beyond the "here and now." Also, these thoughts can influence the adolescent's social and moral cognition.

Propositional reasoning. In contrast to the student in the early elementary grades, who tends to think in concrete, operational terms and mentally manipulate only real objects, the older student who displays formal operational thinking is capable of reasoning using abstract propositions, hypotheses, and quantitative relationships—at least in familiar domains (Flavell, 1963). Hypothetical-deductive and propositional reasoning are the basis for an individual's ability to reason scientifically.

Combinatorial analysis and proportional reasoning. Components of scientific thinking include hypothetical-deductive reasoning, propositional logic, and combinatorial analysis—the component most closely associated with experimental design and data analysis. Tasks that Piaget used to test for hypothetical-deductive reasoning and combinatorial analysis require the student to generate lists of factors that might account for how a physical system functions—for example, the period of a pendulum's swing—and then to determine which factor actually influences the system by testing each factor while holding the other factors constant. Proportional reasoning is a mathematical skill essential to scientific reasoning. A task used to test for this skill involves an object (a stick figure is often used) represented by using two different scales. Lengths of the component parts of one representation are given in some arbitrary unit—the length of an arm in paper clips, for instance. The task is to figure out how long a corresponding part is on the other representation.

Reflective thinking. Awareness and assessment of one's own thinking processes are characteristics of formal operational thinking. This quality of thinking enables students at the middle level to accomplish five tasks:

- describe how they learn best,
- improve their own learning,
- assess the strengths and weaknesses of their problem-solving skills,
- assess the extent to which they understand, and
- assess how well they are meeting the teacher's expectations.

Not only do these and other related skills make it possible for the students to assess their own work; these skills also enable them to improve themselves.

Consideration of issues and situations from different perspectives. The bility to consider issues and situations from different perspectives is characteristic of formal operational thinking. Thus, an adolescent can engage in recursive thought, that is, thinking about the thoughts of others, and contrast sets of perspectives of self and others. Concurrently, young adolescents tend to be egocentric, even as they develop their ability to distinguish between their own concerns and those of others. As their ability to place themselves in a wider social context increases, adolescents begin to see themselves as having a personal and a social destiny (Lipsitz, 1977). Being able to shift perspectives is critical to scientific thinking. For example, spatial reasoning, a particular form of considering different perspectives closely, correlates with scientific achievement. Spatial reasoning implies

- the skills necessary to represent the spatial relationships of objects to each other as they would appear from vantage points other than the one from which the individual is viewing them, and
- the skills necessary to represent how an object would appear from various vantage points or how the object would appear after a linear or rotational transformation.

Another sort of competence in shifting perspective is the ability to consider how others might think about a situation or an issue. The ability to develop an effective scientific argument is dependent on three skills:

- generating possible perspectives that others might take,
- determining which of these a particular individual holds, and
- developing a line of argument to counter or complement alternative points of view.

In traditional Piagetian theory, the logic structures that facilitate spatial reasoning also operate in the ability to understand the perspective of others; more recent research has underlined the importance of context and experience in the ability to shift perspectives.

Instructional implications. As the early adolescent's ability to reason, reflect, and consider other perspectives grows, the educator might be tempted to reduce direct experience with hands-on activities in favor of reading and writing about and discussion of science and technology. Although many students at this age are becoming more adept at abstract thinking, more comfortable using mathematics, and more skilled and practiced in using thinking skills to solve problems, they are nevertheless concrete thinkers most or part of the time. New learning is often elusive, understood at one moment, slipping away at another. Therefore, problem-solving and decision-making skills are best practiced around a concrete, visible, memorable activity or a real experiment, because skills and concepts thus learned can be remembered from a tangible context. Moreover, hands-on experimentation in science provides opportunities for planning, observing, selecting evidence, formulating and ruling out rival interpretations—in short, learning how to impose structure on experience.

Scientific Thinking

Formal operational thinking is a characteristic of scholars in all academic disciplines. It is also a characteristic of the highest levels of social and moral cognition. Piaget's detailed descriptions of formal operational thinking, however are drawn largely from mathematics and the physical sciences.

The nature of scientific thinking. Scientific thinking is the product of formal reasoning strategies operating on a knowledge base. The structure of the knowledge base reflects the nature of the reasoning processes that store information in it. Of particular interest in formal operational thinking are two structural features of the knowledge base that arise from an individual's ability (a) to categorize objects, events, or ideas on the basis of conceptual rather than perceptual features and (b) to understand concepts at a theoretical level. The knowledge base of an individual skilled in formal operational thinking is different from that of someone thinking in concrete terms, because the former is capable of abstract categorization. Concrete operational thinking only requires categorization on the basis of physical attributes —objects on the basis of color, or sounds on the basis of pitch, for example.

Formal operational thinking requires categorizations of objects or symbols using abstract features. Categorizing chemical changes according to reaction type (oxidation-reduction or neutralization), physics problems according to the physical laws that must be applied to solve them, or organisms according to their function in a biological system all present examples of formal operational thinking (Chi et al., 1981).

Individuals thinking in formal operational terms also can understand concepts at higher levels of abstraction than do individuals thinking in concrete operational terms. Concepts can be understood on at least three different levels of abstraction: phenomenological, experimental, and theoretical. At the phenomenological level of abstraction, understanding implies familiarity with the qualitative aspects of phenomena. Density, for example, can be understood in terms of phenomena—objects and substances floating and sinking in liquids and gases. However, understanding at this level does not imply that the individual understands the explanation, only that the individual can completely and accurately describe the phenomena.

At the experimental level of abstraction, understanding density means knowing how to measure volume and mass and, consequently, density. While understanding density at this level involves manipulation of concrete objects, the fact density is a derived quantity, the ratio of two measured quantities (volume and mass), means that understanding density at this level requires proportional reasoning. Formal thinking is also necessary to understand the explanation for floating and sinking phenomena.

At the theoretical level of abstraction, understanding of density implies knowing that density is an intrinsic property of substances, a property that depends upon the mass of the molecules of which the substance is composed and upon the number of molecules in a unit volume of the substance. The knowledge base necessary for formal operational thinking contains integrated information about a concept like density at all three levels of abstraction.

Developing scientific thinking. According to psychological theory, three factors affect intellectual development (including the development of a science-relevant knowledge base and science-related skills): physiological maturation, interaction with the natural world, and social experience. Developmental psychologists tend to downplay the influence of formal educational experiences (a type of social experience) in the development of formal operational thinking. Other psychologists—the neo-Piagetians, for example—admit to the effects of formal education on the acquisition of formal thinking. For this reason, science educators stress the importance of hands-on science work linked to the student's experiences and accompanied by discussion among groups of students as well as with the teacher. They stress the development of conceptual schema through effective education rather than the physiological development of logical structures. Our report is predicated on the premise that this sort of well-conceived school science can contribute to the attainment of formal operational thinking.

Creating educational environments in which early adolescents can capitalize on their expanding capacities requires thinking through the relationship between development and learning. The way in which the relationship between development and learning is construed influences the nature of the school science experience. In much of the educational literature, the distinction between learning and development is blurred and the relationship between them subject to different interpretations. One interpretation is that development of formal operational thinking occurs independently of formal education. According to this interpretation, cognitive developmental level is a critical factor limiting the sophistication of the subject matter that can be learned. This implies that the cognitive demands of learning science subject matter should not exceed the developmental level of the learner.

Another, more constructive view is that learning contributes to cognitive development. According to this interpretation, experiences with the natural world that a child interprets in a social environment contribute in small increments to the child's knowledge base and repertoire of thinking skills. When these experiences are concurrent with the appropriate physiological maturation, one observes the "quantum leaps" in thinking that characterize the transition from one level of cognitive development to another. In practice, based on this interpretation, subject matter is selected for its contribution to the development of formal operational thinking.

The Goals of School Science and Formal Operational Thinking

When the goals of school science are stated in terms of the characteristics of the successful science student, the close correspondence to the characteristics of formal operational thinking is evident. Both the ideal product of twelve years of school science and individuals skilled in formal operational thinking can

- understand scientific concepts, principles, laws, and theories;
- criticize the design of experiments as well as design experiments; and
- understand the sociology of the development of scientific knowledge.

Furthermore, being able to learn on one's own and to assess one's understanding and progress toward achieving a goal also are desired outcomes of school science. Although the goals of science education correspond significantly to the operational definition of formal operational thinking, one critical difference, which has implications for practice in science education, centers around "knowing" something and being able to "figure it out." Piaget operationally defines stages in the development of reasoning skills in terms of the ability to respond to an unfamiliar task, no matter what the domain. Thus, in his view, a correct response, which would include both the "correct" answer and justification for that answer, indicates that the reasoning structures necessary to analyze the task "logically" are available to the student, quite apart from exposure to the subject matter. In contrast, the assumptions underlying the goals of school science start with domain knowledge; that is, the student will know the right answer and be able to justify it after exposure to the relevant subject matter. Even when goals for science education refer to application of knowledge and reasoning skills in unfamiliar situations, the new situations are generally domain specific, that is, they entail scientific knowledge and reasoning skills applicable to academic, personal, and civic problems related to science. Transfer to the domain of analyzing historical exposition, for example, would be considered far transfer, and not expected as an explicit outcome of science education.

Issues and Dilemmas

As one contemplates the possibility of "detraumatizing" adolescence by providing an environment in which the developmental tasks of adolescence can be achieved, one must recognize the impediments to the realization of every student's potential: lack of knowledge about the detailed nature of the formal experiences that help to enhance the development of formal thinking; institutional and structural issues that include teacher preparation and beliefs (see the Center's companion report on teachers and the teaching context at the middle level); and resources—what society is willing to invest to ensure that all but the severely mentally disabled develop formal thinking. Among the considerations with regard to resources is the relative importance of the development of intellectual skills compared to the many other developmental tasks of adolescence. This particular issue creates dilemmas for educational practice in general and for science education in particular. Some dilemmas are philosophical: How does formal operational thinking contribute to the valued outcomes of school science? How does school science contribute to the development of formal operational thinking? Does society value formal operational thinking? If so, how much? To what extent is society willing to devote its resources to achieving formal operational thinking in all youths? Some of the dilemmas also have a theoretical component: Are all "normal" youths capable of becoming formal operational thinkers? What is known about the extent to which the development of formal operational thinking can be facilitated? If development can be facilitated, how is that best accomplished? Is the development of formal operational thinking accomplished best through the study of science? If so, what should be the nature of the science experience? How do experiences with the natural world influence the development of formal thought? Do educational experiences that develop understanding of science concepts at a theoretical level and the ability to design a valid experiment create a formal operational thinker or, rather, is it the case that only the formal operational thinker can come to understand science? What in all this is the role of social interaction? Developers of science curricula and instruction need to consider careful responses to these questions, as well as defining the optimal conditions under which true scientific thinking develops. In the next two chapters, the goals of science education at the middle level (chapter 3) and recommendations for science instruction (chapter 4) are discussed. In these two chapters, special attention is paid to the growing capacities of young adolescents. Also, the recommendations are contrasted to actual current practice.

Chapter III

Goals for Science Education and the Assessment Challenge

Scientific and technologic literacy for all citizens stands high on the list of educational needs for the year 2000 and beyond (National Commission on Excellence in Education, 1983; National Science Board Commission on Precollege Education in Mathematics, Science and Technology, 1983; Task Force on Education for Economic Growth, 1983; Twentieth Century Fund, 1983; however, for a dissenting view, see Shamos, 1988). To summarize the arguments made by advocates of science education: Not only will the economy require an increasing number of scientifically and technically trained professionals and support personnel, but most production and service jobs will require a modicum of quantitative and technical skills (Botkin et al., 1984; but see Levin and Rumberger, 1983, for counter-arguments). Moreover, an increasingly complex interlinking of the man-made and natural worlds makes it important for people to understand the basic parameters of both these worlds and their functioning, so that they can make effective personal and civic decisions. Recent reports have interpreted in some detail the meaning of scientific and mathematical literacy with respect to student learning and proficiency in these fields (American Association for the Advancement of Science, 1989; Mathematical Sciences Education Board, 1989).

The period of early adolescence can be an exciting time in science education. Middle-level science instruction must bridge the introduction of science as a set of accessible activities in the elementary years and science as a sophisticated form of intellectual inquiry in high school. Children exposed to good science in the elementary school years have seen something of the ways in which scientists approach problems, pose questions, and collect and organize information, but they probably have seen little of the formal, systematic knowledge structures that characterize mature scientific disciplines. In the elementary years the students lack intellectual maturity, which limits their ability to work with abstract formal systems. Also, they are just beginning to develop the "tool skills," especially mathematical understandings and symbol systems, necessary to work with abstract scientific concepts. Science instruction in high school is grounded in the scientific disciplines. It is formal, rigorous, and often quantitative. Therefore, middle-level science instruction divides elementary science and high school science by introducing the power, excitement, and utility of formal scientific systems without communicating to children that real science is only comprehensible to the brightest students, the mathematically precocious, or boys.

Education of young adolescents has the dual purposes of providing for their continued personal development and fulfilling the aspirations of society. Early in this century, the literature on junior high schools and in recent decades the literature on middle schools has continuously emphasized the goal of personal development for early adolescent students. While personal development as a goal is appropriate, middle-level educators should not lose sight of the second goal, that of contributing to the society in which the adolescent lives. Putting it more succinctly, the aims of science education in the middle years are to develop the student's capacity to

- think scientifically and use the tools and strategies of science, and to
- apply science knowledge and skills in addressing individual and societal problems.

These broad aims lead to several more specific goals.

Goals for Science Education for Early Adolescents

The goals as spelled out represent general directions. While all students may not attain all goals with equal proficiency and understanding, all students should develop some proficiency and understanding. The goals stated for the elementary years in the Center's earlier reports (Raizen et al., 1989; Loucks-Horsley et al., 1989; Bybee et al., 1989) share common elements with those stated here, as the Center sees articulation of subject matter drawn from science education across grade levels as a critical element of reform. In particular, we continue to emphasize the importance of teaching both science and technology and connections between them as well as the importance of engaging students in activities relating to science and technology. The variations between goals for elementary and middle levels are based primarily on the student's developing capacities, as described in the preceding chapter.

Goal 1: Science and technology education should develop adolescents' ability to identify and clarify questions and problems about the world.

Young adolescents are first and foremost interested in questions and problems that relate directly to them. Constructing a middle-level curriculum could easily begin with such questions as what is normal? Why do organisms behave the way they do? How are things made? Why do things change? What are the relationships among things? These questions are intentionally ambiguous. Young adolescents seldom state questions with immediate personal connections—Why do I change? Am I normal?—although these questions are probably closer to their interests. The point here is to begin with questions and problems that have meaning for adolescents, rather than with concepts and skills that have scientific and technologic significance but seem abstract and removed from life. Although adolescents are likely to have many questions about themselves and their surroundings that are related to science and technology, they may not see this connection until they pursue their questions in greater depth. Asking questions and identifying problems are the first steps in scientific inquiry and technologic problem solving. It is appropriate, therefore, to introduce students to science and technology education in response to their questions and problems.

Goal 2: Science and technology education should broaden adolescents' operational and critical thinking skills for answering questions, solving problems, and making decisions.

As they develop explanations and solve problems, scientists and engineers use cognitive processes and intellectual models that differ from those that people commonly use. Observation, experimentation, and construction of theories in science, as well as consideration of cost, risk, and benefit in technology, are examples of the processes included in this goal. Adolescents should learn what and how scientists and engineers think and why they think the way they do. Students need an introduction to the intellectual rigor and demands of scientific inquiry and technologic problem solving—the need for evidence, the use of logic and creativity. Learning to formulate sound and coherent explanations and developing a nonauthoritarian, skeptical posture are also important. In addition, students need to acquire the social and communication skills appropriate for doing collaborative science activities. This goal connects to other general aims important to middle-level education. Among these other aims are the development of adolescents' operational and critical thinking skills and their physical, social, and emotional capabilities.

Goal 3: Science and technology education should develop adolescents' knowledge base.

Knowledge must be a central concern of science and technology education. Traditionally, the science curriculum (including that designed for young adolescents) has consisted of facts, information, and concepts that represent the life, earth, and physical science disciplines. The criterion for selection and inclusion of subjects has been that the curriculum should represent the accumulated information within each discipline. The task of the teacher has been to present the information. Tests were used to determine what information the students had retained.

The Center recognizes the importance of adolescents' ability to acquire and apply knowledge in personal and social contexts, and the Center's goals reflect this. We recognize the dynamic nature of science and technology and thus recommend presentation of scientific and technologic knowledge as proposed explanations and proposed solutions. The emphasis should not be on trivial facts and isolated information. Rather, the emphasis should be on the acquisition of a knowledge base, the acquisition of concepts that unify disciplines within the sciences, and an understanding of technology. The focus of middle-level education should be on knowledge, concepts, and procedures that have the widest potential for use in one's life, and that help the individual meet family and societal responsibilities. Since learning often proceeds from specific cases to generalizations, the emphasis should be reaching a deep, rich. and rewarding appreciation and understanding of a relatively small number of carefully chosen phenomena that provide opportunity for broadly applicable methods of inquiry.

Goal 4: Science and technology education should develop adolescents' understanding of the history and nature of science and technology.

Adolescents also need to understand science and technology as cultural and social activities. Historical and present-day examples can vividly illustrate how society and culture influence science and technology and how technology and science influence culture and society. Thus, the social context in which scientific explanations and technologic solutions are presented determines their shape. By the same token, some scientific and technologic events have historical significance and have helped shape Western culture. Consider, for example, the revolutions of Nicolaus Copernicus, Isaac Newton, and Albert Einstein; the contributions of Charles Darwin, Charles Lyell, and James Watson and Francis Crick to the current understanding of the processes of evolution; and the roles of such individuals as James Watt in the Industrial Revolution. These developments have had significance beyond their scientific content and technologic products; indeed, they have changed world culture (Bybee, 1990). There is another important reason for spending time on the history and development of science and technology. Students' conceptual understanding of the world sometimes appears to parallel that of history; for example, many students have an Aristotelian view of nature. Presenting different historical perspectives, while affirming that others have perceived the world the way some people do now, can serve to challenge problematical conceptions that students might hold and provide structures for reformulating their explanations. Adolescents should begin developing an understanding of the nature of science and technology. They should see science as a particular way of knowing about the world, and technology as a way people adapt to their environment. How do the sciences and technology advance? What constitutes a valid scientific explanation? How is science different from other ways of knowing, such as history, literature, or religion? Is technologic problem solving different from other forms of problem solving? Science for All Americans (American Association for the Advancement of Science, 1989) provides examples that further clarify both what this goal includes and the conception of science and technology that we hold at the Center. The adolescent should understand that science assumes the world is understandable, that scientific explanations are durable but subject to change, and that science cannot explain all things. Concerning technology, adolescents should understand the interactions between science and technology, that technologic problem solving involves design under constraint, that technology involves control, that technologies can have unintended consequences, and that technologic systems can fail.

Goal 5: Science and technology education should advance adolescents' understanding of the limits and possibilities of science and technology in explaining the natural world and solving human problems.

Science and technology directly relate to contemporary American life. They serve as agents for social change, and, in turn, they are changed by society. Individuals and nations are increasingly being asked to make decisions that influence the quality of life. Understanding the limits and possibilities of science and technology has a direct bearing on the goals for general education in the sciences. This goal encompasses the need to develop personal decision-making abilities. This goal also expands the adolescent's potential for meaningful work and careers and cultivates the adolescent's citizenship responsibilities. These goals represent an integration of our conception of science and technology and the major orientation of middle-level education. The task is to see that young adolescents develop, in a personal and social context, the most complete and accurate understanding of science and technology that is possible at their age and stage of development. Not only is it important for them to understand the processes, the concepts, the history, and the nature of science and technology; it is equally important that these adolescents recognize what science and technology can and cannot do, what they are and what they are not, and how they do and do not influence individuals and society.

The Assessment Challenge

The goals of middle-level science instruction have implications for assessment. The classroom tests teachers develop and use both express their own understandings and also communicate to their students what is important to learn from science instruction. 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. Even in elementary school, children can use classroom tests to help them understand what they should be learning. But during the middle years, with the growth of the capacity for abstract thought and especially for reflection about one's own learning, the messages students receive from classroom tests assume increasing importance.

Classroom tests communicate not only the character of the teacher's intended learning outcomes, but also the level at which mastery is expected. If standards and expectations are set too low (perhaps in a well-intentioned but misguided effort to accommodate the special needs, diversity, rapid physical growth, or presumed cerebral dormancy of young adolescents), the students may infer that they are not really expected to master difficult concepts. Low standards and expectations might retard learning, and consequently the transition to high school science will be needlessly difficult, or, worse, might never occur. Yet, if standards are set too high, the effect can be to reinforce the unfortunate stereotype that science is too difficult for most students.

Early adolescence is a time for exploration and experimentation, a time when students may test their interests and capabilities in a variety of content areas and form enduring impressions of different subject matters. Although career choices are rarely established until much later, the impressions early adolescents form, and the decisions that they, their parents, teachers, and counselors make about tracking and courses to take in high school, profoundly affect their options for postsecondary education and their future vocations. Middle-level students' understanding of their mathematical and scientific abilities, and their decisions about courses they take in these areas, are far too important to be left to chance. In particular, the middle-level science program should acquaint youths with the wonder and excitement of formal science, and aid them in reaching an honest, but optimistic, assessment of their own capabilities to profit from future scientific study in high school. Sound classroom testing practices, including fair and consistent standards and expectations, can further these ends.

If science and technology education successfully address these goals, they will foster three types of outcomes: increased factual and conceptual knowledge; increased laboratory, thinking, and social skills; and increased disposition to apply one's knowledge and skills to unfamiliar situations. Increased learning in these three areas is a prerequisite both for scientific literacy and for preparing to enter scientific or technical careers. The assessment challenge is how to adequately probe the students' competencies in all three of these areas and how to avoid certain adverse effects of testing.

Science Knowledge

The knowledge category includes the "what" of science and technology—knowing facts about the natural and man-made worlds, for example, understanding that sounds are patterns of motion and that the sounds of instruments or one's voice vary as vibrations vary; knowing that rivers are part of the water cycle and knowing how their power is translated into electric energy; and understanding the functions of primate groupings and social interactions. Also included in the knowledge category are the concepts, principles, laws, and theories that scientists use to explain, for example, how vibrating strings relate to sound, how heat energy from the sun drives the hydrologic cycle, and the role of communication among primates. Beyond facts about the natural world, the theoretical knowledge used to compose explanations for these facts, and an understanding of how factual and conceptual knowledge are applied appropriately, the science knowledge category also—as noted in the discussion of goals— includes knowledge about the scientific and technologic enterprises: their history, methods, philosophy, and values and their influence on human existence.

Assessing science knowledge. The first task in assessing the science knowledge acquired by students is deciding which categories of that knowledge are to be probed, and what knowledge within each category should be represented on a test. Once these decisions have been made, testing of factual and theoretical knowledge and knowledge about the scientific and technologic enterprises can be carried out with relative ease, using paper and pencil. Often, short-answer or multiple-choice items are used. This type of assessment format allows a single person to administer the test in group settings; hence, the exercises making up the assessment can be given to a large number of individuals. Because of the relative ease and efficiency of paper-and-pencil tests, particularly those—like multiple-choice—that can be scored by machines, most tests intended for monitoring purposes, that is, providing national, state, or district-wide information on student achievement, take this format (for example, state-mandated tests, commercially available standardized tests, and tests used by the National AssessPment of Educational Progress and in international comparisons). Unfortunately, all too often, multiple choice items test recall of unconnected bits of information, thereby conveying a distorted message about the nature of science. Assessments, however, need not be limited to this form of test. Teachers, in particular, have other strategies available to them. They can design essay questions and review written and oral reports. They can also use more informal methods for gauging their students science knowledge and embed assessment of what knowledge has been learned in more holistic assessment strategies, as described in later chapters.

Interpreting tests of science knowledge. Tests intended to assess science knowledge have a second important characteristic. For well-constructed exercises, the responses can be interpreted with reasonable certainty. A correct response indicates that the individual either knew the information required for the answer, or was able to figure it out using information provided as part of the question. (Of course, it could also be a lucky guess.) Determining the correctness of the response does not need to take into account the thinking skills the individual might have applied in comprehending the written item, in retrieving the fact from memory, in reasoning from the information in the item to the correct answer, or in eliminating incorrect responses. In other words, the concern is not with the means individuals may have used to access the information or the reasons for their conclusions, but only in whether or not they have presented the correct information. Hence, responses to factual items are relatively straightforward to interpret, whereas interpretation becomes increasingly more difficult for items intended to test skills and dispositions.

Skills

Meeting goals beyond knowledge acquisition entails developing four interrelated types of skills: practical laboratory skills, scientific intellectual skills, generic (formal and practical) thinking skills, and social skills.

Assessing practical laboratory skills. Skills development for the middle-level students should build upon their experiences in the elementary years. For example, in their early years, children learn to measure mass by using a double-pan balance to compare different objects with uniform masses, such as paper clips, thumbtacks, weights. At the middle level, they should be able to move on and discuss the accuracy of their measurements and possible sources of error. By the time the students reach high school, they should understand the mechanics of measuring mass (and the connected uncertainties) well enough that they can use a digital balance, and thereby save weighing time and focus instead on interpreting the usefulness or significance of the data.

Middle-level students should be able to measure length, volume, mass, time, and temperature, using instruments capably and quantitative data comfortably. They should also be able to use a microcomputer independently to enter, store, and retrieve data and to simulate experimental conditions.

Assessing laboratory and computer skills requires laboratory equipment, materials, and computers. This sort of assessment distinguishes between knowing about how to do something, which can be probed with paper-and-pencil tests, and having the competence to do something, which cannot. To assess the latter, assessment techniques need to closely match the ability to carry out a given scientific procedure or design task. Obviously, this type of assessment is more difficult to administer and score and requires more material resources than do paper-and-pencil assessments. Nevertheless, NAEP conducted a successful pilot study in this area (National Assessment of Educational Progress, 1987), and Connecticut, New York, and California also are now experimenting with incorporating performance tasks in their science assessments. In science classrooms that include science activities as a regular part of instruction, teachers have many opportunities to observe these skills in action, with the added benefit of being able to do corrective teaching as deficiencies manifest themselves.

At the middle level, observing, classifying, measuring, and other laboratory skills useful for gathering information will recede from prominence, being no longer ends in themselves. This aids the assessment problem to some extent, as the students will be able to record in a journal or notebook observations and data that can be easily monitored by a teacher. The importance of keeping records in accessible forms can be made clear to students by presenting challenging and meaningful problems whose solutions depend, at least in part, on the accuracy of measurements made over time and the careful recording of changes in experimental conditions. Examples of relevant activities are given in chapter 5.

Assessing scientific intellectual skills. These skills include the ability to generate a hypothesis; to design an experiment that is a valid test of a hypothesis; and to collect, reduce, present, interpret, and analyze data (Frederiksen and Ward, 1978). Skills related to technology include the design and building of artifacts intended to perform a specified function. The combination of intellectual skills relevant to science and technology also includes procedural knowledge—knowing "how" to apply the "what," or the factual and conceptual knowledge and laboratory skills one has acquired. Procedural knowledge is the key to addressing unfamiliar scientific questions or operational problems that may arise in the course of one's work, for students as well as for working scientists, engineers, or technicians.

The developing ability of early adolescents to reason deductively; to remove themselves a bit from their experience, and to see how things might look to another observer should enable them to be more flexible and open minded as they examine data. A middle-level student can be expected to understand readily that the failure to be able to report a result must be explained some way. Middle-level students should also understand that they must present the data as they observe them, not as they think the data ought to be. The students should continue their practice from elementary school of estimating and using the words greater, less than, the same as, and they should now use their estimates to question whether measurements or calculations are accurate and reliable. They should be developing enough self-confidence to report what they actually see and to understand the role that honesty plays in the scientific enterprise. By the end of their middle-level education, they should be able to criticize their own work and monitor their own thinking.

Assessing the intellectual skills of science—hypothesis generation, experimental design, data collection, data analysis, and data interpretation—introduces additional confounding factors. Scientific intellectual skills integrate a complex variety of generic thinking skills with the ability to select and perform appropriate practical laboratory skills. The following example starts out with a measurement problem, but quickly expands to a potential test of scientific thinking skills.

In most tests and assessment exercises, scientific intellectual skills are assumed to be generic skills that the student should be able to use in any scientific context. However, many testing experts disagree with this assumption, and they argue that familiarity with the context of the assessment exercise and the science knowledge available to the student are more important factors in the ability to perform an exercise than the scientific intellectual skills. It is certainly conceivable that a student could succeed by using either science and context knowledge or scientific intellectual skills. This makes interpreting a student's performance quite difficult, particularly if the test is externally designed and scored.

Assessing generic thinking skills. Included in this category are problem-solving skills and quantitative, logical, and analogical reasoning. These are component skills of scientific intellectual skills as well as intellectual skills associated with other disciplines (Nickerson, 1988). The problems of designing an assessment exercise and interpreting a student's performance, that we described previously, severely affect the assessment of generic thinking skills. The difficulty lies in interpreting the behavior an assessment exercise elicits, and, again, this interpretation is especially difficult when the assessment is out of context.

When a student performs an exercise and gives an answer, one has no way of knowing the mental processes and knowledge the student used to arrive at the answer. For example, if a student is given a description of a physical event and asked to explain it, the student's correct explanation may be the result of simply being familiar with the situation and knowing the explanation for it. Alternatively, the student might be unfamiliar with the situation, yet be able to recognize that a particular scientific principle applies to the situation. The student can then apply the principle with the appropriate reasoning skills and come to a correct answer. Another possibility is that the student uses incorrect information when developing an explanation, but uses a correct scientific principle and rules of logic while coming to an incorrect answer. On the basis of the answer alone, the examiner cannot possibly know whether the performance represents recall; logical application of correct factual information, a scientific principle, and rules of logic; or right thinking with wrong information.

Assessing social skills. Humans are generally social beings, and young adolescents are particularly so. Consequently, the middle years of education are the best time to complement the students' growing facility in general thinking, science thinking, and laboratory work with skills for working effectively in groups. The need for developing social skills in science grows not only out of an interest in improving the students' social skills per se—although that is an important goal—but also out of the close connections among social skills, learning, and science. Social skills, such as listening carefully and respectfully, exchanging ideas and information, welcoming a diverse array of approaches to solving problems, and acknowledging that a variety of "right" answers (or reasonable interpretations) are possible, are some of the skills the students require. Such skills enable the students to grapple actively and productively with complex knowledge and ambiguous problems. Given a problem or task that is within their capability to solve, students who are working together can be expected to take on challenges that require perseverance and commitment. Moreover, when young adolescents employ their developing skills in science learning and do so in working groups, the classroom becomes a replication of a community of science scholars pursuing scientific knowledge as a social activity. Thus, the students begin to learn about the culture of science and to learn skills valued in the workplace, where the application of science usually proceeds through teams working together.

Assessing social skills is difficult. Written communication, such as laboratory reports, reports on the design and construction of artifacts, or essays on a particular scientific or technologic development, can be evaluated both for their scientific content and the quality and appropriateness of language use. But most communication skills involve direct interaction with other persons, and these skills are best observed during group work. Time for such observation can be short and interpretation difficult, particularly for science teachers in schools with departmentalized structures who may see as many as 125 students in the course of a school day. At the classroom level, spot diagnosis of problems related to social skills (or lack thereof) during normal monitoring of the classroom process may suffice. With respect to large-scale assessments, the need for highly trained observers would make valid and reliable assessments of social skills expensive and feasible, perhaps, only for small subsamples.

Assessing dispositions and habits of mind. Acquiring a knowledge base in science and developing the skills to apply the relevant knowledge to academic problems in school are necessary, but not sufficient in themselves. Unless science education inclines one to apply scientific knowledge and skills to new situations in one's work, daily life, and when one makes personal and social decisions, neither the goal of developing productive science professionals nor the goal of scientific literacy for all citizens will have been achieved.

Science education in the middle years should continue to address dispositions and habits of mind as much as development of content and skills. When planning curriculum, instruction, and assessments, one must take into account the assumptions and attitudes students have about the nature of science and technology and about themselves in relation to science and technology.

Goals for developing scientific habits of mind or scientific attitudes do not change at the middle level per se, but the emphasis should take advantage of the interests, needs, and strengths of students as they move through adolescence. Some of the most important scientific attitudes and dispositions that students should come to understand and practice are

- **Desiring knowledge.** The curiosity and desire to know and understand the world should have been nurtured in the elementary years and should be sustained. The questions a student is asked should increase in complexity. For example, whereas five- to ten-year-old students find physical and chemical changes interesting in and of themselves, a teacher at the middle level might have to plan a discrepant event, such as boiling water in a flask with ice cubes around it, to stimulate questioning about the way the world works.
- **Being skeptical.** Getting students to question authoritarian statements and increasing their confidence in independent thinking should be further developed at the middle level. Because students in these years will become increasingly able to understand another point of view, trade-offs, risks, and benefits; and because they will be increasingly able to take responsibility for their own health and safety, this is an important attitude to cultivate and should greatly interest young adolescents.
- **Relying on data and relying on reason.** Development during the middle years should enable the students to become increasingly able to collect and organize data and to use data to test ideas. Adolescents' ability to reason, to stand back and take another perspective, are strengths that will help them develop this habit of mind.
- **Accepting ambiguity.** Although students often hope for a "right" answer or clear solution or outcome to experiments or problems, in practice data in scientific and technologic problem solving are often ambiguous. The notion that conclusions in science are tentative is a habit of mind that ought to be more clearly understood by students as they move through middle-level science education.

These habits of mind—desiring knowledge, being skeptical, accepting ambiguity, and relying on sound information and reasoning with it—characterize the disposition toward continued engagement with science.

Making judgments about the extent to which students have acquired the habits of mind that dispose them to apply scientific knowledge and skills outside the formal classroom setting adds yet another level of complexity to assessment. One might attempt to assess disposition by the use of a self-report that is, describing situations and asking individuals to indicate whether or not they would take a "scientific" approach to analyze them. This method has not yielded particularly trustworthy information (Gardner, 1975; Munby, 1983; Murnane and Raizen, 1988). A more appropriate method is to observe individuals and determine whether they use a scientific approach to personal and civic problems. This method requires extensive resources and, even when attempted, the direct observations that result are difficult to interpret. Does failing to take a scientific approach indicate that the person has the inclination but not the requisite skills? The requisite skills but not the inclination? Neither? In addition, context has a profound influence on behavior. For example, not being scientific in approach in one situation might be an indication that either the skills or the inclination are not in place. An alternative interpretation is that the person did not deem the scientific approach or the solution suggested by that approach appropriate for that particular situation, but would demonstrate a scientific inclination in other situations.

A possible way around these dilemmas is to measure observable behaviors, for example, the students' interest in voluntarily undertaking science activities beyond prescribed classroom work (and subsequent enrollment in science electives), the students' self-monitoring of their work and their monitoring of peers. Teachers might add observations on these behaviors to the records they keep on their students. Conceivably, some structured performance tasks might also provide opportunity for observing these behaviors, particularly if the tasks call for sustained work. At this stage of understanding, however, much more research is needed in this area to identify behaviors that are reliable indicators of future willingness to continue an engagement with science and continue to apply one's science knowledge and skills.

Many assessments have included measures of attitudes about science as a way of getting at present and future scientific dispositions and habits of mind. These assessments ask whether the students like science lessons or their science teachers, whether they value the contributions of science to society, and whether they have plans for science careers (Hueftle et al., 1983; Mullis and Jenkins, 1988). These sorts of attitude measures have two kinds of shortcomings: results are often paradoxical (for example, "I like my science teacher" but, from the same student, "science class is boringS) and difficult to make sense of (Munby, 1983). Further, the linkages between attitudes about science— even if they could be better assessed—and achievement, let alone later dispositions to engage with and use science knowledge and skills, are open to question (Willson, 1983). Other proxy variables that researchers have used to assess scientific dispositions and habits of mind include impulsivity, attitude toward one's own competence, and fair-mindedness (Nickerson, 1988; Rowe, 1979). Further work will have to be done before the proxy variables can be linked with any confidence to scientific dispositions and habits of mind, including the disposition to apply science knowledge and skills.

The Effects of Age

Age and its correlate, level of cognitive development, is another confounding factor in science assessment. Performance on a problem- solving exercise for a ten-year-old might well be recall of information for the thirteen- year-old. Also, the thirteen-year-old will be able to bring a greater wealth of experience to the exercise. Moreover, the relevant experiences available to one youngster might be very different from those available to another who grows up in a different environment. For example, there is evidence that girls, even at an early age, have different exposure than do boys to certain experiences—fixing simple electrical or mechanical things, playing with motor-driven toys, building tree houses, using scientific equipment—relevant to solving some science problems (Mullis and Jenkins, 1988). As age is easily established, it can be factored into interpretations of assessment of performance, but the role of experience is difficult to take into account unless an assessment specifically collects relevant background information, as does NAEP (Hueftle et al., 1983; Mullis and Jenkins, 1988).

Learning over Time

The problems inherent in assessing complex learning outcomes can be analyzed in a more general fashion. In an article in the *New Directions in Measurement* series several years ago, Snow (1980) described a "continuum of *referent generality*" in both aptitude and achievement measurement. Referent generality refers roughly to the range of situations to which a given aptitude or achievement pertains. At the highest level, there might be aptitudes like general mental ability (RIQS) or the kind of broad, complex developed achievement measured by the SAT. At the lowest level, there might be aptitudes like "speed of response time" or achievements like "two-column subtraction with borrowing." Important science learning outcomes are likely to be higher in referent generality than narrower learning outcomes. Examples are students' understandings of scientific method or of such higher level knowledge as the relationships between structure and function, the meaning of scale, or the concept of systems.

Outcomes higher in referent generality are harder to teach directly, because they must be visited time and again, in a range of contexts, using different materials and different illustrations. They are harder to assess because they are less closely tied to any particular learning activity. The problem is how to assess understanding of the broad organizing principles, the inquiry approaches, and the ways of knowing that characterize science in the context of a particular learning unit, given that these understandings may take years to develop. The problem is not unique to science, nor is it solved in other content areas.


A COMPARISON

Chapter IV

The Context of Science Education in the Middle Years

Good science education must be supported by good assessment, and good assessment is possible only in the context of good science education. Therefore, before discussing strategies that can respond to the assessment challenges posed by the goals and outcomes outlined in the previous chapter, we need to consider how schools and science programs address these goals and outcomes. We begin by briefly outlining the broad context: a widely held and, among many, a deeply felt set of beliefs about the special needs of young adolescents and the role of schools and science programs in accommodating these needs. We then characterize what actually takes place for most students at the middle level: school organizational arrangements, science curricula, science teachers and their working conditions. We note how this context can shape the students' access to science and perhaps influence their attitudes toward and learning of science. We conclude by discussing some of the implications for assessing student outcomes and key features of science programs.

What Is Different about Middle-level Education?

> Middle-level schools—junior high, intermediate, and middle schools—are potentially society's most powerful force to recapture millions of youth adrift and to help every young person thrive during early adolescence. Yet all too often these schools exacerbate the problems of young adolescents. A volatile mismatch exists between the organization and curriculum of middle-level schools and the intellectual and emotional needs of young adolescents. Caught in a vortex of changing demands, many youths' engagement in learning diminishes, and their rates of alienation, substance abuse, absenteeism, and dropping out of school begin to rise.

This assertion of the Carnegie Task Force on Education of Young Adolescents in its 1989 report, Turning Points: Preparing American Youth for the 21st Century, probably strikes those unfamiliar with the literature on young adolescents and middle-level education as unusually strong, even startling. Yet Carnegie's current claim might be little more than the most recent expression of an eighty-year-old effort to

40

acknowledge the unique and challenging needs of ten- to fourteen-year-olds, and the general failure of schools to meet them.

Beliefs about Middle-level Students and What They Need

Long before there was an understanding of the developing cognitive competencies of students in the middle years (discussed in chapter 2), psychologists and educators called attention to the significant and often tumultuous physical, social, and emotional changes in young adolescents. They also promoted the notion that schools should better accommodate the special needs of this age group. They were not without influence. As early as 1909, changes in school organization that separated grades seven, eight, and nine from the later high school years stemmed largely from G. Stanley Hall's (1905) work on the psychology of adolescence. And interest in the special characteristics of young adolescents and what these characteristics imply for school programs has continued throughout the century. A decade ago, for example, the National Society for the Study of Education published *Toward Adolescence: The Middle School Years* (Johnson, 1980). A theme running throughout its chapters is that students in the middle school years are different: they have a high degree of intellectual curiosity; a wide range of skills, interests and abilities; and they prefer active involvement to passive learning activities. Not only were young adolescents found to be different from their older and younger schoolmates, they also were observed to differ considerably from one another—in physical, social, emotional, and cognitive development (Maurice, 1980).

Two syntheses of writings on the special mission and functioning of junior high schools, one in 1940 and the other in 1970, show that throughout the past fifty years, educators have been remarkably consistent in the educational implications they draw from these characteristics of students (Gruhn and Douglas, 1971). In 1987, Hurd summarized the most common recommendations for schools designed to serve young adolescents:

1. **Integration.** Learning experiences should be integrated "into effective and wholesome pupil behavior" as well as link the subjects in the curriculum.
2. **Exploration.** Schools should lead students to discover and explore their own interests, abilities and skills and provide opportunities to include "cultural, social, civic, avocational and recreational interests" as a basis for vocational decisions.
3. **Guidance.** Assistance should be provided to enable students to make intelligent educational and vocational decisions and wholesome social and personal choices.
4. **Differentiation.** Opportunities should be provided that accommodate students of different backgrounds, interests, abilities, and needs.

5. **Socialization.** Learning experiences should be included that will enable students as citizens to participate in and contribute to this country's democratic society.

6. **Articulation.** Schools should provide students with help in acquiring the backgrounds and skills that will help them succeed in later education and adult life.

Recent changes in the social milieu of adolescents have not altered these long-standing beliefs about what schools should provide students at the middle level. Several recent reports have called for programs that focus on the physical, social, and emotional needs of adolescents as well as on their academic learning. These reports also recommend programs that provide integrated curricula, exploratory experiences, and opportunities for close relationships with adults and peers (for example, see the National Middle School Association's report, This We Believe [1982]; the report by the National Association of Secondary School Principals, An Agenda for Excellence at the Middle School Level [1985]; and the report of the Superintendent's Middle-grade Task Force, Caught in the Middle: Educational Reform for Young Adolescents in California Public Schools [1987]). The one prominent recent addition to this long-standing agenda is the goal of helping students learn how to learn and think. Most likely, this new program goal follows from more recent understandings of adolescents' cognitive development.

While the perceived importance of academic preparation (versus meeting the students' developmental needs) has waxed and waned throughout the century, most advocates for ten- to fourteen-year-olds seem to want both, and most believe that both are possible with an integrated, exploratory, and flexible middle-level program. For example, the National Middle School Association (1982:10) asserts that

> The curriculum must carefully balance academic goals and other human development needs. A middle school cannot succeed in fulfilling its educational responsibilities if it ignores non-cognitive objectives. Indeed, it cannot succeed in fulfilling its cognitive objectives if it does not recognize the interrelated affective goals.

To achieve this balance, the association recommends that schools provide a range of organizational arrangements, for example, block scheduling, multi-age grouping, and alternative schedules; varied instructional strategies with an emphasis on small-group methods, peer interaction, independence, and experimentation; a full exploratory program, for example, high-interest, short-term lessons and units, controlled student choice, mini-courses, special-interest activities, independent study projects; comprehensive peer and adult counseling; consideration for the wide variation in the progress that students make; evaluation that emphasizes individual uniqueness; interdisciplinary curriculum planning teams; and a family-like school atmosphere.

Similarly, the recent Carnegie Task Force on Education of Young Adolescents (1989) calls for specific changes in the organization of middle schools and their curricula, teaching force, and relationships with families and communities. The task force intends for these changes to simultaneously promote intellectual and personal growth. Among its recommendations, the task force asks schools for young adolescents to "reorganize into small communities or 'houses' that foster close relationships between students and teachers; teach a core academic program that results in literacy (including scientific literacy), critical thinking skills, healthy living skills, ethical behavior, and responsible citizenship; ensure success for all students by creating heterogeneous, cooperative, flexible, and resource-rich learning environments; improve academic performance by promoting health and fitness, and providing access to health care and counseling."

Noticeably absent from any of these writings on education for young adolescents and what they need from school is the more traditional approach of teaching the disciplines as disciplines. In fact, the concept of the middle-level grades as a "junior" high school characterized by the organization, curricula, instructional strategies, and psychosocial environments of senior high schools is seen as antithetical to the developmental and intellectual needs of students of this age. The fact that most schools serving students in their middle years follow this traditional pattern (as we will report in more detail below) has provided the impetus for the middle school movement.

Science for Students in the Middle Years

What do these beliefs and recommendations for middle-level education imply for science education? In the 1960s and 1970s, reforms in science education for middle-level education focused on the need to produce more scientists. The primary goal of most of the new discipline-based curricula developed during this period was to prepare students for further study of science (Hurd, 1987; Weiss, 1986). The few attempts made to create interdisciplinary curricula for the middle level came late in this reform period and proved less acceptable to the schools than the earlier discipline-based curricula. In recent years, however, a number of science educators and researchers have drawn on the more general middle school literature to recommend new directions for science programs for young adolescents.

For example, Yager (1988:12) has suggested that middle-level science programs can better accommodate the needs and interests of young adolescents by focusing science lessons on the students themselves, "what they bring to the study, what they can do, and evidence of growth in various domains." He has recommended that program goals be oriented toward the student and that curricula and teaching strategies be based on the past, current, and future experiences of students. Similarly, the criteria developed by the National Science Teachers Association

(Yager, 1988) for selecting exemplary middle-level science programs include

- emphasis on learning how to learn;
- learning science in real-life settings that are interdisciplinary, related to society, and related to daily living;
- learning the independent use of inquiry to identify and solve problems;
- learning decision-making strategies;
- developing positive attitudes about science; and learning about current problems that make clear the interdependence of science and technology and their relationships to other human enterprises.

Other science educators, following the work of Flavell (1963), Case (1985), and other developmental psychologists who have built on Piaget's work, have suggested that science classrooms for young adolescents should be dominated by discussion, opportunities for variation in the nature and pace of learning activities, resource centers, and direct interaction with objects and events. They propose that the curricular emphasis be on linking concrete science experiences with events and phenomena familiar to students (Blosser, 1988; ERIC Clearinghouse, 1982).

More specifically, Rakow (1988:1), following Rowe's (1978) finding that hands-on experiences can positively affect students' sense of control, suggests that hands-on science investigations mesh with the need for young adolescents to "become the authority rather than the teacher or the textbook." This happens, he claims, when students gather data and solve problems. Rakow also suggests that understanding the tentative nature of science can help students in the middle grades become more comfortable with alternative hypotheses and multiple solutions. Science education, especially if linked with technology education, also affords young adolescents an opportunity to study real-world problems and allows them to explore and debate issues of immediate concern to them—health, the environment, and energy. Such study, Rakow claims, can help shape positive attitudes toward science.

In specific reference to science education, the Carnegie Task Force on Education of Young Adolescents (1989) suggests that health education be integrated into the curriculum as an element of the life sciences. Pointing to the Human Biology Program at Stanford University as exemplary, the task force claims that if students learn about health in the context of science, they will better understand how their bodies grow and function. As a result of this life-science focus, the task force claims, young people can come to appreciate the value of a healthy diet and exercise and recognize the dangers of illicit drugs, alcohol, and tobacco. However, the report cautions that, to be effective, such a curriculum must also train young adolescents in skills that will enable them to resist pressures to engage in negative health-related behaviors.

Beliefs about what types of science programs will be best suited to students in the middle grades are widely shared by science curriculum developers and middle school advocates. However, as Hurd (1987:29) notes, most recommendations to match science programs to the developmental needs of adolescents have been "more rhetorical than conceptual" and, as such, have provided little specific guidance for curriculum development. He notes that the "literature on the middle school movement is rich in perspectives of what a curriculum should accomplish, but not the curriculum itself (p. 25)." Moreover, Hurd reminds us that, while the literature offers a great deal of hope about better programs for young adolescents, no solid data establish the effectiveness of either the middle school concept generally or of science programs that match this concept. He concludes "the entire issue of science education for the early adolescent remains unresolved. There is no end of statements and committee reports identifying the problem, but as yet no mechanism or concerted leadership to focus and sustain the essential actions for reform (p. 43)."

Somewhat different from, though not necessarily inconsistent with the perspective of meeting the developmental needs of young adolescents is the perspective of science educators developing curricula based on cognitive science research. For example, Anderson and Roth (1988:14) use two criteria to define the nature of scientific understanding: (1) developing "knowledge that is useful for the essential functions of describing, explaining, predicting, and controlling the world around us...[and (2) developing] knowledge that is conceptually coherent and integrated with [one's] personal knowledge of the world." For many fundamental science concepts, this entails that students go through a complex process of conceptual change. In the case of photosynthesis explored by Anderson and Roth, students must reconceive their common-sense notions of food derived from their own experiences and acquire new understandings about the different metabolism of plants, as contrasted to that of humans and pets, with which they are familiar. Science classrooms generally fail to bring about the change that makes it possible for students to absorb and understand a new concept and reintegrate it into their understanding of the world. It is for this reason that individuals often revert to "common sense" interpretations of phenomena when they are met in non-school contexts, even though the canonical scientific explanation was learned in school. To make possible conceptual change, Anderson and Roth argue, requires attention on sense-making, that is, teaching for depth of understanding (narrowing and deepening the curriculum), rather than breadth of coverage (covering a wide range of content superficially). Of course, current testing of science knowledge, particularly when carried out through widely administered standardized tests, rewards precisely the latter teaching strategy—memorizing as many facts and concepts as possible.

A second requirement for true science learning, according to Anderson and Roth, is flexibility and the use of an array of teaching strategies suited to the progress of individual students. Curricula and instruction must be engaging and accessible to students, yet challenge them, through discrepant information, to work hard and think. The teacher's role is not as an expert who provides the right answer, but as a model who enters into scientific inquiry and discussion and who coaches students to do so as well. Moreover, students need to be encouraged to explain and use their newly gained knowledge themselves, finding out in the process something of the nature of scientific dialogue, the application of scientific ideas, and the process of evaluating one's own and other people's work in science so that valid ideas and appropriate solutions emerge. Unfortunately, the reality of schools and science instruction is far from this ideal.

What Students Actually Experience

As Hurd (1987) points out, despite calls for middle schools and science programs that personalize knowledge, integrate subject areas, provide exploratory experiences, build close personal relationships, and allow for student diversity, few specific programs that make these goals concrete have been developed. Moreover, as we describe below, the more frequently found school organization arrangements, science curricula, science training of teachers, and working conditions in schools all promote traditional, academically focused science experiences for students at the middle level. Few students, it seems, have access to the kinds of programs middle school advocates and science educators recommend.

Organizational Arrangements

Grade Span. Students at the middle level attend many different types of schools. Epstein and MacIver (1989) at the Center for Research on Elementary and Middle Schools report that seventh graders may attend schools that include any of thirty different grade spans. The largest number of seventh graders attend grade six through eight schools (representing a 160-percent increase in the number of these schools since 1970), although in many parts of the country grade seven through eight schools are the norm, and in others, grade seven through nine schools are common (Alexander and McEwin, 1989). Some surveys indicate that schools with grade spans including five through eight or six through eight more often exhibit key middle school characteristics (innovative scheduling, cross-disciplinary teams of teachers and students, supportive guidance practices) than do schools that begin with grade seven (Cawelti, 1988; Epstein and MacIver, 1989). However, middle school advocates note that it takes more than a change of grade span to create a middle school. And some claim that the marked recent shift in grade span is more likely to have been spurred by administrative convenience to alleviate the recent overcrowding of elementary schools than by growing interest in the middle school concept (Rothman and Cohen 1989).

Departmentalization. One of the keys to distinguishing the middle school concept from the more traditional junior high school is the degree to which subject areas are housed into distinct departments where students take separate courses in each subject from specialist teachers. While many middle schools are moving toward interdisciplinary classrooms and block scheduling, distinct daily class periods of equal length for each subject remain the norm with, according to one recent survey, 66 percent of schools serving seven through ninth graders using this type of department-oriented scheduling (Cawelti, 1988). Eighty percent of teachers assigned to teach seven through ninth graders are members of a subject-area department and teach classes in that subject to intact classes. Almost none teach self-contained classrooms covering all subjects (Center for Research on Elementary and Middle Schools, 1987). Only about 16 percent of schools are organized into interdisciplinary teams, about half of which include math/science teams (Cawelti, 1988).

The extent of departmentalization, however, overestimates the percentage of students who stay together for part or all of their school day. About equal percentages of seven through ninth graders stay together for all of their subjects, regroup for one or two subjects, and regroup for all of their subjects (Center for Research on Elementary and Middle Schools, 1987).

Grouping by Ability. The predominant method of grouping students for academic instruction in the middle grades is placing them into homogeneous groups by ability. In a recent science survey, fewer than a third of the teachers reported that their classes included students of widely varying abilities; the rest judged their classes to comprise homogeneous groups of high-, average-, or low-ability students (Weiss, 1987). Because grouping by ability is often accompanied by differences in curriculum, teaching, and classroom atmosphere, it can result in uneven science learning opportunities for middle-level students. And importantly, grouping by ability may place poor and minority students at a greater disadvantage, as they are found disproportionately in low-ability science classes—those with the most limited science learning opportunities (Oakes, 1990).

With departmentalization and grouping by ability, middle schools appear to mirror rather rigid high school organizational practices. Only a small percentage of schools have developed schemes that might more easily accommodate the diversity among young adolescents and the spurts in growth that individual students may be expected to experience.

The Science Curriculum

Researchers have documented that the science curriculum for most students at the middle level focuses almost exclusively on academic preparation and largely ignores other middle school goals such as relating science to everyday life, pressing social issues, or the personal concerns of adolescents (Armstrong et al., 1986; Goodlad, 1984; Hurd et al., 1981; Hurd, 1987; Johnston and Aldridge, 1984; Weiss, 1987). For example, one study documented that middle school science teachers found such goals to be "diffused, impractical, remote, and unrealistic," and that the most commonly accepted reason for having young adolescents study science is that they should acquire specific information on science topics (Hurd et al., 1981).

It is not surprising, then, to find traditional content and instructional modes dominating curriculum and instruction at the middle level. Rather than being exposed to exploratory, integrated science, most students are taught a series of traditional science topics. Rather than learning science in connection with other content areas, most students at the middle level take a sequence of specialized courses—life science, physical science, and earth science—or a series of "general" science courses. Lecture, textbook reading, recitation, and tests most frequently characterize science instruction (Goodlad, 1984; Hurd et al., 1981; Weiss, 1987).

Students at the middle level typically use a single text as the source for lessons, activities, lectures, and reading assignments, and most texts are but watered-down versions of those used in high school science. Although some demonstrations and laboratory work supplement these dominant modes of science instruction, students have few opportunities for direct experiences and hands-on activities that engage them in doing science (Weiss, 1986).

Perhaps as a result of a mismatch between the needs and interests of young adolescents and the science curriculum, many students appear to find science difficult, boring, and irrelevant (Goodlad, 1984). The most recent science assessment conducted by the National Assessment of Educational Progress (NAEP) (Mullis and Jenkins, 1988) found, for example, that only slightly more than half of the thirteen-year-olds said that they thought what they were learning in science was useful in everyday life or that they would use science in many ways as an adult. Even more discouraging, fewer than half the students questioned thought that the application of science could help solve such major social problems as world starvation (25 percent), birth defects (34 percent), and reduction of air and water pollution (49 percent). Fewer than half (43 percent) thought that science would help them earn a living or that science would be important in their life's work (37 percent). Prior science assessments have found that nearly three-quarters of the thirteen-year-olds found their classes boring, and more than half reported that they did not like science and planned to quit taking it as soon as they were free to do so (Hueftle et al., 1983).

While middle-level science programs may turn many students off, these programs seem to affect girls more negatively than boys. Information from NAEP and other data consistently reveal gender differences in thirteen-year-olds' attitudes toward science (Hueftle et al., 1983; Mullis and Jenkins, 1988; Zimmerer and Bennett, 1987).

Teachers

The recent report of the Carnegie Task Force on Education of Young Adolescents (1989) put it bluntly: "Many teachers of young adolescents today dislike their work. Assignment to a middle-grade school is, all too frequently, the last choice of teachers who are prepared for elementary and secondary education." If this is right, it may help explain why there are shortages of well-qualified science teachers at the middle level and why, as states require increasing numbers of science courses for high school graduation, this shortage is likely to grow worse.

In 1986, for example, only 68 percent of science teachers in grades seven through nine had taken the number of college courses recommended for middle-level science teachers by the National Science Teachers Association or had degrees in science or science education. Only 73 percent were certified by their states to teach one or more science subjects (Weiss, 1987), indicating that more than a quarter of the science teachers in these grades were teaching out of their field. Compounding the problem, most of those teachers who are qualified to teach science are likely to lack training in working with students in the middle grades (Padilla, 1986). The National Middle School Association notes that only twenty-one states offer special certification for middle school teachers (Alexander and McEwin, 1989). The status and training of teachers of students at the middle level is treated more fully in the

Center's companion volume (Loucks-Horsley et al., 1990).

Teachers' Working Conditions

The working conditions that teachers encounter in their schools exert perhaps as great an influence on middle-level science teaching as the teachers' background and training. Many teachers of young adolescents find that the departmental structure of their schools and the large number of students that they work with each day make it difficult to accommodate the personal, social, and intellectual needs of young adolescents (Center for Research on Elementary and Middle Schools, 1987). Additionally, few teachers in either self-contained or departmental schools have the time and resources to explore exemplary programs and practices or to work with other teachers in designing programs and lessons that cross traditional disciplinary lines.

Many middle-level teachers must work in more difficult physical environments than teachers of older or younger students. Many junior high school buildings are converted high schools—older buildings that might have few of the physical arrangements conducive to integrated, cross-disciplinary programs (such as, connecting of clustered classrooms and space for small-group work). Science teachers face particular constraints with school facilities and equipment that are inappropriate for inquiry-based, exploratory science activities (Weiss, 1986). In a recent national survey, about one-quarter of the teachers of seven through ninth graders reported that inadequate facilities, insufficient funds for purchasing equipment and supplies, and the lack of materials for individual instruction were serious problems at their schools (Weiss, 1987).

Between Reality and the Vision: Some Obstacles

The gap between the vision of what middle-level science might be and the current reality is substantial, and it is a gap not easily bridged. Many obstacles stand in the way of altering the roles of teachers and the nature of the curriculum, instruction, and assessment they provide in classrooms.

For example, the expectation that teachers can provide for the personal and social needs of adolescents presumes that teachers have the knowledge and skills to assist students with sensitive issues—issues traditionally in the purview of guidance counselors and public health workers. But most teachers have had no training in such matters. Nor do many of them relish the task of counseling students about such family problems as divorce, child abuse, drugs, teen sexuality, and other personal and social matters that weigh heavily on many young adolescents. Moreover, the ideal of providing academic instruction and social support through closer, less formal relationships in small communities of students will require a substantial shift in the school conditions under which teachers and students come together. Most teachers now are burdened with too many students, too many bureaucratic demands, and too little time to take on these added responsibilities. A number of schools that serve young adolescents are moving toward a structure that separates students and teachers into small communities or "houses"; however, one cannot assume that close, productive relationships between the students and their teachers will necessarily follow. Most likely, such relationships will require a great deal of additional teacher education and substantial changes in school cultures and peer-group norms (of both students and teachers).

Implications for Assessment of Student Learning

Nowhere are the gaps larger and the obstacles greater than in assessment. Current practices of assessing science learning with "objective," paper-and-pencil instruments focused on the mastery of basic science facts, using individual assessments exclusively, placing students in individual competition for grades, and measuring the quality of programs by aggregating students' test scores stand in stark contrast to assessment strategies that serve the type of science instruction we and others envision for young adolescents. Most paper-and-pencil assessments work against a curriculum that engages students with rich and complex science ideas and embeds these ideas in real-world problems. Individual, competitive assessments make group work seem irrelevant and discourage cooperation. Measuring program quality by test scores alone does little to encourage schools to have the resources and structures in place that can enable programs to become richer and better suited to young adolescents.

Altering the processes whereby students' progress is measured and programs are judged in ways that make assessment compatible with, and indeed serve, the vision of an ideal science program for young adolescents will be a formidable task. Within classrooms, measures must be developed and used that tap into the students' understanding of large concepts and assess their facility in using these concepts to solve a variety of problems. This is difficult because it probably requires collecting samples of the students' work, recording what the students are thinking as they go about giving science explanations or solving problems, and having materials and equipment available to use in assessment exercises. Assessments that serve the vision of science instruction put forward in the Center's report will also require measures of how well groups of students produce science knowledge cooperatively, in addition to measures of individual contributions to group work.

Reporting the results of such assessments to parents presents problems, since parents are accustomed to having their children's work "summed up" in grades that show, not what the child has learned or how, but how well the child compares with others. Reporting the results of what we consider appropriate assessments upward—to principals, districts, and states—is equally problematic. By its very nature, reporting about large groups of students entails undesirable reductionism, whereas the assessment strategies we suggest (discussed in greater detail in the next two chapters) militate against the type of information that can be easily reduced to chunks of information to be aggregated across classrooms, schools, districts, and states.

Another issue that needs concerted attention is to ensure that external assessments—state and national achievement assessments—correspond to the assessments we suggest for classrooms. Unless external assessments come to match the new strategies we suggest, it is unlikely these new strategies will take hold in classrooms. The pressure to do well on external assessments, particularly when these are used to reward or sanction individual schools or teachers, will continue to drive the direction of classroom teaching, learning, and assessment.

Similar barriers stand in the way of more informative and useful assessments of program quality, as described in chapter 4 of the Center's report, Assessment in Elementary School Science Education (Raizen et al., 1989). It is difficult to develop measures of most important "enabling" features of science programs—sufficient, high-quality resources; good teachers, curriculum, and instruction; and professional conditions for science teaching. Not only are such measures difficult to develop; it is hard to bring them to the attention of the public and policy makers who have grown used to using test scores as the most important—and often the only—measure of program quality. Yet without information on program features, policy makers will find themselves without much guidance on how to improve student outcomes.

With all of these difficulties, why should anyone bother? Our vision of science at the middle level and its assessment has two important payoffs. First, and by far more important, there is the opportunity to help adolescents become critical thinkers, in science and in general, a major goal of education. Second, there is the potential to use science instruction to overcome the traditional mismatch between conventional schooling and the needs of young adolescents. To give some reality to this potential will require curricula and instruction, and teachers and teaching conditions that speak to the possibilities presented by the growing capabilities and interests of early adolescents as described in chapter 2 and the goals of science education delineated in chapter 3. It will also require assessment strategies that are consonant with and support the science education envisioned in this and the Center's other two reports on middle-level education. (See Bybee et al., 1990 and Loucks-Horsley et al., 1990).

In the next two chapters, we discuss such assessment strategies at greater length.

Chapter V

Assessment In Middle-level Science: Improving Current Practice

The scenario on the next page illustrates some of the characteristics that science teaching and assessment can foster at the middle level: the emerging ability of students to evaluate what they know and do not know, their expanding communications skills, their interest in collaborating with peers, and their capacity for understanding specialized information (Superintendent's Middle Grade Task Force, 1987). In this chapter, we consider in greater detail the implications for assessment of the developing potential of young adolescents and the goals of science education. Apparently, the needs and capacities of students and teachers at the middle level directly match the opportunities presented by classroom environments that integrate active science learning and performance-based assessment (Carnegie Task Force on Education of Young Adolescents, 1989).

As noted in chapter 2, from a developmental perspective, middle-level students are working toward engaging in sophisticated types of thinking and reasoning. They have a growing capacity to understand the natural world, which is crucial to science learning, and because of their increasing realization of connections between self and the larger world, they are likely to be interested in applications of what they learn to real-world problems. From an assessment perspective, these adolescents can see themselves in place of others and interpret what others think about them, which is central to self-evaluation. Also, because the locus of control is shifting away from the adults around them to a desire for personal authority, middle-level students can be given more responsibility for self-evaluation. During the years of early adolescence, students are striving for increased independence, and this inclination can be fostered in a constructive way by giving them more responsibility for directing their own learning and monitoring their own progress. These students are ready to benefit from a classroom environment that provides challenging contexts and helps them in their search for understanding of the world around them. Unfortunately, as the review in the preceding chapter indicates, recent studies indicate that most teaching, including science teaching, is instead numbingly dull and disconnected from any meaningful context.

Good science instruction that is also consistent with the inclinations of students aged ten through fourteen requires teachers to explore methods of guiding and evaluating their students, progress which is different from mere teaching and testing for rote learning. Hands-on activities that are so vital to understanding scientific concepts should be an integral part of these methods. Students should be given the

Early in the month, Ms. Lopez's class had visited the local museum to observe and draw primate skeletons. She had highlighted many skeletal features, pointing out that limbs were like levers, and that the skull, sternum, and pelvis protected delicate and important body parts. She had shown her students how vertebrae protected many internal organs and supported the body, yet allowed for flexibility. Later in the month, they had visited the zoo. Ms. Lopez wanted her students to become aware of the ways primate and other animal bodies were alike, and how they differed from each other. Over the weeks, as Ms. Lopez and her students continued to investigate bones, they often discussed fitness, health, disease, and ways that scientists answer questions by studying bones.

After her class had visited the museum and zoo, Ms. Lopez wanted to assess what her students had learned. She set up twelve stations in the classroom, and put a bone at each one. At one station she placed a vertebra of a cow, at another, the rib of a mouse, at another, a plastic human arm bone. At each station she penciled a question or two on a piece of paper, which she left next to the bone.

Ms. Lopez's entire classroom was a resource center for her students. Brightly colored tacks pinned sketches of bones the students had drawn at the museum and posters of animal skeletons to the bulletin board. On a table in front of the bulletin board, a plastic human skeleton, no more than two feet high, hung suspended from its mount. As her students answered the questions, they often left their stations to look at the posters, drawings, and the skeleton. Once the students had finished, she collected their answers, which she would later review.

The next week, Ms. Lopez decided to introduce an activity in which the students would study owl pellets. She would use the activity to culminate the student's investigation of bones and to help assess her students' learning. Ms. Lopez preceded the activity with an explanation of how owls eat their prey, digest the soft body parts, and then regurgitate undigested pellets consisting of bones, skin, and feathers or fur. She challenged her students with a simple question: What do owls eat?

Ms. Lopez began the information-

opportunity to use ordinary tools and to make models using common materials, such as wood, paper, plastic, and metal. They should be given experiences with using tools and materials to solve problems or answer questions. These experiences serve to increase the students confidence that they can care for themselves, and they provide practice for such problem-solving techniques as trouble shooting and designing experiments. Attempts to fit young adolescents into the restrictive cycle of lecture, repeat-after-me instruction, and formal testing lead to frustration on the part of both teachers and students and fail to maximize learning at a critical time in the students' academic career.

Although such a shift in philosophy and methods will undoubtedly be difficult, middle-level teachers can also benefit from an activity oriented approach to instruc-

gathering phase of the activity. She distributed owl pellets, which she had obtained from a school science supply house, to groups of two students. The students carefully pulled the pellets apart, revealing tiny bones inside. Using an identification sheet, the students sorted the bones, matched them, and speculated as to which animal the bones had come from. As the students worked, Ms. Lopez moved from station to station with her stack of index cards and noted the way each pair of students went about their work. She looked for ways the students classified the bones and evaluated how the students worked together. She listened as they discussed approaches to answering the question she had set for them.

Once the students had sorted and grouped the bones, Ms. Lopez asked the students to fasten them to a sheet of black paper, laid out so that the skeleton of each animal was reconstructed as accurately as possible in two dimensions. As the students worked, she gained insight into how the students interpreted what they found inside the pellets. She looked to see whether they identified patterns, noted the size and scale of what they had found, and transferred their knowledge of the skeletons of larger mammals to a new set of mammals. She watched for when they ruled out hunches refuted or not supported by evidence.

When at last the students had finished, she asked for each pair to decide on their answer to her question, What do owls eat? On one level, she expected such answers as voles and field mice. But, as each group showed their data to the class, she began to look for more complex scientific thinking. Were the students demanding justification for inferences? Had they questioned whether the data were sufficient to make generalizations? Did the students ask new questions?

As her students discussed each presentation, she took notes. Later, she added these records to the skeleton charts and written summaries of the owl's place in the food chain that the students had completed earlier. She also had their answers to the previous week's questions on individual bones. She now had a rich supply of data to assess her teaching and her students' learning.

tion and assessment. As we note in chapter 4, compared to elementary school teachers, middle-level teachers often are responsible for teaching many classes of students, instead of one or two, which makes it difficult for them to assess accurately the learning of each student. They also have additional demands on their time, such as guidance counseling and managing extra-curricular activities. Thus, an instruction and assessment model based on the learner, which gives students more choice for their own educational success, can help middle-level teachers manage their time more effectively.

Middle-level teachers should capitalize on the growing independence of their students. This independence offers a chance for the teacher to divert energies from managing the classroom to facilitating learning. For example, whereas nearly one-fourth of the elementary school teachers report having to spend more than three hours a week maintaining order and discipline, virtually none of the middle-level teachers report such demands on their time (Mullis and Jenkins, 1988)—almost 80 percent of the middle-level teachers report spending less than one hour per week on these activities. Middle-level students need less direct minute-by-minute guidance about "what to do next" than do younger students, and their need for independence will benefit from opportunities to direct their own learning.

In fact, among elementary, middle, and high school teachers, middle-level teachers have the greatest freedom to build effective evaluation strategies for their students: they are released from the constant supervisory activities so prevalent in elementary school, they are given increased scope in terms of curriculum and student capabilities, and, for the most part, they are only beginning to feel the pressure that grading can exert on the future of individual students. Although in recent years grades have become a central issue for students and administrators, the evaluation policies of middle-level science teachers still have relatively limited consequences for students, careers or plans for higher education. Thus, the middle-level teacher is, to a certain extent, free to experiment with a variety of participatory assessment strategies to find those that both foster the goals of science learning and meet the needs of students in early adolescence.

Assessment in the Service of Instruction

Why do teachers need to assess students? The primary reasons ought to include concerns for instruction rather than the formal reason of assigning grades. For example, teachers need to assess students' prior knowledge in order to know where to begin instruction and to monitor progress to see if certain concepts and skills have been learned. Teachers need to assess what knowledge and skills need to be "retaught" and where to go next.

Reporting to others—parents, school administrators, and the community—about students' progress is also part of assessment, and these more formal activities deserve attention. However, the bulk of teachers assessment activities ought to relate directly to providing appropriate instruction, and should not be separated from instruction, but integral to it. In fact, good instructional tasks and good assessment tasks should be indistinguishable, as the example on page 44-46 illustrates.

Construction rather than Instruction. Although individualized instruction and evaluation are held out as an ideal, the pressures of class size and multiple-preparations may render them infeasible. Middle-level teachers frequently are faced with the demand of teaching many more classes than teachers in the elementary grades, and "getting to know" so many students is often difficult enough without trying to tailor instruction and measure its effectiveness on a one-to-one basis. Thus efficiency in assessment must also be a priority for the middle-level teacher. The inclination of students in the middle school to strive for increased independence can be fostered constructively by giving them more responsibility for directing their

Please turn to page 50

Ms. Lopez's class has been studying primates as the basis for learning about communication and behavior. Ms. Lopez has found that a trip to the zoo to observe monkeys, two resident gorillas, and two pair of orangutans is a good way to introduce thinking about the advantages of group living and social interactions from a somewhat removed position. Her middle-school students find primates interesting and funny, so different from human beings yet eerily familiar. This helps Ms. Lopez engage her class in this animal study. These students can distinguish behaviors in other species that have parallels in their own lives while beginning to understand the perils of anthropomorphism. By learning about—perhaps—less complex primates, they can begin to develop a language and conceptual basis for reflection on their own relationships and interactions. After the zoo visit, the students have worked collectively to interpret their observations. A list of biological and social functions is beginning to emerge from the observed behaviors, which contains such notions as communication, caring for young, protecting themselves, social and family groups, and getting food. In addition to learning about primates, students have been practicing their skills, including observation, recording data, and interpretation.

Ms. Lopez has emphasized six characteristics of primates that they share and that differentiate them from other groups of animals: grasping with fingers and/or toes, the opposable thumb, finger and toe tips for touching, great capacity for "thinking," stereoscopic vision and seeing in color, and group living. Ms. Lopez wants to focus on the last: group dynamics and the advantages of group living—safety in numbers, cooperation in obtaining food and caring for young, and establishing bonds and aid in the health of the group through such interactions as grooming.

How can Ms. Lopez assess her students' learning and understanding?

She wants her students to make use of information-gathering and problem-solving skills, as well as to exercise habits of mind that characterize the doing of science. She asks them to (1) select a behavior or set of behaviors from their journal observations that they found particularly interesting. Next they are to (2) generate ideas about why these behaviors might be useful to this type of primate. They are to then (3) set the task of researching information on the type of primate observed, seeking evidence that will help to (4) decide which of their own ideas are supported and which are not.

Each group of students will (5) report to their classmates on the following topics:

- *This is the behavior we found interesting.*
- *These are some ideas we have about the benefit of the behavior to the primate group.*
- *Here is what we learned from research to support our own thinking.*
- *Here is what we learned that does* ***not*** *support our ideas.*
- *This is a summary.*

In addition to learning about primates and having her students develop crucial thinking skills, Ms. Lopez wants to reinforce some of the larger concepts and themes underlying her science instruction: diversity and variation; patterns, rhythms, and cycles; and models and theories. Thus, to explore wider ranges of animal interaction and strategies for *survival and to emphasize some of these larger concepts, she proposes that the class form small groups to study the behavior of other animals and insects. In this instance, Ms. Lopez asks the students to choose their own groups, and it doesn't take long before the groups are formed and begin suggesting species to study. Rosemary's group wants to learn about dolphins and Joling's about elephants. To ensure diversity, Ms. Lopez suggests that the other two groups study ducks and honeybees. Each group will conduct research on existing information, including visits to the library and interviews with relatives and neighbors who might have special information. Using their experience with primates as the foundation of their study, they are asked to focus on how their chosen species communicates, protects itself, cares for its young, gets food, and organizes itself socially and in family units. Their notes will be assembled for eventual review by Ms. Lopez, with the final goal an oral presentation by one of the group members to the entire class. The presenter will be selected by a random method. As an aid to the presentation, each group is advised to prepare a poster depicting the key points of its findings. After the presentations, the class will discuss the similarities and differences among groups and make generalizations about the behaviors of living things.*

Ms. Lopez's instruction is designed to lead students to increase their science knowledge and skills, but at the same time she is systematically collecting evidence of their learning through the notes, the posters, the oral presentation, and the quality of the group discussion. Further, students are building their social and communications skills and being given responsibility for the success of their own group.

own learning and monitoring their own progress. Teachers will use their time more effectively if they share the burden of evaluation with their students and utilize the students developing abilities in the areas of peer-and self-evaluation.

Laboratory activities and group projects offer opportunities for efficiently evaluating students' learning, while also affording students a chance to work independently and to collaborate with their peers. When the challenges are appropriate to the manual skills, available materials, and stage in development of critical thinking skills, middle-level students are able to use tools and go through the process of solving technological problems. Furthermore, students of this age are apt to find such active learning of interest, and problems of technology often are effective vehicles for developing critical thinking skills.

Students should be encouraged to ask questions, and teachers can use a variety of follow-up procedures to foster such inquiry. As illustrated in the example at the opening of this chapter, students can be asked to work in groups to answer each others questions, or the teacher can periodically place interesting discussion questions in a log. When students display interest in a particular topic, the teacher can ask them to work in this area outside of school and to present their findings to the class.

Teachers can assign more structured tasks and long-term projects in which students are responsible for various phases along the way, such as collecting and interpreting data, reporting their findings, and formulating a researchable question. Rather than evaluating the results of each phase independently, teachers can ask their students to proceed on their own, only seeking help when they need it. Alternatively, the students can share their progress periodically with classmates to receive suggestions about areas in need of improvement or more in-depth treatment.

Because it is difficult to monitor the daily progress of thirty to forty students working on laboratory activities or projects, a teacher can ask the students to keep notes on their investigations, and detail their procedures, findings, and interpretations. These accounts need not adhere to the formal style of laboratory notes, but should go beyond simple narratives and take on some of the characteristics of actual scientific reporting.

These notes and reports of scientific investigations can be assembled into portfolios and used in teacher-student conferences to discuss the student's growth in science knowledge, in understanding the principles of scientific inquiry, and in the ways in which these can be applied. For example, students need to hone their understanding of the differences between observation and inference and speculation.

If teachers are interested in more frequent evaluations of their students' progress, notes and written descriptions of work underway can be turned in on a daily basis and quickly reviewed—not to mark with a red pencil, but to evaluate students' progress and offer feedback as quickly as possible.

By assembling portfolios, students will also have the satisfaction of recording their learning for easy reference and review. The portfolio can give these middle-level students a sense of independent accomplishment and, if the journals are kept by groups of students, this learning activity can also take on a collaborative aspect.

Assessment and cooperative learning. Using cooperative group learning strategies in science instruction and assessment is pedagogically sound and practical. Not only do such procedures include efficient peer-evaluation techniques, but students working together in small groups helps to foster positive, academically oriented peer-group norms. Cooperative learning gives students greater opportunities and incentives to articulate and communicate their understandings. By working in groups, students learn cooperation and build interpersonal skills of intrinsic value. Cooperative and group work offer a practical alternative to individual laboratory work, simplifying logistics and reducing the amount of science equipment required. Cooperative group work has special value for young adolescents as it turns their natural inclination for peer interaction toward a constructive end.

Cooperative learning approaches, however, usually force teachers to set aside some of their assumptions about instruction and assessment, and they must accept and nurture a different sort of classroom climate. In these situations, the most important decision-makers are the students themselves. When students work together, it may be difficult for teachers to quantify learning outcomes, gauge the success of

Mr. Nguyen's class has been studying simple machines. He introduced this activity by having students identify, draw and explain simple machines used in their homes. As a culminating activity he has assigned the construction of an article that does something using a combination of two or more simple machines. He has supplied his classroom with wood, cardboard, glue, saws, nails, screws, wire pulleys, rope, etc. Students who need additional materials put a list on the board, and Mr. Nguyen tries to have them by the next class period. Some students are working outside of class as well. Mr. Nguyen is interested in watching his students to see how they move from an original idea to a final product. He asks them to keep journals that will chart the evolution of this process. He plans to videotape each student or pair of students presenting the process and product for the camera. The journals, the report on camera, the product, and his observation of his students at work will enable him to assess many of the skills, attitudes, and ways of thinking that he is trying to foster.

In her science journal, Sally entered the following notes made during her trip to the zoo:

Observations	***Thoughts and Ideas***
8 monkeys, 2 are much smaller than the rest	*Two are babies*
A small one clings to a large one	*Is this a baby with its mother? Who takes care of the babies?*

Observations	***Thoughts and Ideas***
Another small one keeps darting to the food pan and then back to the same large monkey	*Is this a baby learning to get its own food?*
One large one sits in a corner with its face turned away.	*Do all the monkeys behave this way sometimes or is this behavior unusual?*

As Ms. Lopez checks Sally's entry, she asks her how she might find out some answers to her questions. Sally decides she will check with her partner about his observations and ideas. Then she will not only read about Columbus monkeys but also about some different monkey species to see whether the behaviors she has observed characterize monkeys in general or only Columbus monkeys. She will record what she found out in her science journal. She also plans to go back to the zoo to observe additional monkey behaviors, including those of some very unusual spider monkeys. She might ask Jos, who can draw very well, to go with her and sketch some of the monkey behaviors, because his pictures sometimes tell more than just words. When she goes to visit her grandparents in San Diego for Christmas, she will make a trip to the large zoo there to observe and record the behaviors of additional monkey species. She thinks how interesting it might be to become someone like Jane Goodall, whom she has seen in a documentary on television.

students' interactions, assign marks or grades to individuals, or ensure that all members of each group are participating and learning. Students will have to come to understand their collective responsibility to work and learn together and to assume some of the burden for informal, individual assessment of learning outcomes.

The experienced teacher understands the significance of these new responsibilities. Although adolescents should be developing a strong sense of their own personal strengths and weaknesses through the comments of their teachers and classmates and through their internal criteria for self-evaluation, they are likely to need specific instruction in how to work together and assume collective responsibility. They must learn to plan together, to apportion tasks and duties among themselves, and to involve all members of the group. At its best, cooperative group work leads the students to discover that each has different talents and interests, and each has special knowledge and experience that can be brought to bear. After carefully preparing students for group work, the teacher must assume a different role—that of facilitator, rather than information source.

Of course, the teacher must still monitor individual learning outcomes as well as group processes in order to evaluate the progress of each student. Some specific techniques provide such monitoring and encourage student responsibility at the same time. For example, the teacher might announce that one member of the group will be selected to report on the group's procedures and findings, and some or all of every group member's grade will be based on the quality of that presentation. Another approach might be to give one member of the group the specific responsibility of ensuring that everyone understands the concepts involved, that everyone helps carry out the task, and that everyone demonstrates understanding by participating in the group's deliberations.

Cooperative learning groups, composed of students with complementary skills, can provide a supportive environment for science learning and offer an efficient way to assess students' progress. By giving group members collective responsibility for ensuring that each member of the group has completed a particular assignment, teachers can ease their own burden in a way that is compatible with the needs of middle-level students. Whether the group's assignment is to learn a new concept, understand how to control for variables in a complex experiment, or take account of critical factors in the design of an application, collaborative learning and peer assessment can help to strengthen students' interest, provide less able students a chance to learn from their classmates, and reinforce understanding and skills as students review materials to check their own progress.

It is during the middle-level years that differences in science achievement and perceptions toward science become solidified for girls and minority students. For example, the NAEP science data (Mullis and Jenkins, 1988) indicate that, while average science proficiency for third-grade boys and girls was approximately the same, except in the physical sciences, a performance gap was evident by the seventh grade and increased by high school. Cooperative learning groups in which students have an opportunity to display a variety of talents provide a richer context and greater promise for girls and minority students to develop their own self-confidence in scientific endeavors. In addition, the support of their peers may help increase their science knowledge and skills, and working on science in a social atmosphere may enhance its appeal as an interesting, worthwhile endeavor.

Scientific inquiry as assessment. A project-oriented approach to science learning provides ways to integrate assessment techniques with activities in ways that reinforce and extend learning, while also gathering useful information on students' progress. As middle-level students develop new abilities and interests, and as the curriculum becomes more challenging, teachers are offered rich opportunities to explore different assessment techniques and to find those that are best suited to the diverse needs of their students. Yet, over half the middle-level science teachers report spending one or more hours per week in a typical science class administering short-answer tests or quizzes; considering that the total amount of time the bulk of the teachers (70 percent) report spending on science instruction each week is only three or four hours (Mullis and Jenkins, 1988), this is particularly distressing.

It is true that some things are easier to assess than others, and that paper-and-pencil tests and quizzes appear a quick way to test factual and conceptual knowledge. Yet the ratio of testing to instruction reported by middle-level teachers seems inordinately high for what can be learned from using paper-and-pencil methods to assess the process and thinking objectives that are at the heart of the scientific method. Scientific reasoning, observation, and experimentation must be grounded in a nontrivial knowledge base of the phenomena investigated, not in the end-of-chapter quizzes found in most science textbooks.

Middle-level science teachers should seek out ways to assess the application of scientific processes in the context of the learning units they create, but following the textbook and simply doing activities does not necessarily offer appropriate opportunities to do so. Unfortunately, activities in science classes often have little to do with real science. Textbooks publishers unintentionally tend to portray science as a static body of facts to be recalled and rules to be applied in solving artificial problems; laboratory exercises all too often entail fixed, step-by-step procedures that attain results known in advance. Students seldom are asked to give justifications for their answers, explain how their experimental procedures and findings support their inferences, demonstrate that their designs serve the intended functions, or otherwise make their reasoning explicit. Science teaching must cross over from telling students about science and technology to having them in some small way participate in science and develop pieces of technology. The scenario on pages 55-56 illustrates the integration of instruction and assessment through an extended unit that addresses several of the points we make in this chapter.

To meet the requirements of recording grades or providing written documentation of each student's individual progress, teachers can assign meaningful, individual tasks on which to base their evaluations of each student's performance—for example, designing and building models, conceiving of and conducting experiments, carrying out demonstrations for the class, or conducting sophisticated oral or written presentations on the development and results of their investigations. The assignments can fit into larger group or class investigations or be complete efforts on their own. In either case, each individual assignment will yield a product or record of student achievement that can be evaluated, and each will represent a much more significant accomplishment than a perfect score on a limited paper-and-pencil test.

At this point, the reader might agree that assessment needs to serve classroom instruction, but, nevertheless, question the applicability of the suggested assessment strategies for large-scale assessments. However, even short-term assessment exercises—whether for use in a single classroom or in large-scale assessments—should include hands-on performance tasks that allow students to demonstrate their

Ms. Washington's treatment of the topic of rivers emphasized energy relationships in the generation of electricity, the problems of the design and construction of dams, and in such geological processes as sediment transportation. Her assessment plan for the unit on rivers had multiple objectives: (1) to determine how well the students understood the energy relationship; (2) to assess the students acquisition of formal thinking skills; (3) to assess how well the students were developing skills in self-assessment of their own learning; and (4) to assess acquisition of the social and communication skills for working in groups.

Ms. Washington presented the assessment exercise to the class in the form of a "what if" question. But first, she reviewed with the class the interactions of the water in the river with objects, substances, and living things it encountered along its path of flow, focusing on energy transfer and conservation in these interactions. Then she asked the students to think about what would happen if their city decided to remove the power dam. Would it be possible to return the river to its original state? What would be the consequences of removing the dam on the local river? With Ms. Lopez's help, they could form groups to consider how their observations of the river would have been different if a single physical property of the water were differentPspecifically, its density. What if water were as dense as mercury, how would the river be different? Each student thought about the question overnight and came to class the next day with a list of possible differences and some indication of which differences he or she was particularly interested in investigating. With Ms. Washington's help, the students categorized their ideas and reduced the long list they had generated to several questions, which they used in forming groups to investigate individual questions.

After briefly reviewing the class procedures for working in groups on extended projects, which included the importance of planning and keeping records, the groups began work on their individual questions. One group formed to investigate how the dam affected the movement of sediment down the river. Had the dam trapped sediment behind it? What were the consequences of the trapping? After developing a preliminary plan for investigating these questions, this group began its work by (1) building a small-scale stream, (2) seeing how water velocity changed as the stream entered the reservoir, and (3) learning about how fast the water has to travel in order to move the sediment.

Another group was interested in what would be done with all the concrete that had been used to build the dam. They began their study by asking how materials are recycled and whether old concrete is useful. They learned that some chemical reactions are not reversible. They went on to learn about the technology of demolition and did some brainstorming on new uses for old concrete.

A third group of students was particularly interested in how the old reservoir site could be restored to something like its original conditions. In the process of planning they discovered that restoration ecology is an important new area of science.

A fourth group investigated the prob-

lem of alternative energy sources. If the dam was no longer used for electric power, where would the growing community get energy? In their planning they discovered that systems analysis would help them understand the larger problem. They did several experiments that showed that people tend to waste electricity and that energy conservation is an untapped "source" of energy.

As the students worked on their problems, Ms. Washington circulated about the classroom, offering information and encouragement, rewarding desirable behavior, and making systematic observations. She augmented her observations with information about her students progress obtained from their weekly progress reports to the class and from their journals.

When planning this assessment activity, Ms. Washington developed several observation sheets organized by her for objectives for the river unit. These would guide her observations, which she could record systematically. She focused on the following:

- ***Acquisition of subject-matter understanding.** Because energy was a subject-matter focus for this unit, Ms. Washington listed the forms of energy and energy particles that the class had studied on her observation guide. Since her objective was to assess the students' development of understanding, she noted instances of the application of these concepts and principles, sometimes recording the students statements (both correct and incorrect) about the forms of energy and energy principles.*
- ***Acquisition of formal operational thinking.** On another observation guide, Ms. Washington had listed aspects of formal operational thinking to guide her observations of her students cognitive development. For example, the creation of the scale model of the stream allowed her to assess how well her students were learning to reason proportionally. The river assessment exercise provided many opportunities to observe the students ability to identify variables, to test which were relevant, to hypothesize, and to design tests of their hypotheses.*

Ms. Washington was particularly alert for instances of reflective thinking, instances when the students were talking about how they were thinking about their problem or when they were identifying thinking processes that they had explored but that proved deficient in some way. She also was watching for instances when her students commented on the extent of their understanding of a concept or principle or when a student noted how his or her understanding had improved or was called into question by the addition of new information. This group exercise also enabled Ms. Washington to observe her students' skills in viewing explanations from different perspectives as they challenged each other's ideas and observations.

- ***Acquisition of skills in self-assessment.** Self-assessment skills depend on the capacity to think reflectively. A significant difference is that self-assessment requires students to judge their own understanding against a standard. In the case of the river unit, it was the standard set by Ms. Washington's expectations. Ms. Washington intended to be alert to how well the students understood her expectations and their ability to assess their performance against her expectations.*

- ***Acquisition of skills in group process.*** *Ms. Washington noted several behaviors in her observation guide she would track as groups proceeded in their work: listening to the ideas of others, challenging them appropriately and considering the evidence presented, putting forward one's own ideas clearly and defending them while considering countervailing evidence, being willing to put aside one's own approach in favor of a more promising one proposed by others. She also recorded progress in oral and written communication made by the students, the latter through their journal entries, laboratory notes, notes on background reading they had done, and records of their planning and procedures. As groups reported on their progress to the rest of the class, she noted the quality of their presentation and visual aids and also the responses elicited from the audience of their fellow students.*

proficiencies in laboratory and science-thinking skills. One step in the right direction is to pose a problem for students to solve in a limited amount of time, using equipment that permits alternative solutions. (A larger, more difficult step is to involve students in finding their own questions to pose and investigate about some phenomenon, which makes life more complex for the teacher.) In one such problem, based on an exercise first developed in Great Britain (Assessment of Performance Unit, 1984-85) and that has been adapted for large-scale testing, the students are asked to determine which of three brands of paper towels holds or soaks up the most water. They have available samples of the three kinds of towels, beakers, a scale, a pitcher of water, and other materials. Appropriate solutions include saturating and then weighing the towels; saturating the towels and then wringing them out and measuring the water released; and soaking up as much as possible of a known quantity of water, then seeing how much of the water is left.

Grading for such an activity might be based simply on whether the students were able to arrive at an acceptable solution, or the solutions might be rated as to adequacy and sophistication, with more or less credit given, depending on the solution employed. Another such problem, "Survival," was originally developed in Great Britain and was adopted by the National Assessment for Educational Progress (1987) for its pilot study on using performance tests for assessing higher order thinking. Students are given two or more different kinds of materials and fasteners and scissors, aluminum and plastic cans of several sizes (which they can fill with hot or cold water to stimulate persons), a fan (to stimulate wind), and various measuring devices—thermometers, a ruler, a stopwatch, and graduated cylinders. They are to determine which of the fabrics would keep them warmer on a mountainside on a cold, dry, windy day. In this exercise, students need to identify the variables to be manipulated, controlled, and accurately measured and recorded. They need to be able to draw a reasonable conclusion from their data and justify it.

In both these instances, the teacher should ask the students to explain how their conclusions followed from their procedures and observations, and the teacher could rate the quality of their explanations. Accuracy of the procedural and measurement techniques might be rated separately, as might be the adequacy of the records kept.

In short, when using scientific inquiry as an assessment approach, the teacher should emphasize both the approach and the product, how a student obtained an answer or carried out a hands-on activity, as well as the "appropriateness" of the answer or performance.

For the teacher's purposes, such performance assessments can form a natural part of instruction. The solutions students propose can be discussed as they occur and evaluated for logic and rigor. The teacher can discover much about each student's strengths and weaknesses that might be relevant to subsequent instruction.

Using knowledge and skills. The principle of use and application in a wider context (sometimes called transfer) is critical to the assessment of science knowledge and process objectives. In order to find out whether students can apply what they have learned in new contexts, teachers may need to deliberately avoid asking some questions or giving some examples in their classroom instruction and reserve these materials for assessment questions exercises, or special projects that will yield evaluative information. Just as an English teacher might ask the students to analyze a poem they have not seen before, the science teacher might ask students to discuss the special environmental adaptations of an organism they have not studied, or ask them to predict the chemical properties of a compound with which they have not worked.

It is often possible to imagine a continuum of problems, increasingly dissimilar from those actually studied. If students have studied the motion of frictionless pucks, a question about billiard balls and bumpers would introduce few new elements—but, a problem about the collision of objects in three-dimensional space might involve significant transfer. Still more remote would be problems involving the simultaneous interaction of more than two bodies in space. In contrast to giving the students problems, the teacher could ask them to think up their own questions about applications of friction. Both strategies would give teachers a good idea about where the students might fall at different points along the continuum of being able to apply their science knowledge and skills to unfamiliar problems and situations.

The nature of this continuum in any particular case would depend on the kinds of activities the students had engaged in and the instructional goals of those activities. The basic principle, in any case, would be that the difficulty would be determined by the number of common versus dissimilar elements in the instructional versus assessment situation, where element is loosely defined. For example, in an assessment exercise about a situation including only a few new elements, the students might be asked to adapt solution procedures previously learned and arrive at precise answers. For more novel situations, they might be asked to estimate or roughly describe what might happen. In the example given above about friction in three- dimensional space, the students probably could do no more than to indicate the nature and direction of changes that friction would introduce and, perhaps, to justify their answers. For the most unfamiliar problems and contexts, the expectation might be that students could indicate what basic principles would still be expected to hold, for example, the conservation of energy and momentum or the balancing of forces.

Broadening the definition of what counts as assessment. Teachers can collect "hard evidence" of performance in many ways, without administering a single test or quiz. To collect evidence of learning in these more constructive and meaningful ways, middle-level teachers must learn how to rely on their judgment and powers of observation. Teachers should question why, for one or another student, their subjective assessment of achievement might not agree with the test-score averages maintained for that student in their grade books. This may be because the tests teachers give on which they base students' grades are not refined or comprehensive enough to reflect the numerous and complex facets of student learning.

Much—probably most—of the information teachers use to guide their instructional decision-making comes, indeed should come, not from formal tests, but from informal classroom observations. Although multiple-choice and short-answer tests and quizzes can go beyond measuring facts and convergent thinking, such instruments are very difficult to develop, even by professional measurement experts. End-of-chapter questions and multiple-choice and short- response tests can provide a measure of the students' surface understanding of processes and their ability to define terms. However, over-reliance on these tools can lead teachers to overlook other highly valued aspects of learning, such as the inclination to question, the ability to transfer learning to new situations, and the capability to analyze complex interrelationships.

Good instruction requires that teachers be sensitive to their students most subtle signs of progress. Monitoring facial expressions for flickers of understanding or puzzled looks is perhaps an obvious example of one way for a teacher to "tell how I am doing." Viewed differently, it is also a way for teachers to tell how their students are doing.

There are ways of doing such observations better and, at the same time, increasing the credibility of these observations as a source of information for marking and grading, and communicating with parents. First, such observations should be somewhat systematic—done on a regular basis. Teachers might carry around a packet of index cards to jot down observations on what particular students do from time to time. They might spend a few minutes at the end of each day (at the very least, every few days) to file these observations for future retrieval. Teachers should be alerted to the human tendency to note the atypical and neglect the commonplace (Almy and Genishi, 1979). Routine observations are of value. Also, it is important that informal observations systematically cover all the students in the classroom. In short, teachers should be scientific observers. Informal observations on the face of it seem more valid—less artificial and contrived—than more formal, written measures. However, unfortunately, they may also, on the face of it, appear less reliable and lacking in the comfort provided by a row of "objective" scores in the grading book. Reliability comes through standardization and replication. Multiple-choice tests are reliable, in part, because they are standardized across learners, but also in part because they involve the summation of many independent responses to a variety of questions. The same principle can be used to enhance the reliability and the status of informal observations, so that they will provide the same security under fire as the row of test scores.

Reliability can be increased by aggregating the results of observations over multiple occasions. Validity will be highest when multiple observations also involve a degree of "convergence" or "triangulation," a synthesis of evidence from different contexts, applying different modes of representation. Although the informal observations themselves do not yield tangible pieces of evidence that can be put on the bulletin board or sent home to parents, documentation of these evaluations for students together with the behaviors that support the evaluations can be very valuable and powerful. Parents and administrators, as well as teachers themselves, should remember that these assessments are individualized and made by the person in the best position to make the judgments. Teachers should not feel uncomfortable with the idea that their perceptions count as assessment. Such written records can provide a way to chart the progress of individual students and the class as a whole.

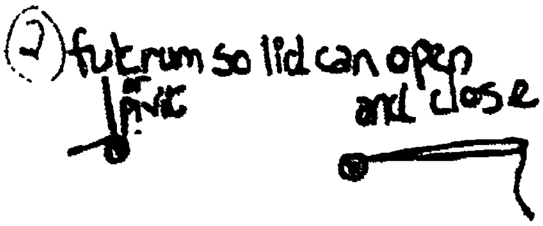
As they refine their monitoring skills, teachers can listen to and reflect on the questions that students ask, not just to provide an answer, but to evaluate what such a question could mean from the perspective of students' misconceptions or misunderstandings. Interpreting conversations among students in cooperative learning situations and using this information to make determinations about the quality of their understanding can also be an illuminating strategy and provide the basis for reiterating material with greater specificity or in a new context. Simply observing students as they work on projects can be another very useful way to assess their learning. If students are having difficulty applying their science knowledge and understandings to hands-on situations, this may indicate that they have an inadequate or flawed understanding of the underlying concepts.

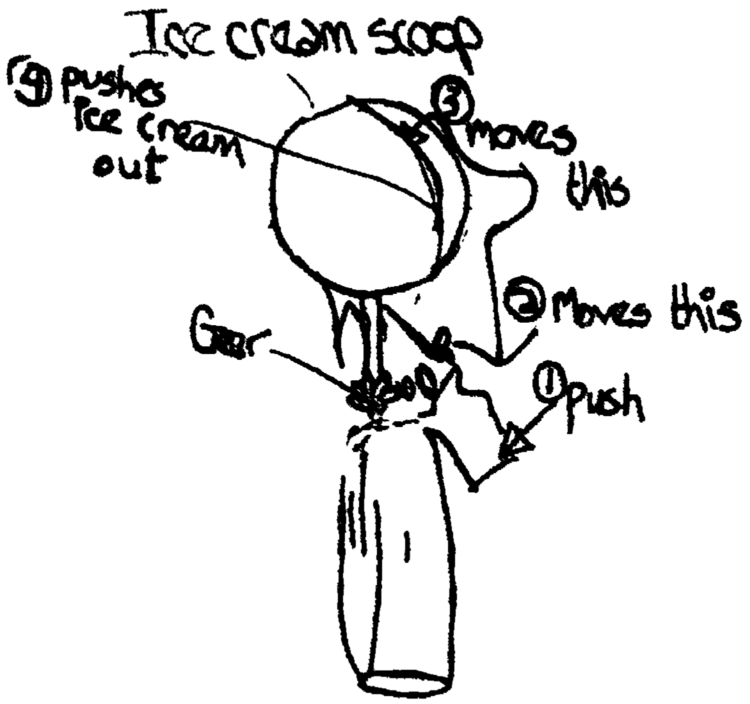
A caution is in order here. Rote hands-on activities are no better than rote memorization. It is entirely possible for students to perform the steps of an experiment or to follow instructions that lead them through an activity without having even a glimmer about the purpose of the experiment or project. Teachers, therefore, must be aware of the concepts or deep learning they are trying to foster and continuously check each student's understanding against these goals. Simply asking students if they know the scientific principles involved in an activity can be revealing. Beyond that, asking them to apply those principles to a new situation might yield interesting and informative replies.

Teachers have many ways to assess students by using their considerable powers of observation and confirming these observations through questioning and dialogue. Documentation of these assessments will provide useful and valid evidence of students learning, or the lack thereof. Teachers should recognize, however, that such assessment strategies can raise more questions than they answer. For one thing, for some students, problems with language and inability to express themselves clearly may be confounded with problems in science understanding. For another thing, in noting a particular student's misconceptions or lack of understanding, a teacher might wonder if that student has other misconceptions in related areas and also need to begin to explore these.

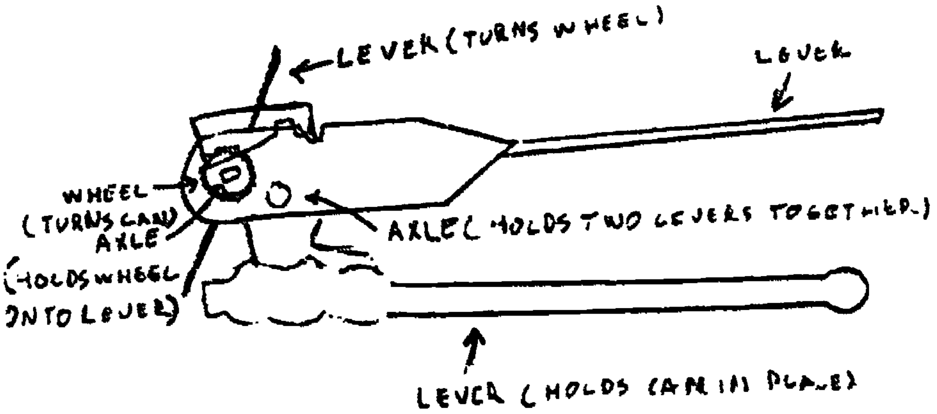
However, as middle-level teachers begin to merge instruction and assessment—watching how students approach and perform tasks, monitoring the questions they ask, and evaluating the final products of sustained learning—they will have ample information to assign a defensible grade. More importantly, they will have developed assessment procedures that facilitate their instruction and respond in a more comprehensive way to the needs and abilities of their students.

Taking assessment out of the classroom. Out-of-school projects and homework can be a very productive and efficient way to monitor the students' progress. Asking the students to answer a question in depth or to conduct activities





Mr. Nguyen has introduced his class to simple machines, giving them experiences and challenges with inclined planes, levers, pulleys, wheels and axles. Eventually he wants his students to construct some machines of their own. Before he goes on to this part of the unit, however, he wants to emphasize the usefulness of machines in everyday life. He has found that drawing can be a powerful tool for assessing understanding. He assigns a piece of homework asking the students to find one or two household objects that are simple machines and draw the object, and to name the simple machines. When the students hand in their homework the next day, Mr. Nguyen is excited. Looking at the drawings, he is pleased that the students are able to identify and name simple machines. He is struck by the care and accuracy of the drawings, which suggest to him that the students were invested in this assignment, that the school-real life connections have been made. He puts the homework drawings on the walls of his classroom and notes the pride with which the class looks them over. The students are reporting finding simple machines everywhere—flagpoles, road machinery, kitchen gadgets, roofs, and ramps. He speculates that the students' feeling of competence and the connection between school work and the real world have enhanced a disposition to acquire additional knowledge.



at home and write up their results for review conserves classroom time and resources, while at the same time giving teachers the opportunity to gauge the students' learning. (An extended example of this strategy is given in the next chapter.) Many avenues of assessment open to teachers beyond the commonly adopted worksheets and end-of-unit tests, including investigations, projects, journals, in-depth written reports and even drawing—as illustrated on pages 61–62—can be undertaken outside of the classroom— in the home, school, or community—and this adds to their potential appeal and relevance.

Methods of assessment that are based on projects and explorations and that are an extension of student's own interests and learning experiences may be viewed as challenging, relevant, and even fun, rather than tedious. For example, investigations that involve controlled experiments can provide opportunities for the students to apply their understandings of science in settings relevant to their daily lives. Watching a growing plant or animal can provide the opportunity to keep a daily record or journal, which the student can submit after several months for the teacher's evaluation.

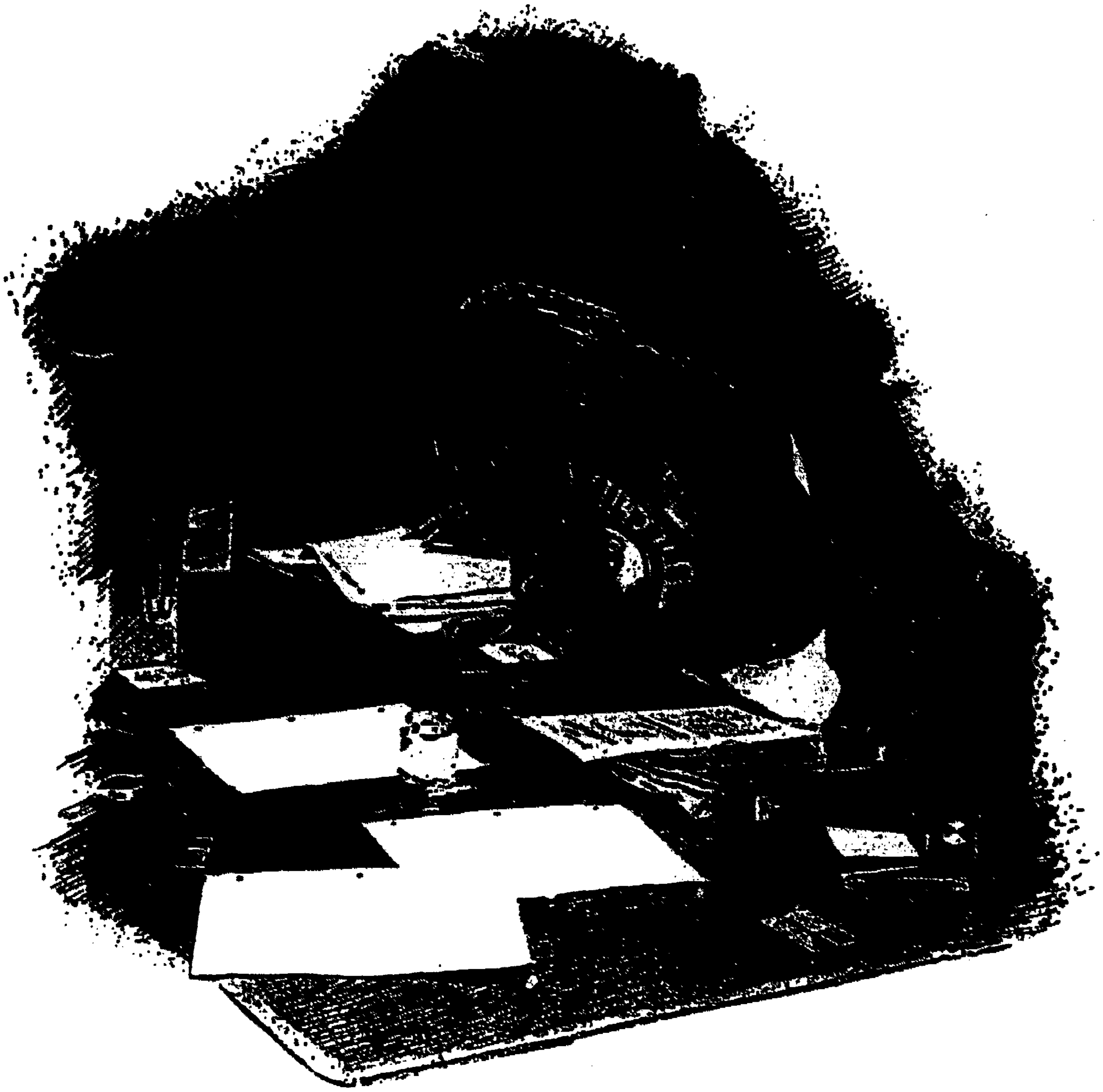
Projects for which the students must collect information from their friends, neighbors, and relatives or solve a school or community problem might be of relevance to their own concerns and interests. For example, investigating other's points of view about pollution or waste disposal can help the students understand the ways in which such issues intertwine with science and technology and touch people's daily lives. Students should learn to appreciate that potential solutions involve taking account of technological advances, scientific principles, and underlying political complexities. If students are asked to interpret their findings at regular or significant intervals, then their changing interpretations provide a long-term perspective on their learning.

Finally, taking assessment out of the classroom instills values that are consistent with the goals of science learning and with educational goals for the middle school as a whole. It prepares students for the larger world outside the classroom. The more closely instruction and assessment mirror the processes of scientific exploration, inquiry, interpretation, and documentation, the more likely it is that the students will come to understand how scientists think, how science develops and technologies are created, and what their contributions are to society and to individuals.

Striking a Balance

The primary message in this document and it's predecessor report on assessment in elementary school science (Raizen et al., 1989) is that assessment should be integrated with instruction. This works to the advantage of the students. When assessment is integrated with instruction, teachers can use the results in a diagnostic and formative way—to alter their instructional sequences and emphases. They know when the students are confused and they can structure appropriate learning experiences. However, in science instruction as in the assessment of student learning, a balance must be maintained among the various strategies used.

The approaches described above are intended to broaden horizons and present assessment strategies practiced in good science classrooms. However, we urge a variety of approaches. For example, while hands-on and laboratory work is integral to learning science, there will continue to be a place for reading and written work. Cooperative learning strategies have many strengths, but groups and individuals exhibit competencies in different ways, and individuals should also be assessed according to their own particular accomplishments. Finally, because of the reality of giving grades, teachers informal assessments should be accompanied by standardized procedures. In addition to the value of monitoring student activities and the information gained from evaluating products of long-term projects, there is still a place for strong, well-designed paper-and-pencil examinations that include essay questions assessing cumulative learning and attainment of instructional goals.



Chapter VI

Innovative Assessments: New Directions

In the last chapter, we discussed assessment strategies designed to support good instruction. These strategies are available to any teacher willing to move beyond multiple-choice tests and short-answer quizzes. Some educators argue, however, that assessments should go beyond these instructional functions, that they should be learning experiences for students in their own right, and that tests should model good instruction (Sizer, 1986; Wiggins, 1989; Archbald and Newmann, 1988; Feuerstein et al., 1987; de Lange, 1987.). In this chapter, we describe several innovations in assessment that are working toward these new expanded goals for assessment. The first two are examples from other countries, the next few are examples from state assessment programs in the United States, and then we describe an example developed at a university for the purpose of doing research in classrooms. The last section of this chapter discusses the potential of the computer in assessing science learning, based on its role in science instruction.

Innovative Curriculum and Assessment from the Netherlands

The first example of innovative assessment is taken from secondary mathematics in the Netherlands. The reader might wonder what relevance this has for middle-level science programs in the United States. In our view, although many of the surface elements are different, the problems and issues are very much the same. The assessment we shall describe was developed for the Mathematics A program in the Netherlands, a program designed for high school students who would not be mathematics majors but would use mathematics as a tool in their careers (e.g., economics, medicine) and in their lives. Therefore, the primary focus of this curriculum would be its usefulness. In middle-level science programs in the United States, teachers often have very similar goals. They recognize that only a small proportion of their students will become scientists but that a much larger number should have the competence to enter science- related careers and use science in other aspects of their lives. Therefore, they want science to be useful. They are also aware of the disappointing fact that, for many students, their contact with chemistry and physics in the middle school will be their last formal learning experience in these subjects.

As they explored the appropriateness of different strategies for assessing the effectiveness of the Math A curriculum, de Lange and his colleagues (1987) were influenced strongly by Gronlund's (1968) assertion that achievement testing should be a learning aid. After looking at more than one hundred tests used in twelve participating schools, the researchers concluded that, for complex problems like those in the Math A curriculum, the students needed sufficient time to read, reflect, mathematize, and interpret results. Yet, strict time limits were set on commonly used written tests, which placed severe constraints on what could be asked of students. Therefore, these tests incorporated only a limited scope of the goals of the Math A program, making feedback to the educational process minimal and defeating a major purpose of the tests—to help improve instruction.

A recent report, *Everybody Counts* (Mathematical Sciences Education Board, 1989:68) warns: "Tests stress lower rather than higher order thinking, emphasizing student responses to test items rather than original expression and thinking." This captures de Lange's (1987:177) concerns: "Examinations in mathematics which consist only of timed, written papers cannot, by their nature, assess ability to undertake practical and investigational work or the ability to carry out work of an extended nature. They cannot assess skills of mental computation or ability to discuss mathematics, nor, other than in very limited ways, qualities of perseverance and inventiveness." He suggests that these qualities can only be assessed in the classroom over an extended period. He warns that tests lead teachers to emphasize in their classrooms activities that are directly related to the type of questions used in the examination "...which means that practical and investigational work finds no place in day-by-day work in mathematics" in cases where timed tests represent the only method of assessment. As *Everybody Counts* states (p. 69): "What is tested is what gets taught. Tests must measure what is most important."

Assessment Principles Used in the Netherlands

De Lange developed the following five principles for effective assessments, which serve as his criteria for judging different assessment approaches:

1. **Tests should improve learning.** A properly designed test or task should not only motivate students by providing them with short-term goals toward which to work, but also by providing them with feedback concerning their learning progress.
2. **Tests should allow students to demonstrate what they know (positive testing) rather than what they don't know.** Otherwise, students may lose confidence, which should be avoided at all times.
3. **Tests should operationalize the goals of the Mathematics A curriculum.** More specifically, tests must be developed that provide the freedom of response required for measuring certain complex outcomes. These include the ability "to create, to organize, to integrate, to express, and similar behaviors that call for the production and synthesis of ideas."
4. **Test quality is not in the first place measured by the accessibility to**

74

objective scoring. de Lange accepts the fact that competent, independent judges may score differently—but within certain limits.

5. **Tests should fit into the usual school practice.**

Alternative Assessments Used in the Netherlands

The assessments designed by de Lange involve four different strategies that can be combined as appropriate: tasks that are to be accomplished in stages, tasks to be accomplished outside school hours, essays, and oral performance.

The two-stage task. Inspired by the ideas of Van de Blij (as referenced in de Lange, 1987), the two- stage task uses both short-answer and essay questions and provides students with two separate grades, one for each stage. The first stage, essentially a preparation for the second stage, consists of a traditional, time-restricted, written test administered to all of the students simultaneously for completion in class. The students are expected to answer as many questions as possible with an orientation toward having students find out what they don't know rather than demonstrating to the teacher what they do know. Most of the attention is given to the "lower goals" of computation and comprehension. Scores are as objective as they can be under these conditions. The teacher scores the first-stage papers, and hands the tests back to the students with the biggest mistakes (and only those) and the scores disclosed.

In the second stage, the student repeats the work at home using the teacher's feedback. Interactions (e.g., outside advice and library research) are permitted between the two stages. At a designated time, perhaps three weeks later, the students turn in their work, and the teacher scores the tests a second time.

The second stage follows the five principles listed above. The test improves learning; it emphasizes what students do know; it gives attention to the higher goals of interpretation and reflection; it uses subjective but reliable scoring; and, by allowing the students to work at home, it fits with usual school practice.

Three findings are worth noting:

1. There is a relatively wide spread in scores for the first stage, from very poor to excellent. At the second stage, the spread of scores is greatly reduced with more students doing well.
2. Girls perform relatively more poorly than boys in the first stage. At the second stage, this difference disappears. In fact, the best results were scored by girls.
3. Students experienced enhanced self-confidence when they were able to improve in the second stage.

The take-home task. Following a fifty-minute written task, a sample of students were allowed to choose one out of five subjects to work on at home. They could either work alone or in pairs.

The essay task. The students were given a newspaper article on the problem of overpopulation in the Republic of Indonesia. The article contained a great deal of numerical information, but made no use of graphic representation. The students were asked to rewrite the article, making optimal use of graphic representation. This task called on the students to find relevant mathematics in the text, find relationships between different facts, and reflect upon different aspects of both in rewriting the article and developing the appropriate graphs, tables, and charts.

The oral task. Once a standard part of examinations in the Netherlands, this type of assessment has become less popular recently. de Lange reinstituted it in order to study its effectiveness as an assessment procedure for Math A. All interviews took twenty minutes and involved the student, the teacher, an external independent examiner, and an observer. The first question differed from student to student, using questions that seemed appropriate to the expected performance level of the student. de Lange noted that one advantage of the oral exam over all forms of written tests is that one is able to find out how much relevant information a student really needs to start solving an assigned problem. He cites as other advantages the observation that, because of hints provided by the interviewer, the students "never got stuck." On the negative side, some students felt rushed due to time constraints, felt nervous because of the presence of officials, and felt uneasy at not being able to do actual computations.

Some Conclusions

For three reasons, de Lange recommends using a combination of untimed assessment strategies to assess the Math A curriculum. First, de Lange discovered that the correlation between the restricted-time, written test, the take- home task, and the oral test were low, indicating that these tests actually measure different dimensions. (This is consonant with findings by Applebee et al. [1989], on assessing different dimensions of writing competence.) Second, the different testing strategies yielded different patterns of results for boys and girls. Specifically, boys performed considerably better than girls on the time-restricted, written tests; on the stage-two and take-home tasks, boys and girls performed at more or less the same (high) level; and on the oral tests, boys and girls performed more or less the same and at a level between the time-restricted, written tests, and the take-home tasks. Third, the untimed strategies more closely paralleled the goals of the Math A curriculum, which is strongly process oriented, focuses on higher thinking skills, and attempts to enable students to engage in the mathematization process, all of which require time to enable students to engage in reflection and generate creative and constructive thought. Perhaps the most important message of de Lange's most important findingwork, however, is that assessments can be developed that are truly criterion referenced, where the goal is to assess achievement of a student's learning rather than spreading out individual test scores along a predetermined distribution curve.

The Use of Profiling and Moderating Panels in Great Britain

Great Britain currently is developing a new national curriculum in many areas, including science. According to the report of the Task Group on Assessment and

Testing (Department of Education and Science and the Welsh Office, 1987), the emphasis of the new assessments will be on developing profiles of each student on a set of between four and six profile components. For each component, there are twelve attainment targets that have been identified, which are identical for all grade levels tested (ages seven, eleven, fourteen, and sixteen) but take into account the expected growth in knowledge and skills. The assessments proposed by the Task Group are like teachers day-to-day assessments: they are directly concerned with what is being taught, and they are designed to reveal the quality of each pupil's performance irrespective of the performance of other pupils.

> For example for 7 year olds, and largely for 11 year olds, it is proposed that the tests will take the form of topics for children to work on. These will be designed so that they look like interesting pieces of work ordinarily met in class. In the course of doing them, children will be able to display a range of achievements which teachers can assess, by observing children's activity and by marking work—artistic, written, oral—that they produce using standard procedures|Teachers will be able to select tasks from a "bank," choosing subjects and contexts suitable for the background and interests of the pupils involved|because children are more likely to do full justice to themselves in contexts which are familiar and interesting to them (pp. 11-12).

The results of these assessments can be used by teachers for their instructional and evaluative purposes and will also be aggregated at the school level. To ensure comparability, teachers will use "moderation" meetings with teachers from other schools to discuss the progress of their groups of children, including consideration of the spread of results from the national tasks compared to the spread of results from their own assessments. The final responsibility for decisions about the progress of individual pupils will rest with their teacher.

Performance Assessment in California

Since 1989, California has been field testing for eventual incorporation into its state assessments programs a series of performance tasks and open-ended questions emphasizing the key concepts and principles of science for grades six and twelve. Performance tasks focus on science process skills embedded in the content areas of the life, physical, and earth sciences. The open-ended questions focus on engaging students actively in the use of hypotheses and the design of scientific investigations and processes, as well as providing opportunities to respond to societal and ethical issues related to science.

In the spring of 1990, California will pilot a new testing program, the California Golden State Examinations. These are optional tests that students may elect to take in order to receive a special endorsement of their diplomas. In chemistry and biology, these examinations will include a combination of different modes. The multiple-choice questions will be conceptually based (thirty minutes): the open-ended questions will ask students to respond to a prompt by interpreting or entering data on a chart or graph, drawing a picture to answer a question, or writing a short analytical paragraph (ten minutes); and performance tasks will use a laboratory setting (fifty minutes).

California's Assessment Program (CAP), intended to probe student achievement statewide, has identified nine overlapping characteristics of a desirable assessment program:

- **Emphasis on production rather than discrimination.**
- **Modeling good instruction.** Tests should be mirrors of instruction and assessments should provide a learning opportunity in and of itself.
- **Focus on integration.** The tasks should be multidimensional in skills assessed, multisensory in stimuli presented, and multimodal in response formats.
- **Fewer tasks of greater depth and breadth.** The right kind of exercises would take considerably longer than normal multiple-choice questions, hence that portion of the test (assuming also a short-answer portion) would consist of a relatively small number of tasks.
- **Interdisciplinary learning and assessment.** Complex multidimensional tasks would cut across disciplinary lines, providing opportunities for students to write about science, and tell how they would solve a social problem. Instruction and assessment would focus on large real-life problems, such as deforestation or hunger.
- **Communication.** Exercises would demand demonstration of how clearly the students could communicate learning.
- **Face validity.** The tasks should be credible to the teachers, parents, and students.
- **Learning and assessment in groups.** The ability to interact, negotiate, and cope with different opinions to achieve common ends should be part of the assessment.
- **Renewed emphasis on speaking and listening.** Oral examinations could take several forms, for example, student debates, peer problem-solving sessions, examinations of small groups that have done research together, or examinations of individual students.

California is also experimenting with collecting portfolios of the students' work in mathematics. In the spring of 1989, fifty-five teachers teaching grades three and six were asked to collect their students' work over two or three months. Each portfolio included three or four pieces of individual work, one report on a group project, and a reflective or imaginative piece that asked the students to reflect, in writing about the work done in mathematics class. Subsequently, the teachers met to review each other's portfolios, compare criteria for assessment of their students' work, and exchange both teaching and assessment ideas. The experiment will be repeated during 1989-90, with more teachers participating.

Also in mathematics, the California State Department of Education (1989) has been including open-ended questions in its most recent twelfth grade mathematics tests. The purpose is to provide the students an opportunity to think for themselves, construct their own responses, demonstrate the depth of their understanding, and encourage the students to solve problems in several ways. For 1987–88, the first year that open-ended questions were included in the test, a random sample of 2,500 responses (500 to each of 5 questions) out of a total of 240,000 responses were reviewed by a special committee. Performance on a large percentage of the responses (well over half for all but one of the problems) was rated as inadequate. The committee members surmised that the results indicated that students were not used to writing about mathematics and had little experience in reflecting on or describing their thought processes as they solved mathematical problems.

Performance Assessment in Connecticut

In summer 1989, the Connecticut State Department of Education received a grant from the National Science Foundation to work collaboratively with the Coalition of Essential Schools and the state departments of education in Michigan, Minnesota, New York, Texas, Vermont, and Wisconsin to develop performance assessments in science and mathematics. In August, 1989, three dozen high school teachers from these states met and formed the Connecticut Multi-State Performance Assessment Collaborative Team (CoMPACT) to develop performance tasks that would be tried out in their classes during the 1989-90 school year. Criteria for developing effective tasks included the following:

- **The tasks should be based on essential rather than tangential aspects of the curriculum.** They should represent "big" ideas or significant themes.
- **The tasks should be authentic rather than contrived.** They should use the processes that scientists or mathematicians use, and the outcomes of the tasks should be of value to students.
- **The tasks should be rich rather than superficial.** They should cause the students to raise related questions, consider other problems, and make new connections.
- **The tasks should be engaging.** They should be thought-provoking and foster persistence.
- **The tasks should require the students to be active rather than passive.** The students should construct meaning and deepen their understanding as they solve complex problems. On a subset of tasks, the students would be encouraged to work collaboratively with other students.
- **The tasks should be integrative rather than fragmented.** The students should be expected to bring together many separate pieces of knowledge in the completion of a given task.

Connecticut will use tasks designed to meet these criteria to assess the progress of students on the goals set forth in *Connecticut's Common Core of Learning* (Baron et al., 1989), an exposition of the state's educational expectations for its students. The tasks will assess students' scientific attitudes and dispositions as well as skills, processes, and knowledge. Toward these multiple ends, the first set of tasks to be developed will focus on sustained group tasks. These tasks could take anywhere from part of a class period to several weeks to complete. It is intended that the students will work together to plan and conduct investigations and solve real-world, multistep problems.

The participants in the summer workshop recognized and readily acknowledged that effective performance tasks closely resemble effective instructional activities. In fact, one of the goals of the assessment is to model good instructional tasks. However, the component which makes a task appropriate for use in assessment is the existence of accompanying scoring guides. In science, students will be scored on their understandings and applications of scientific knowledge and concepts, as well as on their scientific attitudes and dispositions, their effective employment of the skills and processes of science, their ability to use scientific tools and apparatus safely and appropriately, their ability to work effectively as a member of a group, and their ability to communicate their findings effectively.

The CoMPACT also developed draft criteria for determining whether a performance task is appropriate for group work. Two classes of problems seem particularly well suited. First are those that are too large or too time-consuming for an individual working alone to complete. Related to the jigsaw approach (Aronson et al., 1978), each student in the group would collect different data or do a different piece of research, thereby fostering private independence among the members. Each individual also would have to integrate all of the various pieces so as to be able to contribute to the completion of the group task. This approach has been successful in enhancing the self-concepts of the students who come to see themselves as indispensable to the work of the group, and it often raises their level of performance accordingly (Aronson et al., 1978). A second category of appropriate group tasks are those to which each individual brings only a partial understanding of the scientific phenomena under consideration. Working together and sharing ideas has the potential to deepen the students' collective and individual understanding (Cobb et al., in press).

Teachers who believe in the value of group work often become frustrated when it comes to assessing the contributions of the individuals within the group. Because of their need to assign grades to individual students and be able to justify these grades, teachers need to have valid and reliable techniques available to them for assessing the achievement of each student. The CoMPACT is currently exploring different scoring approaches to group work that will allow teachers to check for the understanding of individual group members. This will permit the teachers to assess the performance of the individuals while at the same time rewarding the efforts of the group.

A Three-Stage Performance Task

One of the many strategies being considered by the CoMPACT is the development of a three-stage performance task, which was stimulated by de Lange's two-stage testing. In this model, teachers would obtain an *initial assessment* of each student's knowledge and understanding at the beginning of the group task. For example, this initial assessment might call for a written prediction about what might happen during the course of an investigation, with the students asked to provide reasons or related descriptions and explanations of their predictions. Or, the students might be asked for a preliminary design with an accompanying rationale. A second series of measurements of each student's understanding would be taken during the interval in which the group is working. These would be *on-line checks for understanding*, accomplished through such informal means as students journals and logs, oral interviews, and a paragraph turned in at the end of the class. The third stage would occur at the completion of the group work. Each student would be asked to complete independently a *near-transfer or extension task*, example, something closely related to the knowledge and processes used in the group task with an appropriate degree of novelty. If students used the group experience to enhance their understanding of the scientific concepts and principles inherent in the task, they would be able to succeed on related but unfamiliar problems or tasks set in a novel context.

Portfolios in Vermont

Vermont, as California, is pilot testing the use of students' portfolios in its statewide assessment of writing and mathematics in grades four and eleven and is considering the inclusion of portfolios in other areas, including science. According to an August 1989 draft developed for teachers by the Vermont State Department of Education:

> Portfolios will be used to provide data in areas not reasonably addressed through standardized tests. The content of student portfolios in mathematics should reflect evidence of the ability of students to solve both routine and non-routine problems, in both group and individual situations. There should be evidence of a student's ability to communicate and reason mathematically. Portfolios should show student growth in understanding and using connections among various mathematical topics and between mathematics and other disciplines. There should be examples of the student's work in exploring problems and describing results using a variety of models or representations. Reflections on the student's own thought processes in solving problems and on the feelings and attitudes of the student, as well as a self-assessment of strengths and areas needing improvement, should also be included.

The portfolio should contain a few examples of a student's best work collected over a period of more than one year. It might include the following:

- A solution to a problem assigned as homework or given on a test or quiz. The solution should show originality or deviation from the usual procedure, not just a neat set of figures. Several different solutions to the same problem could constitute one entry.
- A problem made up by the student, with or without a solution, depending on the complexity.
- A paper done for another subject that contained some mathematics, such as an analysis of data presented in a graph, particularly if the data were collected by the student.
- A report of some group activity or project, with comments as to the individual's contribution (e.g., surveys, reviews of the use of mathematics in the media).
- A picture made by the student of his or her work with manipulative, or two- or three-dimensional figures as a solution to a problem, or a description of a mathematical concept or situation.
- Art work done by the student involving mathematics, such as drawn designs, coordinate pictures, scale drawings or maps, etc.
- A videotape of a student or a group of students giving a presentation involving mathematics.
- A report on the history or application of some mathematical concept.
- An entry or entries from the student's journal.

During the 1989–90 school year, teachers are being recruited to participate in a pilot study of the use of portfolios in order to provide their reactions, and their students reactions, to generating portfolios. The teachers will note the advantages they see, problems they encounter, uses they made or envision for the portfolios, and suggestions they have for further consideration. They also will keep track of the things they do and the time they spend with the portfolios, recording how often they review the portfolios with the students and the extent to which they use portfolio materials when meeting with parents.

Using Naturally Occurring Problems to Assess Students' Understanding

Linn and Songer (1988) have described a research program consisting of a thirteen-week unit on thermodynamics, which used a series of microcomputer-based laboratory experiments with real-time data collection to collect, record, and instantaneously display laboratory data. The research team used successive curriculum reformulation to deepen eighth-grade students' understanding of the difference between heat, energy, and temperature. In order to examine whether the students really understood the role of the different variables involved (e.g. starting temperature, volume, surface area, and insulation), the team developed a set of transfer tasks. Even though the students understood the effects of the variables in laboratory settings, they had considerable difficulty applying their understandings to a series of graded, natural-world problems. The specification of appropriate transfer tasks represents an innovative approach to assessment because it takes into consideration single- and multi-variable situations and the degree of similarity between the naturally occurring problem and the laboratory tasks in the curriculum. Using problems in assessment that occur naturally in the world outside school may have the great benefit of developing models for instructional activities designed to "teach for transfer," a term that has found many advocates but little application because of the lack of cogent exemplars.

The Potential of Computers for Assessment

In science classrooms, computers have been used in four major ways (see Guertin et al., 1987:Abeles, 1989:Linn, 1988), and each use has implications for assessment. First, department heads and teachers have used computers to assist them in classroom management activities. These include keeping inventories of materials, budgeting, computing students' grades, and preparing tests. Second, they have been used in various ways to assist instruction, including drill and practice, tutorials, simulations, and research (e.g., databases, spreadsheet analyses of data, and word processing for report writing). A third category involves telecommunications, with students from different physical locations inputting and sharing data on a phenomenon or common problem. Fourth, computers are used in microcomputer-based science laboratories as a way to collect, portray, and analyze real-time data.

Recent advances in computer hardware have made possible quite sophisticated instructional and assessment activities. The use of the microchip has reduced costs and processing time while increasing computer memories. Optical laser disk technologies combine the power of the computer with the remarkable storage capabilities of laser disks. One of these disks, the CD-ROM, is only about five inches in diameter and stores 270,000 printed pages of information. This means that one or two CD-ROM disks can provide an entire semester of study consisting of text and assessment materials; slides and movies can be computer controlled and integrated with the materials stored in the CD-ROM disks. Obviously, at the current time, it is not the technology that limits its applications to science education but the failure to generate software and learning opportunities to take advantage of the existing capabilities. Other impediments to widespread use involve the cost of computer hardware and the training necessary to enable teachers to feel comfortable using computers in their classrooms (Cohen, 1988).

In this section, some possible uses of the computer for monitoring what students know and can do in science are described. The perspective taken is that of the classroom teacher. However, the data generated through computer-based assessment can potentially be aggregated for policy makers at the school, district, state, national or even international levels. We discuss four possibilities: item banks, simulations, telecommunications, and microcomputer-based science laboratories (MBLs).

First, some teachers are currently using computers to build item banks for their unit tests and final exams. The format of these items usually is multiple-choice. Some teachers keep item statistics to see how well students do on individual items over a period of time. In some cases, teachers even develop and print alternative forms of their tests for use in classrooms where students sit close together. The promise of this use lies in the potential for exchange and quality control of items among teachers, provided the items are openly available.

Second, the use of simulations makes possible a very different kind of learning and assessment. Simulations can be classified into two types: whether the student plays an active or passive role. *Passive simulations* are like teacher demonstrations in that the student observes the scientific phenomena. The advantages of using a computer are several: that it can substantially reduce—or elongate—the time it would take for phenomena to occur; it allows students to observe phenomena that would require expensive, unwieldy, unavailable, or dangerous equipment; it can enlarge or reduce the scale of phenomena to make them observable in the classroom. For example, within a part of a class period, the students can observe many generations of genetic offspring, ecosystems with prey and predators, or geologic or astronomical phenomena covering vast regions and taking thousands of years to occur. Also, the light and sound emanating from the computer often provide a more motivating learning situation for some students than the same material presented in textbooks or classroom lectures. Furthermore, such simulations may deepen the understanding of students about how things actually occur, making possible more complex analysis and evaluation activities. The simulations can become the stimuli for assessments that ask students to explain the phenomena and make and justify predictions about related phenomena.

Active simulations create a hands-on environment that provides an opportunity for students to manipulate many of the variables involved in scientific phenomena and to observe their effects. For example, computer programs exist that simulate the operation of a nuclear power plant. Using control rods, the student can control the amount of heat generated by the reactor and thus the amount of steam formed and electricity generated. In mechanics, simulations exist to teach students about combining gears to perform desired functions. Thinkertools is an example of an environment created to teach students many of the more difficult and abstract concepts related to the laws of motion (Raizen et al., 1989). In assessment contexts, the students can be asked to solve complex problems involving the manipulation of variables in these simulated environments. Different levels of abstraction, transfer, and application to real-world contexts can be incorporated in the assessment problems. The technology currently exists to track students' thinking by programming the computer to keep records of the strategies students use to try out their solutions.

This permits the teacher to monitor the understanding of the students—to see whether their problem-solving strategies are deliberately constructed or random trial-and-error.

Third, telecommunications permit students from different locations to work together on a common problem. Sometimes this entails studying the effects of environmental variables on natural phenomena. For example, one middle school application has students in different countries at different altitudes reporting the temperature at which water boils. A second one involves students at many points along a river taking water samples and reporting the data from the analyses of these samples to one another. In some sense, the students are functioning like managers of databases (Guertin et al., 1987), allowing them to discover relationships, note trends, and form hypotheses. These databases often challenge the students to investigate a topic further and report findings back to the class and to others on the telecommunications network. In the event, the students are motivated to use their textbooks, their teachers, and other experts in their search for understanding. The assessment opportunities within these learning events are unlimited in that teachers can monitor virtually any combination of the students' scientific understandings and research and communication skills. (See March 1987 issue of *Classroom Computer Learning*).

Fourth, the use of sensors or probes in microcomputer-based science laboratories (MBLs) is particularly well suited for science education, because it allows the students to conduct hands-on investigations, with the computer assisting in gathering and presenting data. The students set up the apparatus and perform manipulative operations just as in a traditional laboratory, but the data are presented in graphic form as they are collected in real time. The students see the data displayed on a graph or table as they are collected and see relationships as they happen. According to Abeles (1989), there are several advantages to MBLs. Data collection is less tedious, more accurate, precise, and efficient. And data can be gathered about phenomena not readily available before (e.g., reading the thermometer inside the freezer section of a refrigerator every three minutes for twenty- four hours).

These motivating and engaging qualities of MBLs give the students the opportunity to gain a broader perspective on what is taking place in an investigation and enable them to pay much more attention to the phenomena and the concepts being studied rather than, as some teachers have put it, "getting lost in the data." For example, according to Guertin et al. (1987:6-7):

> ...While the MBL product is continuously measuring temperatures during a cooling curve experiment, students can watch the sample rather than the thermometers and observe that the crystallization coincides with the temperature 'plateau.' Students have time to detect an effect and look for its cause. They are motivated to seek explanations for the relationships they observe. For example, should the temperature drop suddenly during an experiment, students would be alerted immediately and might discover that a draft had been created from an open window. Using traditional methods, the students might not have detected the data anomaly until they graphed the data after the laboratory session.

Many researchers have found that real-time graphing increases the students' ability to interpret graphs (Mokros, 1986; Linn, 1988). Guertin et al. (1987) note that an entire experimental cycle can be completed within the attention span of the student because the time between data acquisition and data analysis and display is brief. For this reason, MBLs have great potential for use in investigative science with younger students. Faster data acquisition also allows the students flexibility to repeat the experiments, explore more cases of an experiment, and do more experiments. Furthermore, many students are motivated to ask "what if" questions and change the parameters of the investigation. In addition, data can be collected over periods of time longer than the school day, extending from several hours to overnight to several days or weeks. Finally, the students can see the same data displayed and printed in a variety of tabular and graphic formats, and they are able to analyze it at a later time.

Once the students become familiar with the technology, they are free to ask their own questions, generate their own hypotheses, and explore them in their own way. They control both the nature and the pace of their experimentation. They collect their own data and portray it in the graphic mode that seems best for their purpose. They are encouraged to verify, replicate and make sense of their data, use a variety of approaches, and communicate their findings to others. This represents an extension of the opportunities inherent in the active simulations previously described, because the use of the data-gathering probes allows for the students to collect data in the real world. The students are free to explore their environments, manipulate variables, and observe outcomes. They have an opportunity and increased motivation to explore a phenomenon deeply. This new technology should also enhance transfer of learning and notions about the relevance of science.

Furthermore, scientific dispositions are likely to come to the fore in the use of MBLs. MBLs enable the students to feel like scientists. The students come up with problems and collect real data. They then portray the data professionally. They also have opportunities to work collaboratively, and the classroom can become a community of inquirers. In this way, the students can develop and display many of the attitudes, attributes, dispositions, and habits of scientists.

Applications of sound and light probes exist in all branches of science, often making visible those processes that previously could only be read about. For example, body functions and reactions can be studied by viewing and analyzing heart and respiration rates, skin resistances and temperatures, electrocardiograms and electromyograms. Even brain waves can be viewed and recorded. In physics, photogate probes are used for measuring velocity and acceleration, sonic transducers for measuring distances, strain gauges for measuring force, thermistors for measuring temperature. Chemical reactions can be studied through colorimetric or potentiometric techniques. As noted by Guertin et al. (1987), the ability of the computer to act in place of voltmeters, freeze-frame oscilloscopes, thermometers, pH meters, light meters and a host of other laboratory instruments, coupled with the fact it allows large quantities of data to be quickly collected, organized, stored, graphed, and analyzed, means that it might be viewed not as an expensive tool but rather as the "best buy" in town.

Guertin et al. (1987) note that one of the strongest arguments for the use of the MBL in school science classrooms is that college laboratories, research institutions, medical laboratories, and industry now use digital data acquisition devices rather than traditional laboratory methods. Hence, experience with the MBL in school laboratories will prepare the students for advanced MBL activities in higher education and in scientific and technical occupations. We would urge, however, that simultaneously with such experience, students at the middle level also develop a strong foundation in understanding how basic measurements are made and the various uncertainties attached to different types of measurements, as discussed in chapter 3.

The assessment opportunities arising from the application of MBLs are limited only by the teacher's time and inclination to make use of them. The records kept by the student can provide a rich base for assessing their operational and conceptual knowledge as well as their thinking skills. Problems of some complexity and sophistication can be developed, which can be addressed in a classroom period (or over several periods) using MBL-generated data. Social skills can be assessed as groups conduct inquiries using the MBL. Almost every strategy for assessing science that goes beyond paper-and-pencil, short-answer formats can benefit from the use of the MBL.

Chapter VII

Assessments and Policy

The emphasis in the preceding two chapters was mainly on improving assessments carried out by the classroom teacher in support of good science instruction and to evaluate the students' learning and performance. In this chapter, we take up issues related to assessments carried out for broader policy purposes. Indeed, educational administrators, school board members, legislators, and other educational policy makers are increasingly turning to tests for information to assist them in monitoring outcomes, setting goals, allocating resources, and, most important, holding districts, schools, and even individual teachers accountable for the learning of their students. Currently, most of the tests given for educational policy purposes are entirely separate from the ones that teachers select or create to use in their own classrooms. They are often referred to as *externally mandated* tests, because they are administered under the aegis of some authority beyond the classroom, either within or outside of the educational system. Tests used to inform policy and increase accountability are given on a much larger scale than classroom tests, often to all of the students in a district or state, to nationally representative samples, and sometimes even across international borders. They are given less often than classroom tests and generally sample the contents of a year or more of the curriculum.

Externally mandated tests often focus on critical transition points in the schooling process, including the middle-level years. Thus, testing during eighth grade is ubiquitous, occurring at the district, state, national, and international levels. Because they are almost always used for making comparisons among the classrooms, schools, districts, states, or nations tested, these externally mandated tests are carefully standardized and must be given at a specified time regardless of the pacing of instruction by a particular teacher.

Both the students and teachers are likely to regard externally mandated tests as less important than classroom tests. The standardization and breadth of coverage of these tests usually mean that they are less closely tied to the curriculum than classroom tests, and the students are often told that their report-card grades will not be affected by how well they do. Many external testing programs employ elaborate safeguards to protect the anonymity of teachers and schools, as well as of individual students. Nonetheless, these tests can have major direct and indirect effects on curriculum, instruction, and learning, and they merit close attention by anyone concerned with science assessment.

High-Stakes Testing: Impacts for Good or Ill

To understand the importance of externally mandated tests, consider the following examples:

- A principal reviews the annual test score means for each of the school's fifth grade classrooms;
- A school board compares school means on the district-wide tests;
- The local newspaper publishes average scores, by school, on the state assessment;
- Legislators anxiously monitor the press coverage concerning the "Wall Chart" put out by the Secretary of Education in Washington;
- The National Assessment of Educational Progress (NAEP) introduces state-by-state comparisons, and both educators and political analysts ponder the potential for good or ill;
- International comparisons show eighth graders in the United States near the bottom of the countries tested in science and mathematics knowledge, and educational leaders respond with calls for increased education funding, better teacher training, and more rigorous curricula.

When policy makers or the public begin to use test scores to make comparisons and judgments, the scores become important in their own right, and the testing becomes "high stakes." At all levels of the educational system, the test scores become a factor as decisions are made about budgets, textbooks, curriculum frameworks and guidelines, and ultimately about the ways that students spend their time in classrooms.

Teachers might understand very well that "less is more"—that true scientific literacy and critical thinking would be helped more by a deep, extended, multifaceted treatment of a few topics than by a superficial survey of many topics. But the teachers who teach a few things well run the risk that their students will be unfamiliar with most or even all of the questions on high-stakes tests that sample factual knowledge on dozens of topics. The best textbooks for improving scores on such tests may be those that cram in the most content, at whatever cost to depth of understanding. Classroom time spent on generating questions, planning experiments, learning to observe and record, and learning to discuss the gains and losses involved in alternative solutions to technological problems might do little to improve test performance.

If educational administrators and policy makers, parents, and the public insist on treating high scores on tests of factual knowledge as ends in themselves, scientific literacy and critical thinking will suffer.

Despite their possible negative effects, externally mandated tests also can be a force for the improvement of science learning. In the short term, even poorly designed tests can focus public attention on the need for educational reform and increased financial support. By showing what is possible in the best schools or under the best conditions, tests can help raise the sights of educators and policy makers elsewhere. Of course, much greater benefits could be expected from sound, comprehensive tests providing valid information about the full range of intended learning outcomes. Such tests could guide the allocation of educational resources to areas of greatest need and could help in the formulation of specific goals for improvement at the classroom, school, district, or state levels. They could also be used in large-scale evaluations of alternative curricula, instructional practices, and educational policies. Finally, valid and comprehensive tests could illustrate for teachers and students the kinds of outcomes and levels of attainment expected.

Measuring the Full Range of Learning Outcomes

Most externally mandated testing programs, even when done on a sample basis, involve large numbers of students. Therefore, it may appear prohibitively expensive to employ testing formats other than paper-and-pencil tests with multiple-choice items. In the long run, however, the costs of *not* employing a broader array of testing formats and response modes may be even higher. The only way to minimize the risks and maximize the benefits of high-stakes testing is to assess a full range of important learning outcomes. In middle-level science, this is likely to require that the students use scientific apparatus and that they respond to some kinds of questions that call for open-ended responses, rather than a selection among a small, fixed set of alternatives. In addition, even within the constraints of written, forced-choice tests, there may be room for substantial improvement in the range of learning outcomes measured.

Alternative Testing Materials and Response Formats

Most of the research on alternative testing formats, forced-response versus essay tests, for example, has shown that different kinds of tests tend to rank order students in about the same way. Such findings have been used as a justification for continued reliance on relatively inexpensive testing formats. If tests that are more costly to administer and the scores yield no more information than inexpensive forms of tests, why use them? There are two reasons.

First, as discussed above, testing is *reactive*. The educational system can change in response to accountability mechanisms, including tests. Teachers and students will look to the test content for messages about the forms of learning outcomes expected, and curriculum and instruction will evolve in the direction of greater emphasis on those outcomes tested.

Second, both logic and empirical research affirm that some of the most important outcomes of middle-level science education *cannot* be measured adequately using paper-and-pencil, multiple-choice items. Frederiksen (1984) has investigated students' ability to pose plausible hypotheses to explain patterns of experimental findings. The format is simple: An experiment is described and its results are presented, often with the aid of a simple graph or figure. The students are then asked to pose as many reasonable explanations for the findings as they can, and their responses are rated for quality. After developing and validating this "formulating hypotheses" test, Frederiksen attempted to create a multiple-choice test measuring the same ability. He and his associates found that the multiple-choice version of the test failed to measure the same abilities as the free-response version. Similarly, assessing the quality of a student's writing according to several different constructs of writing performance yields scores that, while correlated, appear to capture different competencies. Moreover, the parameters of the tasks required in a test constrain writing performance (Applebee et al., 1989). Even without appealing to empirical research, it is clear that the ability to design simple experiments or use scientific apparatus safely and correctly will be difficult or impossible to test fully using multiple-choice questions.

A recent science assessment of fourth graders in the state of New York demonstrated the feasibility of large-scale testing using simple apparatus. A series of stations were set up in each classroom; at each station the students were to use a ruler, a simple pan balance, or other equipment to answer questions on a test they carried with them from station to station in five-minute rotations. The skills being assessed included measurement, prediction based on observation, categorization, inference, and forming hypotheses. The students' scores were recorded only at the school; the school's scores were reported at the state level in terms of percentage of item difficulties. The assessment may have fallen short of testing all the forms of scientific reasoning that might be hoped for even at the fourth grade, but it did yield very significant information about the science program in New York schools and particularly about the students' limited exposure to hands-on science. The more ambitious plans for state assessments, including performance items and more extended exercises being formulated in California and Connecticut, were described in the preceding chapter.

Better Multiple-choice Questions

Multiple-choice questions are often thought of as testing no more than factual recall, but they have been used successfully to measure a much broader range of outcomes. On science tests, for example, a multiple-choice item might pose a scientific question and then describe several different experimental setups or procedures that might be used to investigate it. The correct answer is the one in which experimental and control groups differ only with respect to the matter at issue. This multiple-choice item format tests the important conceptual understanding Piaget termed *controlling variables*, which is often presented in middle-level science as a fundamental principle of scientific method. Multiple-choice questions can test more complex kinds of reasoning, as well. Imagine an item describing the history of a rock formation and presenting a diagram showing different geological strata identified by letters or numbers. Students' understanding of basic principles might be assessed by questions asking about the order in which the strata were laid down, where one might look for particular kinds of fossils, or which of several explanations accounts for some stratum not being horizontal. Obviously, other questions about the same figure might test vocabulary or other recall forms of knowledge.

These examples illustrate that multiple-choice items can be constructed to measure reasoning and understanding of scientific method. Note, however, that such understandings are best tested in the context of an actual problem, in conjunction with relevant factual knowledge. When individuals reason, they reason *about* something. It follows, as we noted in chapter III, that poor performance on such items can be due to deficient understanding of either the processes of scientific reasoning called for or the factual information required to apply that reasoning in a particular situation. (Of course, poor performance can also reflect poor motivation, failure to understand test instructions, poor reading ability, or other causes.)

These examples also illustrate that multiple-choice items testing higher order thinking skills will almost always require the presentation of more elaborate stimuli than most questions measuring factual recall. More text, figures, charts, graphs, and diagrams will be required to describe the problem situation the students are asked to reason about. As a result, such test questions may require a higher level of reading ability than questions that test knowledge of facts and principles, as well as the ability to interpret graphs and other kinds of displays. They may also call for greater effort, attention, and motivation on the part of the test taker. Any of these additional requirements might serve to lower some students' scores (that is, percent of items answered correctly versus expectations, depending on the conditions of testing.)

Multiple-choice questions employing more complex stimuli will also take longer to answer so that fewer items can be administered in a given period of time. For all these reasons, reliable and valid multiple-choice tests will be more difficult to construct for higher order thinking skills than for factual knowledge. In addition, items on such tests will tend to be harder for the students to answer and harder for instructors to teach to, and so some students and teachers might resist movement in the direction of testing higher order skills. *It follows that significant improvements in externally mandated tests, including multiple-choice tests, are unlikely unless concerned parents, citizens, educators, and curriculum specialists insist on better tests, measuring a broader range of important learning outcomes.*

Information Needs of Decision Makers: Achievement and Context

Validity should be regarded as a property not of tests, or even of test scores, but of test-score interpretations. Valid interpretation of achievement scores, for individuals, classrooms, schools, or larger aggregations, always requires supplemental information about at least some of the many factors known to influence achievement. Teachers using tests in their own classrooms bring a wealth of background information about individual students' earlier performance levels, interests, and other characteristics, as well as knowledge about their own curriculum and instruction. This information enables them to set reasonable expectations for achievement levels and evaluate the plausibility of alternative explanations for low or high scores. Even more important, it enables teachers to make better use of scores when deciding what to change to improve learning, for individual students or for the class as a whole.

Assessment for Improvement

Valid interpretation of scores on externally mandated achievement tests likewise requires contextual information if policy makers are to use the results to improve the students' achievement. For externally mandated testing programs limited to a single school or a small school district, decisionmakers might already have access to sufficient information to interpret the scores appropriately. School principals and district personnel would need to know or find out about the curricular goals, textbooks and other instructional resources, and teaching practices in the classrooms tested; the students' performance in prior years and in other content areas; and something about the communities served by the different schools where the tests were given. Depending on the focus of the testing program and the particular score interpretations intended, additional information might be required. Policy makers might want to look in greater detail at the students' opportunity to learn the facts or concepts covered in each test question. They might ask about the teachers' formal training in science, or about the number of years the present textbook series has been in use. Without such information, it is impossible to say from achievement levels alone which teachers or schools are doing poorly and which are doing well. One school may be doing extremely well in the light of its students' language backgrounds and other contextual factors, and still have lower scores than a mediocre school serving more advantaged learners. Another consideration is the match between the curriculum and the test. Where achievement is poor, contextual information is essential to determine to what extent the content of the test corresponds to the curriculum. (Of course, good correspondence merely indicates that the test is appropriate to the curriculum, not that the test—or the curriculum, for that matter—reflects good science instruction.) These examples could easily be multiplied. As the scale of testing programs increases, it becomes less likely that decision-makers will have the needed contextual information at hand. Thus, it becomes increasingly important to collect it in conjunction with achievement-test data. This is most often accomplished by giving the students Sbackground questions" in conjunction with test items, and by providing separate questionnaires to be completed by their teachers and by knowledgeable personnel at the school level.

Baron and Forgione (1989) discuss the kinds of background information it is useful to collect in large-scale assessments. Discipline is required when assembling such questions—Sit would be nice to know" is not a sufficient justification for using precious questionnaire space and testing time. They recommend that all background questions at the student, teacher, and building levels should satisfy at least one of three criteria for inclusion: First, information must be collected on demographic factors that will be used to organize and report the results. These include questions on gender, racial and ethnic identification, socioeconomic status, and other demographic characteristics. This sort of information is important for understanding how educational resources and classroom processes—that is, the opportunity to learn—as well as student achievement and attitudes are distributed across different groups of students. Second, information should be collected concerning schooling factors known to influence achievement, including indicators of classroom process. Finally, background questions can be included, because they assess educational outcomes important in their own right, apart from academic achievement. These include questions about attitudes, beliefs, and behaviors. In an eighth-grade science assessment, for example, students in Connecticut responded on a scale from "strongly agree" to "strongly disagree" to statements such as "Careers in science are more appropriate for men than for women" and "My knowledge of science will be of little value to me in my day-to-day life" (Baron and Forgione, 1989:189). As a further measure of attitudes, the students were also asked how many years of high school science they expected to take.

To learn something about instructional practices in science, younger children were asked in separate questions if they had ever used a magnifying glass, a metric ruler, a thermometer, or a magnet in science, and whether they had ever made a simple electrical circuit or an electromagnet. Eighth-graders were asked on a scale from "never" to "more than ten times" how often they had used a triple-beam balance, a graduated cylinder, or a microscope, and how often they had set up an electrical circuit. The eighth-grade assessment included actual use of a triple-beam balance to weigh an object. Responses to the experience question were strongly related to success on the performance task. (See also the discussion above on the more extensive test of manipulative skills administered to all fourth graders in New York).

Teachers and principals can be asked parallel questions about use of equipment and so forth, as a check on the studentsU responses. In the Connecticut Assessment of Educational Progress there were also specific questions about the availability of good science teachers, the amount budgeted specifically for consumable science supplies, amount budgeted specifically for purchase of new science equipment, and numbers of microcomputers available for science instruction. Respondents were also asked to rate the seriousness of such problems as "a general belief that science is less important than other subjects," "out-of-date teaching materials," "lack of materials or equipment," "inadequate budget for science," "lack of student interest in science," "lack of teacher interest in science," "teachers inadequately prepared to teach science," "lack of support of administration," "teachers views not incorporated into curricular decisions," and "lack of opportunity and/or support for in-service." Teachers also reported on whether science equipment was available to them and, if so, whether they had to share it. In addition, they were asked to indicate how well trained they felt they were to teach science (Baron and Forgione, 1989:206-210).

Similar kinds of questions have been asked in connection with the science assessments conducted by NAEP, by several major studies conducted by the National Center for Education Statistics (Schools and Staffing Survey, National Education Longitudinal Study of 1988), and in the 1985-1986 National Survey of Science and Mathematics Education (Weiss, 1987).

In short, we suggest that policy makers need to understand and document the status of some of the conditions that influence what students actually learn in school. According to Oakes (1989), the following three categories of variables are important to examine:

Access to Scientific Knowledge. Availability of instructional materials, laboratories, computers, and equipment; teachers' qualifications and experience in science; scheduling (for example, departmentalized, discrete classes, or interdisciplinary teams); classroom assignment practices (grouped by ability or mixed instructional groups) and the curriculum associated with each group; availability of academic support and enrichment programs (tutoring, after-school remediation, science fairs, field trips, museum programs); and parental involvement in science instruction or science activities.

Press for Science Achievement and Participation. Opportunities for school-wide recognition of science participation and accomplishments; curriculum and instructional activities focused on challenging, real-world scientific concepts and problems; faculty beliefs about the students' ability to learn science (for example, whether all students are capable of learning science); faculty emphasis on science as an interesting and important subject for students at the middle level; instructional leadership in science—the extent to which a significant person or group at the school advocates and supports science curriculum and instruction; and the degree to which noninstructional constraints interfere with science activities.

Professional Conditions for Science Teaching. Teachers' salaries; teachers' student load and class size; clerical support staff available for noninstructional tasks; time available for professional, non-teaching work; time spent on collegial goal setting, program planning, and instructional improvement; participation on the staff in school-wide decision making; administrative commitment and involvement in staff development in science; and administrative support for professional risk taking and experimentation.

As with student outcome measures in science, assessment of several of these important program characteristics must in part rely on human judgment. Obviously, assessments of science programs cannot possibly provide the complex data researchers need in order to understand fully the relationships among program characteristics and science outcomes. They can, however, provide useful clues to policy makers about strengths and problem areas. The challenge is to design assessments that provide the most central information with a parsimonious set of indicators. Relevant kinds of questions about programs, asked in conjunction with student assessments that involve (1) multiple-choice questions testing both lower order and more complex reasoning skills, (2) hands-on performance exercises requiring actual use of science equipment, and (3) assignment of problems involving extended and in-depth work can begin to illuminate the complex questions facing educational decision-makers. If large-scale assessments are to help inform policy and ensure accountability, *both* the tests themselves and the background information providing a context for interpreting test performance must be sound, reliable, and comprehensive.

Erosion of Validity

Valid interpretation of test results will become more difficult as mandated assessments grow, particularly when they involve high-stakes testing. As noted, validity inheres not in a test itself, but in an intended test interpretation, an inference based on a score. There may be different logical bases for such inferences, calling for different strategies of test design and validation. Consider three examples: A college admissions test, a typing test for applicants for a secretarial position, and an achievement test administered by a state or district. The warrants for using the SAT or similar tests to help reach college admissions decisions include both logical arguments from the tests content and design and empirical arguments from their observed correlations with college grades and other indicators of success. In contrast, the typing test directly samples performances that are a part of the work the person hired will be expected to do. The achievement test probably would be intermediate between these first two examples. To the extent that it directly sampled some domain of proficiencies that the students were expected to acquire, as, for example, use of a thermometer or an equal-arm balance, it would be like the typing test. To the extent that it was intended to show what children were likely to do or be capable of doing in non-test situations, its validity would have to rest on logical or empirical grounds—areas that need much further exploration and work in the case of science tests (Frederiksen, 1986).

Erosion of validity may be said to occur when, as an indirect result of using the test, the warrant for the intended score-based inferences is weakened. In the case of college admissions tests, coaching that concentrates on test-taking skills or practice with feedback in answering multiple-choice items may improve test scores without bringing any concomitant improvement in the complex, developed aptitudes the test is intended to reflect. If such coaching improves the scores of some examinees, the correlation between test performance and subsequent college success is likely to be reduced, thereby eroding the test's validity as a predictor. (Of course, a longer term program of coaching that focused on the underlying skills the test was intended to assess might improve both test performa criterion performance. That would not affect the test's validity.) In the case of the typing test or reading a thermometer, it is more difficult to imagine any kind of training that would substantially improve test performance without also improving criterion performance. A work-sample test is highly resistant to erosion of validity.

When policy actions that affect individual schools, administrators, teachers, or students are taken on the basis of assessment results, assessments become very important. Teachers are more likely to teach to the test, and it is naive to ask them to avoid doing so. Thus, the issue from an assessment perspective is to improve the quality of such tests so as to make instruction based on their content worthwhile.

Similar cautions are in order with respect to some conceptual or background questions and educational goals. If background questions are interpreted as indicators of educational quality, they may be subject to the same erosion of validity as cognitive questions. This may happen whenever the answers to questionnaire items are treated as ends in themselves. Based on the illustrations just given, for example, a well-intentioned but misguided teacher might decide to teach the use of a magnifying glass or a triple-beam balance as an isolated skill. Likewise, questions about the number of homework or writing assignments completed may invite the proliferation of brief, meaningless assignments. Questionnaire developers should be sensitive to such reactive effects of background questions and, whenever possible, should word questions so as to discourage treating activities as ends in themselves.

Summary

Policy makers need content-valid outcome assessments set in the schooling context. These assessments must mirror the goals of instruction. They should sample the kinds of hands-on activities and extended problem assignments found in the best classroom science instruction and should also provide information on program and schooling features.

Due to the reliability, versatility, and efficiency of multiple-choice items, such items are likely to continue to play a role in such assessments, but care should be taken that tests call for scientific reasoning and the application of scientific principles, not just factual recall. This is likely to require multiple-choice items with more complex, extensive stimuli than simple knowledge items. It is critical to recognize that some of the most important science learning outcomes may be nearly impossible to test with forced-choice items of any kind. As an instance, students should gain skill in formulating plausible explanations for experimental findings. Testing this skill may only be possible with free-response items requiring hand scoring.

Contextual information about teachers and learners, classroom resources and practices, as well as information about important affective learning outcomes may be obtained using background questions for students and separate questionnaires for teachers and principals, but care should be taken to discourage respondents or test users from treating instructional activities as ends in themselves.

Chapter VIII

Recommendations

The recommendations in this chapter are grouped around six principles which summarize our view of how student learning in science at the middle level should be assessed:

1. Assessment must be challenging and interesting. Classroom, school, and large-scale science assessments must reflect the educational purposes at the middle level and the growth and development of young adolescent.
2. Assessment must reflect science instruction, which itself should reflect the goals for science learning, which in turn should reflect good science. Assessment must include both science knowledge and the laboratory, intellectual, and social skills crucial to the learning and doing of science.
3. Reporting systems should reflect science assessments with fidelity.
4. Educators involved at every level need to understand the new conception of assessment and carry out relevant strategies, and their clients's and audiences' need to understand the purposes and results.
5. Improving the quality of the science program in a school or district requires information on context as well as on outcomes.
6. Further knowledge and new techniques must be created so that assessments of science learning and performance are faithful to the goals of science education and to the nature of science.

For each of the principles, we provide a brief discussion as necessary and then some action steps for bringing about the kinds of assessments that would support good science instruction in the classroom and help policy makers in their efforts to improve science education for young adolescents.

Principle: Assessment must be challenging and interesting. It must reflect the educational purposes at the middle level and the growth and development of the young adolescent.

Discussion

Classroom assessments should be opportunities for students and teachers to discover together how learning has progressed. It should be recognized that students as well as teachers are important users of the information that assessments provide.

Students will find assessments challenging and interesting if, in solving the problems posed, they discover new uses for the ideas and methods they have learned. Performance assessments can be highly engaging and instructive, and they can test learning outcomes that are difficult to measure in other ways. Projects and productions of many different kinds can serve as the basis for performance assessments.

Recommendations

1. Assessments should include some long-term projects that involve the integration of the knowledge, laboratory skills, and thinking and reasoning competencies the students are expected to acquire.
2. Some classroom assessments should call for new applications of the material that has been learned. After learning about environmental adaptations, for example, the students can be asked to "design an animal" to survive in a specified environment.
3. Even though some departmentalization is typical during the middle-school years, some science assessments should be integrated with assessments in other content areas. Oral and written reports can demonstrate literacy and communication skills as well as scientific understanding. Laboratory workbooks should demonstrate growing skill in using mathematics as well as science.
4. When middle-level students work collaboratively on group projects, at least part of the assessment should address the quality of the group's effort. The students should not be asked to work cooperatively but then only be assessed individually.
5. Large-scale assessments should strive to support the efforts of classroom and school assessments.

Principle: **Assessment must reflect science instruction, which itself should reflect the goals for science learning, which in turn should reflect good science.**

Discussion

Both in this report and in the earlier one on elementary school science, we have stressed the need for a correspondence between good science instruction and assessment. Without this interlinking, the growing emphasis on assessment can only serve to exacerbate the current poor condition of science education in U.S. schools.

Recommendations

1. Teachers should be prepared to model the integration of science knowledge, science laboratory skills, and science thinking skills in their instruction.
2. Teachers should integrate assessment and instruction, gathering assessment data as students are engaged in classroom science activities.
3. Teachers should clarify the goals of their instruction, making sure that their students are equally clear about these goals—across the course and in the context of each instructional activity.
4. The boundaries of classroom assessment must be expanded to integrate instructional goals and information, including teacher observations, oral presentations, production of computer and constructed models, drawings, and research efforts in and out of laboratory settings.
5. Teachers should design science instructional and learning activities that incorporate the collection of concrete evidence of learning, including models that the students have built, reports, laboratory logs, computer output, essays, and records of oral presentations.
6. Teachers should plan science instruction and learning activities that incorporate both individual and group tasks. This will provide a wide variety of products to assess progress, both to inform future instruction and to give grades.
7. Teachers should design science instructional and learning activities that provide students ample opportunities to assess the quality of their own work, including encouraging students to keep written journals of their progress towards the learning goals.
8. Teachers should provide opportunities for the collaborative interpretation of the evidence accumulated and perceived by students on their learning and their own judgments and records of progress.
9. Teachers should ensure that this collaborative evaluation of student progress results in a product that can be shared with parents, which gives middle-level students the responsibility for keeping their parents informed about their progress on a regular basis.
10. Teachers should use assessment data to modify instruction and plan future activities.
11. Teachers, principals, and science supervisors should become partners, with principals and science specialists providing regular checks on the effectiveness

of the instruction and the progress of students. Such independent observations of the students' learning provide additional perspective, enhanced opportunity for staff development, and a way to keep principals informed about the school science program.

12. Superintendents have the obligation to support teachers and principals, both in terms of providing the necessary resources for facilitating district goals in science, including appropriate staff development in assessment, and in terms of educating the community and local school board about the strengths of the new approach that integrates instruction and assessment.

Principle: Reporting systems should reflect science assessments with fidelity.

Discussion

It is important that the messages sent to teachers and parents in assessment reports about what is important in science education not be antithetical or contradictory with the message to use instructional opportunities as assessment opportunities. Therefore, if there is a prescribed body of content and skills expected of the students, the teachers need to be able to incorporate them into their curriculum, instruction, and assessment. In that way, there will be a positive correspondence between the data collected at the classroom level and the goals of the school and the state department of education.

In many cases, there is no articulated curriculum that teachers feel a need to follow, and they appear to be free to develop instructional opportunities for students according to their own goals. Sometimes, however, a curious and unfortunate situation occurs in which national or state tests are administered to the students in a school, and the test results are then used to hold the school accountable for knowledge and skills that were not shared with the teachers. If the tests that are administered do not, in fact, represent the school's goals, then the teachers and school authorities should not treat the test results and report them to the community as if that were the case. The same holds true at a state and a national level.

Teachers and administrators cannot control what an external evaluation agency might do, but they should exert their influence in interpreting test results and striving for better assessment. For example, if a nationally normed assessment of school science does not match the state's or a school's goals, the message sent to parents and teachers should not be that the school or state has not succeeded simply because the test results are poor. Rather, the school or state should be able to present a convincing case of how it has succeeded on the goals it has been striving toward.

The reporting implications that follow from this discussion are that schools, districts, and states need to be clear about the goals they are assessing. Then, they ought to give considerable thought to what the observable indicators are (and on which data should be collected) that would provide evidence that their goals have been achieved.

Recommendations

1. Teachers should be involved in developing strategies to gather, analyze, and portray assessment information that will be meaningful to parents, communities, and policymakers.
2. State departments of education should provide technical assistance to enable teachers to gather, analyze, and portray data that will be meaningful to parents and communities. Assistance should also be provided to help teachers and school officials to aggregate and report data useful for policymakers.
3. Districts should help teachers develop alternative report cards in the form of profiling, as distinguished from grading. These sorts of report cards, providing descriptive information about each student's strengths and weaknesses, would be useful for both formative and summative evaluation.
4. National and state agencies should seek ways to aggregate data collected at the school level. This may require parallel and complementary data collection efforts, checks using standardized assessment questions or tasks, or "second opinions" by outside observers to ascertain the reliability and validity of the data collected locally.
5. For certain core learnings, there should be a consensus effort to agree upon assessment strategies and reporting strategies throughout a state. States will need to provide a technical assistance component to ensure that comparable procedures are used for administration of assessment exercises and interpretation and reporting of results.

Principle: Educators involved at every level need to understand the new conception of assessment and carry out relevant strategies, and their clients and audiences need to understand the purposes and results.

Discussion

Higher education institutions that educate prospective teachers and are charged with inservice staff development, associations of principals and superintendents, teachers' groups, associations of parents and school board members, and educational writers all have important parts to play in fostering improved science learning and assessment.

Recommendations

1. Higher education institutions should model, explain, and give teachers the tools to reproduce the principles of learning and assessment described in this report. Some specifics follow:

    - Appropriate emphasis (or a premium) should be placed on fostering a depth of understanding of scientific concepts and principles, in the teachers' own science preparation as well as in what they are expected to bring to their students.
    - Teachers should be taught and have opportunities to practice a variety of strategiesfor monitoring their own level of understanding via individual journals, small- and large-group discussions, and opportunities to compare their own thinking, through discussion and further reading, with that of practicing scientists.
    - Higher education environments should model communities of inquiry in which teachersare encouraged to generate new questions, ask clarification questions, and discuss their tentative hunches and hypotheses with others—both with respect to science subject matter and pedagogy for teaching science.
    - Institutions of higher educationshould foster and encourage persistence and the assimilation of new information and experiences by giving teachers long-term assignments which require revisiting the same concepts, as they will be expected to do with their students. They also should foster and assess the acquisition of the dispositions of scientists, including the stimulation of intellectual curiosity, open-mindedness, and tolerance for ambiguity.
    - Higher education classrooms should provide opportunities and rewards for teachers to do true investigations (in contrast to verification exercises) in which teachers generate and clarify the problem to be researched, develop a strategy for data collection, analysis and portrayal, and communicate their findings to their classmates, their instructors, and possibly to other audiences, including members of the science community. In short, prospective teachers should be taught and assessed as they will be expected to teach and assess their students.

2. Groups of superintendents, principals, and teachers meeting alone and with one another must work to improve the quality of science learning and assessment. Several specific recommendations follow:

    - These three groups should discuss strategies for examining the standardized tests used to assess the middle level science program in order to determine what messages they are sending about their goals in science education.
    - The same question should be asked about teacher-made tests and other information used to determine report-card grades. What besides tests is used to determine report-card grades? What might the students conclude about the relative importance of breadth of knowledge and depth of understand-

ing? How are scientific skills, processes, and dispositions factored into the determination of grades?

- These three groups also should reflect on strategies to be used at the school level to ascertain whether the school's science curriculum addresses what they truly believe to be important for students to learn. Some questions to consider include the following: Does the curriculum as it currently is being delivered produce students who see the relevance of science in their lives? Does it motivate students to take more courses in the biological, physical, and earth sciences in high school? What percentages of students are taking more than the minimum number of required science credits? What additional skills and knowledge would be required of the teaching and administrative staffs in the school in order to design learning and assessment activities likely to address needs identified by the above questions?

3. Parents' groups and school board members should ensure that superintendents, principals, and teachers are free to design better learning and assessment opportunities for young adolescents. Superintendents, principals, and teachers claim that it is the parents and the school boards who want to know how the local school's students compare to students in other, similar schools or across the nation. Sometimes this perception results from the existence of school board policies calling for annual testing in science at the middle level. Parents' groups and school board members should confront the fact that the pressure to perform well on standardized, norm-referenced tests pressures teachers to "cover the curriculum" represented in overstuffed textbooks rather than to provide a set of more time- consuming learning and assessment experiences that are aimed at a conceptual understanding of science.

   We recommend that parents' groups and school board members reevaluate the goals they have for science education, how these goals are to be achieved, and how achievement of the goals will be assessed so as to preserve their intent.

4. Education writers also have an important part to play in bringing about improved learning and assessment opportunities for young adolescents. If the problems and potential solutions described in this report were made available to a public considerably larger than the one likely to read this report, parents and school board members could become more aware of the magnitude of today's science education dilemmas. Education writers can help by carefully examining current testing practices and reporting on their limited significance so that the public will demand that educators try some different approaches to developing learning and assessment opportunities at the middle level rather than using inappropriate and constraining practices.

Principle: Improving the quality of the science program in a school or district requires information on context as well as on outcomes.

Discussion

Improvement of science education at the middle level hinges on an understanding and tracking of the process through which student learning in science as well as other outcomes are produced. This kind of information is available to the teacher for the individual classroom, but not to policymakers at more aggregate levels, unless it is specifically collected.

Recommendations

1. National policy makers should set the tone for assessing the context in which science learning takes place by highlighting national data about essential program characteristics: science program facilities and equipment; teachers' backgrounds and qualifications; curriculum; instructional strategies; and professional teaching conditions in schools.
2. State policy makers should include context assessments among the indicators of science program quality they use for school and district accountability or for triggering program improvement initiatives.
3. State education agencies and the research community should assist in the development of valid and useful measures of essential science program characteristics and schemes for reporting the results of such assessments.
4. State education agencies and local school district administrators should provide technical assistance to schools as they attempt to implement measures of the school context in valid and reliable ways and as they begin to use the results of such assessments to frame improvement strategies for science programs.
5. Local district administrators and school boards must work with parents and the community to help them understand the importance of assessing and reporting information about the context of science programs. They must show the community that such information can highlight problems and provide clues about potential solutions. They must communicate loudly that viewing science test scores in the context of information about science programs can help communities move beyond self-congratulation or hand wringing by providing useful directions for school improvement in science education.

Principle: Further knowledge and new techniques must be created so that assessments of science learning and performance are faithful to the goals of science education and to the nature of science.

Discussion

Throughout this report, we have noted instances where knowledge and understanding are inadequate. Examples drawn from earlier chapters include the extent to which capacity for formal operational thinking can be developed in all young adolescents and the science experiences and programs that enhance such development; measurable attitudes and behaviors that are valid proxies for future engagement with science and application of scientific thinking skills; and identification of policy-mutable program variables that are strongly linked to desired student outcomes for science education at the middle level. Further experimentation with and development of valid assessment techniques that are sufficiently reliable for use in large-scale assessments is urgent. Similarly, better means for collecting relevant program and contextual information must be developed. These needs imply support both for basic research and for development.

Recommendations for Research

1. The National Science Foundation, the United States Department of Education, and private foundations concerned with science education should sponsor research programs designed to investigate how instruction, and what kinds of science activities and content teaching specifically, can help develop formal operational thinking in young adolescents with different backgrounds, competencies, and preceding educational experiences.

2. Interdisciplinary teams of researchers drawn from science education, the relevant science disciplines (that is, those generally included in middle or junior high school science curricula), psychology, and educational measurement should investigate the relationships currently posited among scientific attitudes and behaviors exhibited in school (or reported on a questionnaire) and disposition beyond the science classroom to apply science knowledge and thinking skills and continue one's engagement with science.

3. Federal agencies and private foundations supporting research in education should invest in fine-grained longitudinal studies to establish linkages between science programs and teaching variables and science learning outcomes for different student groups. (This is in contrast to large-scale longitudinal studies which, perforce, have to use gross process and outcome variables.) What is the role of different instructional strategies (hands-on and laboratory work, collaborative group work, long-term projects, oral and written presentations, use of the microcomputer-based laboratory)? What is the role of the textbook, trade books, other written materials, guest appearances by

scientists, and science fairs? How important is parent involvement, and how can it be engendered? How do the effects of these factors vary for girls? Boys? For students from different ethnic and socioeconomic groups? To what extent can science programs, to be successful with young adolescents, deal with subject matter and abstractions important for science learning but far removed from their experiences and ostensible interests? All these are questions that need better information than is available at present, when too much of science instruction continues to be based on unverified practice and opinion.

Recommendations for Development

4. Assessment strategies consonant with the goals of science education and exemplary science in the middle grades must be developed for use both by individual teachers and in large-scale assessments. In particular, the National Assessment of Educational Progress and individual states should attach to each science assessment they conduct and evaluate some experimental assessment exercises that will probe complex and important science learning outcomes not addressable through tests using multiple-choice or other short-answer formats. (See, for example, National Assessment of Educational Progress, 1987). Experimentation should include not only the design of such exercises but also innovative scoring protocols and other rating methods to explore their feasibility and reliability. Attention also needs to be given to cost implications recognizing that the improvement of assessment will require investment of additional resources or redeployment of current spending.

5. Similar experimentation needs to proceed with respect to the measurement of program variables and teaching conditions. We suggest, however, that—unlike the experimentation with better outcome measures recommended above—this experimentation take place separately from the large-scale assessments of student learning. The reason for this separation is that these assessments are already very complicated and cumbersome and therefore not a good vehicle for the careful exploration of how best to track the characteristics of science programs and school conditions that have been shown (through the research recommended in number three above) to be strongly linked to student outcomes.

6. The best of assessment strategies will fail unless supported and adopted by the persons ultimately responsible for the students' development in science—the classroom teachers. We therefore urge that preservice and inservice teacher education materials be developed that empower teachers to carry out assessments that will serve good science education in their classrooms.

References

Abeles, Sig
1989 The macroworld of the microchip or "son of sand" in the science classroom. *Connecticut Journal of Science Education* 27 (1) Fall, Winter.

Alexander, William M., and McEwin, C. Kenneth
1989 *Earmarks of Schools in the Middle: A Research Report*. Gainesville, Fla.: Author.

Almy, Millie, and Genishi, Celia
1979 *Ways of Studying Children, An Observational Manual for Early Childhood Teachers.* New York: Teachers College Press.

American Association for the Advancement of Science
1989 *Project 2061: Science for All Americans.* Washington, D.C.: American Association for the Advancement of Science.

Anderson, Charles W., and Roth, Kathleen J.
1989 Teaching for Meaningful and Self-Regulated Learning of Science in *Advances in Research on Teaching Vol. I*, J. Brophy, Ed. Greenwich, CT: Jai Press.

Applebee, Arthur N., Langer, Judith A., and Mullis, Ina, V.S.
1989 *Understanding Direct Writing Assessments: Reflections on a South Carolina Writing Study.* Princeton, N.J.: Education Testing Service.

Archbald, Doug A., and Newmann, Fred M.
1988 *Beyond Standardized Testing: Authentic Academic Achievement in the Secondary School.* Reston, Va.: NASSP (National Association of Secondary School Principals) Publications.

Armstrong, J., Moyer, K., and Wilkins, J.A.
1986 *50-State Survey of Middle/Junior High School, Science Initiatives and Issues.* Denver, Colo.: Education Commission of the States.

Aronson, E., Blaney, N., Stephen, C., Sikes, J., and Snapp, M.
1978 *The Jigsaw Classroom.* Beverly Hills, Calif.: Sage Publications.

Assessment of Performance Unit
1984–1985 *Science Report for Teachers:* 2–6. Department of Education and Science and the Welsh Office; Department of Education for Northern Ireland. Distributed by Garden City Press Limited, Letchworth, Hertfordshire SG6 1JS, Great Britain.

Baron, J.B., and Forgione, Jr., P.O.
1989 Collecting and Profiling School Instructional Variables as Part of the State-NAEP Results Reporting: Some Technical and Policy Issues. In U.S. Dept. of Education, *Report of the NAEP Technical Review Panel on the 1986 Reading Anomaly, the Accuracy of NAEP Trends, and Issues Raised by State-level NAEP Comparisons.* National Center for Education Statistics. Technical Report CS/89–499:171–212. Available from the U.S. Dept. of Education, Washington, D.C.

Baron, Joan Boykoff, Forgione, Jr., Pascal D., Rindone, Douglas A., Kruglanski, Hannah, and Davey, Bruce
1989 *Toward a New Generation of Student Outcome Measures: Connecticut's Common Core of Learning Assessment.* A paper presented at the American Educational Research Association, Symposium 28.01. Hartford, Conn.: Connecticut State Department of Education.

Blosser, P.E.
1988 *Teaching Science to Middle School Students. Science in the Middle School,* XI(1):25–28.

Botkin, James W., Dimanceseu, Dan, and Strata, Ray
1984 *The Innovators: Rediscovering America's Creative Energy.* New York: Harper and Row.

Bybee, Rodger W. (Chair), Buchwald, C. Edward, Crissman, Sally, Heil, David R., Kuerbis, Paul J., Matsumoto, Carolee, and McInerney, Joseph D.
1989 *Science and Technology Education For the Elementary Years: Frameworks for Curriculum and Instruction.* A report of the National Center for Improving Science Education. Andover, Mass.: The NETWORK, Inc.

Bybee, Rodger W.
1990 *Science and Technology Education in the Middle Years: Frameworks for Curriculum and Instruction.* A report of the National Center for Improving Science Education. Andover, Mass.: The NETWORK, Inc.

California State Department of Education
1989 *A Question of Thinking: A First Look at Students' Performance on Open-ended Questions in Mathematics.* Sacramento, Calif.: State Department of Education.

Carey, Susan
1985 Are children fundamentally different kinds of 1985 thinkers and learners than adults? In S. Chipman, J. Segal, and R. Glaser (eds.), *Thinking and Learning Skills: Vol. 2, Research and Open Questions.* Hillsdale, N.J.: Erlbaum.

Carnegie Task Force on Education of Young Adolescents
1989 *Turning Points: Preparing American Youth for the 21st Century.* Washington, D.C.: Carnegie Council on Adolescent Development.

Case, Robbie
1985 *Intellectual Development—Birth to Adulthood.* Orlando, Fla.: Academic Press.

Cawelti, Gordon
Nov. 1988 Middle Schools a Better Match With Early Adolescent Needs, ASCD Survey Finds. *Curriculum Update.* Alexandria Va.: Association for Supervision and Curriculum Development.

Center for Research on Elementary and Middle Schools (CREMS).
June 1987 Special Report on Middle Schools: A Description of Organizational Structures in Middle Schools and Their Effects on Student-Teacher Relations and Instruction. Baltimore, Md.: The Johns Hopkins University.

Chi, M. T. H., Feltovich, P. J., and Glaser, R.
1981 Categorization and representation of physics problems by experts and novices. *Cognitive Science* 5(2):121–152.

Cobb, P., Wood, T., and Yackel E.
In Press Classrooms as learning environments for teachers and researchers. *Journal for Research in Mathematics Education Monograph.*

Cohen, David K.
1988 Educational Technology and School Organization. In Nickerson, R.S. and Zodniates, P.P. (Eds) *Technology in Education: Looking Toward 2020.* Hillsdale, NJ: Lawrence Erlbaum Associations.

Committee on Adolescence, Group for the Advancement of Psychiatry
1968 *Normal Adolescence.* New York: Charles Scribner's Sons.

Committee of Inquiry into the Teaching of Mathematics in Schools (Dr. W. H. Cockcroft, Chairman)
1982 *Mathematics Counts.* London, Great Britain: Her Majesty's Stationery Office.

Department of Education and Science and the Welsh Office
1987 *National Curriculum: Task Group on Assessment and Testing: A Report.* Great Britain: Department of Education and Science and the Welsh Office.

de Lange, J., Jzn.
1987 Mathematics Insight and Meaning. Utrecht, The Netherlands: OW & OC.

Epstein, Joyce L., and MacIver, Douglas J.
1989 *Education in the Middle Grades: Overview of a National Survey of Practices and Trends.* A paper presented at the American Education Research Association, March, 1989 under a grant from OERI No. G-86-90006. Baltimore, Md.: The Johns Hopkins University Center for Research on Elementary and Middle Schools.

ERIC Clearinghouse for Science, Mathematics, and Environmental Education.
1982 *Teaching Science to Middle School Student.* Information Bulletin, No. 2.

Feuerstein, Reuben, Rand, Yaacov, Jensen, Morgens Reimer, Kaniel, Schlomo, and Tzuriel, David
1987 Prerequisites for Assessment of Learning Potential: The LPAD Model. In Carol Schneider Lidz (ed.), *Dynamic Assessment: An Interactional Approach to Evaluating Learning Potential.* New York and London: The Guilford Press.

Flavell, John H.
1963 *The Developmental Psychology of Jean Piaget.* Princeton, N.J.: D. Van Nostrand. Reprinted 1968.

Frederiksen, N.
1984 The real test bias: Influences of testing on teaching and learning. *American Psychologist* 39:193–202.

1986 Construct validity and construct similarity: Methods for use in test development and test validation. *Multivariate Behavioral Research* 2:1–24.

Frederiksen, N., and Ward, W. C.
1978 Measures for the study of creativity in scientific problem solving. *Applied Psychological Measurement* 2:1–24.

Gardner, P. L.
1975 Attitude measurement, a critique of some recent research. *Education Research* 7:101–109.

Goodlad, John
1984 *A Place Called School.* New York: McGraw-Hill.

Gronlund, N.E.
1968 *Constructing Achievement Tests.* Englewood Cliffs, N.J.: Prentice Hall.

Gruhn, W. T. and Douglas, H. R.
1971 *The Modern Junior High School* (3rd ed.). New York: The Ronald Press.

Guertin A., Pease, R. and Smith, W.
1987 A guide to computer use by the science teacher. A Curriculum Development Project of the Project to Increase Mastery of Mathematics and Science. (PIMMS) and the Connecticut State Department of Education, Hartford, Conn.

Hall, G. Stanley
1905 *Adolescence: Its Psychology and its Relations to Physiology, Anthropology, Sociology, Sex, Crime, Religion, and Education.* 2 volumes. New York: Appleton-Century.

Hueftle, S. J., Rakow, S.J., and Welch, W.
1983 *Images of Science: A Summary of Results from the 1981*82 National Assessment in Science.* Minneapolis, Minn.: University of Minnesota, Minnesota Research and Evaluation Center.

Hurd, Paul de Hart
1987 *The Middle School as an Institution: Implications for an Education in the Sciences.* New York: Carnegie Corporation.

Hurd, Paul de Hart, Robinson, J. T., McConnell, M. C., and Ross, N. M., Jr.
1981 *The Status of Middle School and Junior High School Science* (Volume II Technical Report). Louisville, Colo.: Center for Educational Research and Evaluation.

Johnson, M. (ed.)
1980 *Toward Adolescence: The Middle School Years.* The National Society for the Study of Education 1980 Yearbook, Part I. Chicago, Ill.: University of Chicago Press.

Johnston, K. L., and Aldridge, B. G.
1984 The crisis in science education: What is it? How can we respond? *Journal of College Science Teaching* Sep/Oct:20–28.

Levin, Henry, and Rumberger, Russell
1983 *The Educational Implications of High Technology.* Institute for Research on Education, Finance and Governance. Project Report No. 83-A4. Stanford, Calif.: Stanford University.

Linn, M. C., and Levine, D. I.
1976 Adolescent reasoning: The development of the ability to control variables. In *Advancing Education through Science Oriented Programs.*: Report PSc-9. Berkeley, Calif.: Lawrence Hall of Science.

Linn, M.C.
1988 Science Education and the Challenge of Technology. In Association for the Education of Teachers in Science *1988 Yearbook Information Technologies and Science Education* Eric Clearinghouse for Science, Math and Environmental Education.

Linn , M. C., and Songer, N. B.
1988 *Educational Reformulation: Incorporating Technology into the Curriculum.* Available from authors, University of California, Berkeley.

Lipsitz, Joan Scheff
1977 *Growing Up Forgotten.* Lexington, Mass.: D.C. Heath.

1978 Perspectives and Recommendations for the Science Education Directorate. In *Early Adolescence: Perspectives and Recommendations.* National Science Foundation Directorate for Science Education—Office of Program Integration, September 1978, SE78-75:69.

Lipsitz, Joan Scheff (ed.)
1979 *Barriers: A New Look at the Needs of Young Adolescents.* New York: The Ford Foundation.

Loucks-Horsley, Susan, Carlson, Maura O., Brink, Linda H., Horwitz, Paul, Marsh, David D., Pratt, Harold, Roy, Kenneth Russell, and Worth, Karen
1989 *Developing and Supporting Teachers for Elementary School Science Education.* A report of the National Center for Improving Science Education. Andover, Mass.: The NETWORK, Inc.

Mathematical Sciences Education Board
1989 *Everybody Counts: A Report to the Nation on the Future of Mathematics Education.* Washington, D.C.: National Research Council.

McEwin, C. K., and Clay, R. M.
1983 *A National Comparative Study of Practices and Programs of Middle and Junior High Schools.* Boone, NC.: Appalachian State University.

Mokros, J.E.
1986 *The Impact of Micro Computer-Based Labs on Children's Graphing Skills.* Cambridge, MA: Technical Education Research Centers.

Mullis, Ina V.S., and Jenkins, Lynn B.
1988 *The Science Report Card. Elements of Risk and Recovery.* National Assessment for Educational Progress. Report No. 17-S-01. Princeton, N.J.: Educational Testing Service.

Munby, Hugh
1983 Thirty studies involving the Scientific Attitude Inventory: What confidence can we have in this instrument? *Journal of Research in Science Teaching* 20(2):141–162.

Murnane, Richard J., and Raizen, Senta A., eds.
1988 *Improving Indicators of the Quality of Science and Mathematics Education in Grades K*12.* Committee on Indicators of Precollege Science and Mathematics Education, National Research Council. Washington, D.C.: National Academy Press.

National Assessment of Educational Progress (NAEP)
1987 *Learning by Doing.* Report No.: 17-HOS-80. Princeton, N.J.: Educational Testing Service.

National Association of Secondary School Principals
1985 An agenda for Excellence at the Middle School Level.

National Commission on Excellence in Education
1983 *A Nation at Risk: The Imperative for Educational Reform.* Supt. of Doc. No. 065-000-00177-2. Available from the U.S. Government Printing Office. Washington, D.C.: U.S. Department of Education.

National Middle School Association
1982 *This We Believe.* Columbus, Ohio: National Middle School Association.

National Science Board Commission on Precollege Education in Mathematics, Science and Technology.
1983 *Educating Americans for the 21st Century.* CPCE-NSF-03. Washington, D.C.: National Science Foundation.

1987 *Science and Engineering Indicators—1987.* NSB 87-1. Washington, D.C.: National Science Foundation.

National Science Resources Center
1988 *Science for Children.* Washington, D.C.: National Academy Press.

Nickerson, Raymond S.
1988 On improving thinking through instruction. *Review of Research in Education* 15:3–57.

Oakes, Jeannie

1986 *Educational Indicators: A Guide for Policymakers.* OPE-01. Santa Monica, Calif.: The RAND Corporation.

1989 What Educational Indicators? The Case for Assessing the School Context. *Educational Evaluation and Policy Analysis* 11(2)181–199. Washington, D.C.: American Educational Research Association.

1990 *Multiplying Inequalities: The Effects of Race, Social Class, and Ability Grouping on Students' Opportunities to Learn Mathematics and Science.* Santa Monica, Calif.: The RAND Corporation.

Offer, Daniel, Ostrov, Eric, and Howard, Kenneth.

1981 *The Adolescent. A Psychological Self-Portrait.* New York: Basic Books.

Padilla, M. J.

1986 *Middle School Science Curriculum: Some Critical Issues* (unpublished). A paper commissioned for the 1985 Middle School Science Project.

Raizen, Senta A., Baron, Joan B., Champagne, Audrey, Haertel, Edward, Mullis, Ina V. S., Oakes, Jeannie

1989 *Assessment in Elementary School Science.* A report of the National Center for Improving Science Education. Andover, Mass.: The NETWORK, Inc.

Rakow, S. J.

1988 Images of Science Through the Eyes of Middle School Students. *Science in the Middle School,* XI(1):1–8.

Renner, John W. et al.

1976 *Research, Teaching and Learning with the Piaget Model.* Norman, OK: University of Oklahoma Press.

Rothman, R., and Cohen, D. L.

1989 Middle Grades called "Powerful" Shaper of Adolescents. *Education Week* Vol VIII, Number 39 (June 21).

Rowe, Mary Budd

1978 *Teaching Science as Continuous Inquiry: A Basic.* New York: McGraw-Hill.

1979 Externality and Children's Problem Solving Strategies. Paper prepared under NIMH grant no. R01MH 25229. Available from the University of Gainesville, Florida, Gainesville.

Shamos, Morris

1988 The lesson every child need not learn. *The Sciences* July/Aug 1988:14–20.

Shayer, Michael, and Adey, Philip
1981 *Towards a Science of Science Teaching—Cognitive Development and Curriculum Demand.* London, Great Britain: Heinemann Educational Books.

Sizer, Theodore
1984 *Horace's Compromise: The Dilemma of the American High School.* Boston: Houghton Mifflin.

1986 Changing Schools and Testing: An Uneasy Proposal. In *Redesign of Testing for the 21st Century.* 1985 ETS Invitational Conference Proceedings. Princeton, N.J.: Educational Testing Service.

Snow, Richard E.
1980 Aptitude and Achievement. In W.B. Schrader (ed.), *Measuring Achievement: Progress over a Decade.* New Directions for Testing and Measurement. San Francisco: Jossey-Bass.

Superintendent's Middle Grade Task Force
1987 *Caught in the Middle: Educational Reform for Young Adolescents in California Public Schools.* Sacramento, Calif.: State Department of Education.

Task Force on Education for Economic Growth
1983 *Action for Excellence: A Comprehensive Plan to Improve Our Nation's Schools.* Denver, Colo.: Education Commission of the States.

Twentieth Century Fund
1983 Report of the Twentieth Century Fund Task Force on Federal Elementary and Secondary Education Policy. New York: The Twentieth Century Fund.

Weiss, I. R.
1978 *Report of the 1977 National Survey of Science, Mathematics, and Social Studies Education.* Research Triangle Park, N.C.: Research Triangle Institute.

1986 *Developing Options for Managing NSF's Middle School Science Education Programs.* Research Triangle Park, N.C.: Research Triangle Institute.

1987 *Report of the 1985*86 National Survey of Science and Mathematics Education.* Research Triangle Park, NC: Research Triangle Institute.

Wiggins, Grant
1989 A True Test: Toward More Authentic and Equitable Assessment. *Phi Beta Kappa* 70(9):35–39.

Willson, V. L.
1983 A meta-analysis of the relationship between science achievement and science attitudes: Kindergarten through college. *Journal of Research in Science Teaching* 20(9):839–850.

Yager, Robert E.
1988 Criteria for Identifying and/or Developing Exemplary Science Programs in Middle Schools. Science in the Middle School XI (1):9–14.

Zimmerer, L.K., and Bennett, S.M.
1987 *Gender Differences on the California Statewide Assessment of Attitudes and Achievement in Science.* Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C.

Assessment Panel

Joan Boykoff Baron
Consultant
Bureau of Evaluation and Student Assessment
Connecticut State Department of Education
Hartford, Connecticut

Edward Haertel
Associate Professor
School of Education
Stanford University
Stanford, California

Jeannie Oakes
Associate Professor
University of California, Los Angeles, and
The RAND Corporation
Santa Monica, California

Audrey B. Champagne*
Senior Program Director
Office of Science and Technology Education
American Association for the
Advancement of Science Washington, D. C.

Ina V. S. Mullis
Deputy Director
National Assessment of Educational Progress
Educational Testing Service
Princeton, New Jersey

Senta A. Raizen, Chair
Director, the National Center for Improving Science Education
Washington, D. C.

---

* Currently co-director, National Center for Improving Science Education

120

Advisory Board Members

William Baker
Chairman (Retired)
AT&T Bell Laboratories
Murray Hill, New Jersey

David P. Crandall
Executive Director
The NETWORK, Inc.
Andover, Massachusetts

Richard Duncan
President
Association of Presidential Awardees in
Science Teaching
Whitford Intermediate School
Beaverton, Oregon

Douglas Lapp
Executive Director
National Science Resources Center
National Academy of Science
Washington, D. C.

Andrew Porter
Director of the Wisconsin
Center for Educational Research
University of Wisconsin
Madison, Wisconsin

Kenneth Russell Roy
National Director
National Science Supervisors Association
Glastonbury Public Schools
Glastonbury, Connecticut

Thomas P. Sachse
Manager
Mathematics, Science and Environmental
Education
California Department of Education
Sacramento, California

Laurence Vickery
General Director for
Regional Personnel Administration
General Motors Corporation
Warren, Michigan

Richard M. Berry
National Science Foundation (Retired)
Bowie, Maryland

Sally Crissman
Science Teacher and Lower School HeadT
h)TShady Hill School
Cambridge, Massachusetts

David Kennedy
President
Council of State Science Supervisors
Olympia, Washington

Jerome Pine
Professor of Biophysics
California Institute of Technology
Pasadena, California

Mary Budd Rowe*
Professor
College of Education
University of Florida
Gainesville, Florida

F. James Rutherford
Chief Education Officer
American Association for the
Advancement of Science
Washington, D. C.

Mare Taagepera
Director
Science Education Program
University of California
Irvine, California

* Currently Professor, School of Education,
Stanford University

Index

130

The National Center for Improving Science Education, funded by the U.S. Department of Education's Office of Educational Research and Improvement, is a partnership of The NETWORK, Inc. and the Biological Sciences Curriculum Study (BSCS). Its mission is to promote changes in state and local policies and practices in science curriculum, science teaching, and the assessment of student learning in science. To do so, the Center synthesizes and translates recent and forthcoming studies and reports and develops practical resources for policy makers and practitioners. Bridging the gap between research, practice, and policy, the Center's work promotes cooperation and collaboration among organizations, institutions, and individuals committed to the improvement of science education.

The NETWORK, Inc.
Andover, Massachusetts
Washington, D.C.

BSCS
The Biological Sciences Curriculum Study
Colorado Springs, Colorado