Journal of Science Education and Technology, Vol. 1, No. 3, 1992

Instructional Assessments: Lever for Systemic Change in Science Education Classrooms

Brian Gong,[1,3] Richard Venezky,[2] and David Mioduser[2]

The Educational Testing Service/University of Delaware Science Education project has addressed the challenge of promoting systemic change in science education through the use of instructional assessments. Instructional assessments integrate performance-based assessments with curriculum and instruction at the classroom level. Instructional assessment provides a promising lever for promoting professional development and improving middle school science curriculum, instruction, and assessment.

Key words: Assessment; instructional assessment; science; science curriculum.

By the year 2000, U. S. students will be first in the world in science...achievement.
—America 2000, p. 63

SYSTEMIC CHANGE: "I KNOW AN OLD WOMAN WHO SWALLOWED A FLY . . ."

A children's song begins,

I know an old woman who swallowed a fly.
I don't know why she swallowed a fly.
Perhaps she'll die.
I know an old woman who swallowed a spider
that wriggled and wiggled and tickled inside
her.
She swallowed the spider to catch the fly,
but I don't know why she swallowed the fly.

The subsequent verses expand on this food chain as the woman, to the singer's delight, ingests larger and larger and more and more preposterous animals to help her solve her problem.

The United States has great needs and lofty goals for citizens of all ages to be more literate in science and technology. A central challenge to achieving these goals is to improve the science education that takes place in the schools, but many historians of science education and current advocates of reform agree that improving American science education embodies a systemic challenge approaching proportions like those faced by the troubled but energetic woman in the song. Working on one part inevitably raises a need to work on another part in an interconnected educational system. Science education well illustrates this point: to meet the governors' and president's resolution that American children will be number one in the world in science, the science curriculum must be refocused to connect education with the real world. However, to implement such curricula, teachers must be trained both in increased subject matter understanding and new pedagogical approaches. Instructional practices must be radically transformed from the "teach by mentioning" and "read about science; don't do science" approaches predominant today. Teacher preservice institutions must rethink the relation between "education" and the science discipline areas. Teacher inservice must reach millions of teachers with more than a half-day workshop in science per year. Teachers must work with parents to foster sustained interest in science among students in greater numbers

[1] Educational Testing Service, Division of Cognitive and Instructional Science, Rosedale Road, Princeton, New Jersey 08541.
[2] Department of Educational Studies, University of Delaware, Newark, Delaware.
[3] Correspondence should be directed to Brian Gong, Educational Testing Service (ETS), Rosedale Road, Princeton, New Jersey 08541.

and greater diversity, and, to support competent science teachers and students to practice in the schools, assessments, expectations, funding, and organizational structures must be revised.

As studies concur that piecemeal solutions have constrained impact and little staying power, the need for systemic reform is rapidly becoming accepted: curricular revision must include attention to teacher training or be doomed from the beginning; organizational restructuring of schools must be accompanied by provision for adequate resources; changes in assessment must support desired changes in practice (NSF, 1990). A key task is to identify where to start in the system: Where can we focus our efforts that will lead to systemic improvements, where efforts in one area will support and magnify contributions in other areas?

ASSESSMENT AND SCIENCE EDUCATION

Assessment may provide a strategic leverage point for systemic reform, many people in education, research, politics, and business agree. In the press to improve the American educational system, assessment has been cast in the role of "driving instruction and curriculum" increasingly frequently (cf. Department of Education, 1991; Resnick, 1991; Shepard, 1991; Frederiksen and Collins, 1989). It has been true for science education reform as well. In the past decade, dozens of reports by various commissions and boards have called for reform of science education (e.g., AAAS, 1989). Many of these reports have noted that improved assessment will play a crucial role in supporting the needed changes. However, in most of these discussions, assessment appears as a precursor or adjunct to classroom learning. Rather than being assigned an active, ongoing role, assessment is relegated to providing a statement of goals, motivation, or general information for accountability rewards and sanctions.

Although there is wide agreement that assessment is a necessary component of systemic reform in science education, there is less agreement about what form assessment should have and what role it should play. Much debate and action have revolved around reforming assessment instruments. The current debate over multiple-choice response formats and "authentic assessments" is perhaps the most evident and intense part of this focus on instruments (e.g., Archbald and Newmann, 1988). However, the question of the role for assessment, especially in relation to other parts of the educational system, precedes the technical questions of format (e.g., multiple choice versus non-multiple choice), medium (e.g., pencil-and-paper versus performance), and administration mode (e.g., standardized versus work sample). The more essential issues have to do with what purposes the different types of assessment are being used to achieve, how the assessment is related to the other components of the educational system, and how the information is used to improve learning and teaching.

Perhaps because of the focus on national and state-level reform, there has been less discussion of how assessment might be incorporated by teachers and students in the classroom to enhance learning. Discussion of the role of assessment in supporting educational change must be extended to envision teachers and learners more actively assessing and improving their own goals and performances. A challenge for educational reformers is to envision classrooms that are self-improving systems and to conceive of classroom-based assessments that are powerful in the hands of learners and teachers, coherent in relation to the best of curriculum and instruction, and practical in the contexts of classrooms today.

Promoting Systemic Change in the Classroom through Instructional Assessment

For the past several years we have focused on instructional assessment as a leverage point for promoting systemic change at the classroom level. Instructional assessment can powerfully improve science education by providing mechanisms that make a system more adaptive, self-monitoring, and self-improving: Instructional assessments help provide useful information to teachers and students, engage teachers and learners in active assessment to improve classroom teaching and learning, motivate purposeful learning, and adapt instruction to the progress of learners. Instructional assessments can merge performance assessments into the on-going instructional decisions of the classroom.

Our underlying model of assessment holds that assessment should be informative and empowering: it should inform intelligent action, and the information should support those educational agents who need to act and make decisions, notably teachers and students. At the same time, instructional assessments that are developed within a context of systemic change at the classroom level have implications for other important aspects of the educational system—particularly curriculum, instruction, and teacher development.

Instructional assessment is too large a topic to discuss comprehensively in a single article. In this paper we focus on establishing a perspective for instructional assessment and describe some practical developments that put instructional assessments into classrooms. In doing so, we draw heavily on our experience from the project we have conducted for the past three years. Funded by ETS, the ETS/University of Delaware Science Education Project has worked with collaborating teachers and schools in grades 4–8 to develop, implement, and research the effects of instructional assessment materials and practices on science education.

In the first part of the paper we argue that instructional assessment is an appropriate approach for assessment within educational reform. In the second part of the paper we describe the design of instructional assessments in general and give some examples of instructional assessment activities being developed for middle school classes in science. The impact those activities have had on promoting teacher development, curriculum, instruction, and assessment is described in the third part. We conclude by offering a tentative analysis of the characteristics of instructional assessments that contributed to the teacher change we have observed, and discuss directions for further research and development.

Characteristics of Assessment that Helps Learners and Teachers Improve

Various types of assessment can play several roles in promoting educational change. We focus on the role of informed empowerment for learners and teachers through instructional assessment. Instructional assessment is characterized as serving classroom teachers and students by being classroom-based, educational, and self-administered.

We focus on classroom-based assessment because we believe that assessment information is invaluable in informing adaptive learning and teaching. Instructional assessments arise from the educational interactions between students and teachers, to provide information about student status and progress that teachers or students can act on to improve student learning. Assessment—and in particular assessment that is directed by learners and teachers towards improving learning and teaching—is essential for intelligent action in a complex environment. Classrooms are characterized by changing, complex conditions that require learners and teachers to assess in order to know what to do. Assessment is an integral and essential part of learning and action to improve and should be incorporated by design into instruction and curriculum. Good design (e.g., curriculum) and implementation (e.g., instruction) are essential, but are insufficient when conditions change or are underspecified. In the classroom, instructional assessment provides the feedback loops required to make the system self-monitoring and self-improving.

We think it most appropriate that the assessment within an educational system should itself be deeply educational. That is, the assessment should pay attention to fostering changed perceptions, knowledge, skills, attitudes, and actions. The mechanisms for systemic improvement must be woven into the fabric of the system itself for change to be pervasive and on-going. In other words, educational systems, including the student, teacher, classroom, and school, should be self-monitoring and self-improving. Assessment and responsibility for improvement should be advocated for all involved; there should be no gulf drawn between those who learn and those who evaluate, those who teach and those responsible for program improvement. Students and teachers should be users as well as providers of assessment information. Assessment criteria and results should be valuable and available to both students and teachers. Students and teachers should be expected to seek and use assessment information to help them achieve their purposes and goals. Instructional assessments help classroom teachers and students become the assessment agents. They become primarily responsible for achieving consensus about the assessment standards and criteria, for gathering and analyzing the assessment information, and applying the information to make a difference. Instructional assessments are consistent with our deeply held conviction that students and teachers are responsible for their own learning, at least as much to themselves as to an external "accountability" agent. In contrast, standardized tests usually are administered by an entity external to the classroom, the scores and reports are usually generated outside of the classroom, and the information usually is weakly related to either what has been learned or what should be done to help individual students.

Assessment intended to improve classroom practice must provide information that can guide meaningful action by learners and teachers. In our experience, teachers and students need explicit support beyond clear goals and accountability mechanisms. To contribute powerfully to improved classroom learning and teaching by teachers and students, feedback must be provided that deals with instructionally relevant units, in instructionally sensitive time frames, within instructionally meaningful conceptual frameworks. While traditional accountability assessment can communicate goals and values, evaluate quality or worth, and dispense rewards and sanctions, such actions are not essentially educative. Similarly, classroom-based assessments that focus on selection, placement, and grading deal with precursors and adjuncts to learning. Instructional assessments are intended to find life within the fluid atmosphere of a classroom, where the main goal is to make decisions and take action that will enhance the learning of students. Accountability assessments and program evaluations that focus beyond particular classrooms or use nonclassroom criteria (e.g., national norms) do not provide information that addresses the specific needs of students and teachers. This is particularly true of pre-high school science, where most teachers have little training in the subject matter area. Clear goals and strong sanctions will not necessarily lead to improved science education; to solve the challenges of science education reform, teachers and students also need to know what to do and how to do it.

Instructional assessment is integrated with learning itself. The main purpose of instructional assessments is to promote learning by providing experiences and information that support classroom learning and teaching. While instructional assessments do provide information that may be useful for classroom evaluation (e.g., grades), we have sharply distinguished between instructional assessments and assessments for the purpose of accountability beyond the classroom, e.g., evaluations of students, teachers, classes, or programs by "external" audiences for noninstructional purposes. The main reason for maintaining this separation is the fact that the motivations and operational conditions that make instructional assessments so valuable as an integral part of classroom life become distorted or are insufficiently rigorous for high-stakes evaluations.

We turn now to describe the design of instructional assessments that achieve these purposes of informing and improving classroom learning and teaching.

DESIGN OF INSTRUCTIONAL ASSESSMENT

We have developed a working model for the design of instructional assessments, several units of prototype materials centered in science, and staff development programs for implementing these materials. The units, tools, and teacher development program work together in several ways to help foster improved science education that can contribute to systemic change. The units serve as concrete examples of the goals and procedures of the program. The design of instructional assessment activities serves as a focal point for incorporating significant curriculum revision as well as assessment. The instructional assessment activities provide a platform for teachers to examine their assumptions about student learning and motivation. As they have been implemented in schools, they also provide a platform for a research program.

Units

Our decision to develop example units reflects several considerations. We assumed that instructional assessment is context-sensitive, that is, this type of assessment makes sense only in relation to a curriculum and instruction that can be guided by the assessment outcomes. We concluded that we did not want to develop instructional assessments around current textbook-based curricula, which offer highly constrained opportunities for instructional assessment. Textbook-based approaches to middle school science instruction are typically didactic and inflexible, depending primarily upon a linear sequence of "read and test," with little emphasis on inquiry or student activities and with breadth preferred over depth. In contrast, instructional assessment requires a coherent curriculum based upon student engagement in science, but with clearly delineated conceptual goals and a variety of instructional options. The reformed science curricula guidelines that emphasize depth and organization of important knowledge (e.g., AAAS *Project 2061: Science for All Americans* and the 1990 California Department of Education *State Science Curriculum*

Framework) reflect better the objectives we established for science teaching and which the cooperating teachers also favored. However, few materials for middle school are available that reflect the new guidelines.

We have developed sample instructional units that deal with central science themes, explanatory principles, and causal models within the context of water resources, tools and technology, meteorology, and control systems. The units are interdisciplinary across science disciplines and incorporate non-science areas such as art, social studies, language arts, and mathematics. Each unit is problem-centered and includes a socially relevant or contemporary context for learning the knowledge and applying the skills.

The units incorporate iterative cycles. This is consistent with the assessment design and also with a constructivist instructional model. Rather than having the students learn a string of pieces and then try to put them all together at the end, we have designed the units so that they have repeated practice throughout the unit seeing the large patterns and putting the pieces together. This is especially true of cycles of assessment and inquiry.

Description of Two Example Units

The *Tools and Technology* unit focuses on machines, tools, and technologies viewed as the result of human problem-solving and design activities aimed at satisfying specific needs. The knowledge and skills to be learned through the unit are grouped around two main processes: (1) the *functional analysis process* of technological solutions, and (2) the *design process* of technological solutions.

Through the functional analysis activities, the students deal with functional aspects of a given tool or machine (e.g., its main purpose or network of necessary subfunctions); its structural components, materials, and mechanisms; and the contextual reasons (e.g., social, cultural, technological, or economic reasons) that affect how the particular object came to be designed. In the unit activities the students analyze and design real tools and artifacts, including "antiques."

Through the design activities, the students acquire the knowledge and skills required to carry out a methodical and systematic process of designing and implementing a technological solution. The students are involved in stating a problem and defining the specifications for its required solution, doing research and experimentation on alternative solutions for the problem, implementing the solution, and evaluating the extent to which it satisfied the stated problem requirements and specifications. Note that this analysis–design process so pertinent to technological problem-solving differs somewhat in context and criteria from the conceptual hypothesis-testing approach of the "scientific method" often taught in elementary and middle school science classes.

The *Water Resources Unit* deals with the issues of quality and distribution of water and the role of technology in managing our water supply. The unit is designed to require an evolving problem-solving process: to supply the people in a new settlement with sufficient water of appropriate qualities for their different needs and uses. The learning is carried out primarily in small groups. Each group investigates a particular geographic area, with its different constraints and problems related to water quality, uses, and supply (e.g., Buffalo, New York; Fresno, California; Mexico City, Mexico). The large problem is divided into six subproblems:

- *What Makes Water Water,* where the students learn to deal with the physical properties of water, water components, pollutants, and quality standards;
- *What to Do about the "Unwater" in Your Water,* which deals with purification processes of filtering, desalinization, etc;
- *Moving Water Around,* which deals with physical and technological aspects of water transportation and distribution;
- *Water at Work,* which deals with water's use in energy production, etc.;
- *Water's Delicate Equilibrium,* which deals with issues of water allocation and uses; and
- *Water's Beauty Contest,* which deals with aesthetic uses of water such as kinetic art and fountains.

Solving each of the problems requires the students to perform laboratory experiments, build conceptual and physical models, retrieve information from different sources, and communicate their progress in various ways. The character and specifications of each subproblem are defined on the basis of various social, historical, or technological contexts and constraints, such as the availability of given technologies in different historical periods or of economic considerations affecting choices among alternative solutions.

As the integrative assessment activity in the unit, the student groups prepare an Area Report documenting their analysis and recommended actions. They also construct a three-dimensional mock-up containing models and devices built through their work on the solutions of the different problems, such as stating the quality of the water in their area or using water for energy generation or aesthetic purposes. Students present oral reports to the class, and the whole class interacts in discussing each group's performance as represented in their presentation, report, and model.

Instructional Assessment Activities

Each unit contains a number of instructional assessment activities. The instructional assessment activities synthesize instruction and assessment into educational experiences that serve both functions. The assessment function comes because the experience is designed and used to gather information about the status and progress of students. The experiences are instructional because they are designed to provide valuable learning experiences, incorporated into the curriculum and instruction of the classroom. They are not added on after instruction solely to provide evaluative information to the teacher, as tests often are. Each instructional assessment activity must contribute to learning commensurate with the time and effort that it takes.

The challenging research and design issues surrounding development of good instructional assessments include devising purposes, goals, and frameworks that undergird the assessment process; creating activities and tasks that reflect specifications and criteria drawn from the frameworks and that will invite the desired performances and experiences for learners; devising means to record performances by learners, including processes and justifications; developing means to interpret the performances in ways that are insightful, reliable, and valid; conceptualizing ways to manage the assessment information and to report the assessment results that are practical and useful; developing means to tie the assessment results back to educational action and integrate them meaningfully into the instructional cycle; and developing models for implementing instructional assessment in the variety of institutional contexts that characterize middle school science education. We comment on each of these features briefly below, and give a short example in a subsequent section.

Purpose or Goal

The task must have a worthwhile educational goal that is motivating to the learner. It should provide the foundation for further learning and study, but should demonstrate power or worth at the time it is learned, if possible. Often the purpose may encompass a significant problem to be solved. The purpose sets the framework that establishes why the content to be learned is significant and worth the time.

Instructional assessment activities serve multiple purposes: they provide valuable learning experiences as well as opportunities to gather information about student status and progress useful for informing instructional decisions. Instructional assessments have the purpose of informing instructional decision-making and action on a level familiar to teachers and students. Traditional assessments are usually more concerned with communicating values and goals, informing institutional decisions of selection and placement, monitoring programmatic effects, and justifying rewards and sanctions.

Activities

The activities structure the experiences through which the learner and teacher achieve the goals or purposes. When the goal is to solve a certain problem, the activities are the means for enabling the learner to solve the problem. The activities also structure the experiences that provide the basis for instructional assessment. The project specifies developing activities of four sizes or scope: short, focused diagnostic activities that seek to pinpoint specific skills, knowledge, or attitudes; moderate-scope benchmark activities that deal with clusters of central concepts or skills; integrative activities that span a unit and require student generation, organization, and presentation or knowledge; and reflective activities that involve students and teachers in reviewing and analyzing their learning products and processes. These instructional assessment activities take place before, during, and after instruction in a unit.

To achieve their purpose, instructional assessments must balance multiple demands. They must provide useful assessment information and also themselves be worthwhile learning experiences. The instructional assessment activities in the units achieve this by being designed to:

- Structure classroom experience with strong attention to instructional process and content, in a coherent, flexible developmental sequence;
- Provide multifaceted information about student progress and status that is reliable, valid, and consistent with research on learning and teaching;
- Inform the learning process throughout the unit, from planning through evaluation and follow-up;
- Provide the basis for consensus of public standards;
- Establish a common role of self-evaluator for learner and teacher;
- Contribute to professional development of teachers.

Instructional assessment activities and their associated follow-up actions are integrated within a curricular and instructional context. Interpretations of assessments must be instructionally sensitive and instructionally relevant; similarly, instruction and the curriculum should be assessment-sensitive. That is, assessments should be tied to useful decisions and actions within the instructional context, and instruction and the curriculum should be open to change and adjustment based on assessment information. The instructional assessment tasks should focus teachers' and learners' attention on the critical dimensions and concepts of the discipline and encourage an instructional partnership of teacher and learners in the classroom.

One reason these instructionally relevant assessments are powerful is that their content is important, realistic, and engaging (Lesh *et al.*, 1992):

- Activities emphasize real-life problem solving situations.
- Objectives emphasize modeling-based views of science—and deeper and higher-order understanding and processes that are associated with development, assessment, and refinement of conceptual models.
- Problem-solving and decision-making issues are included that allow students to use realistic tools and resources.
- Responses involve multiple types and levels of correctness (Fig. 1).

Instructional assessments partake of the full range of good instructional activities and administra-

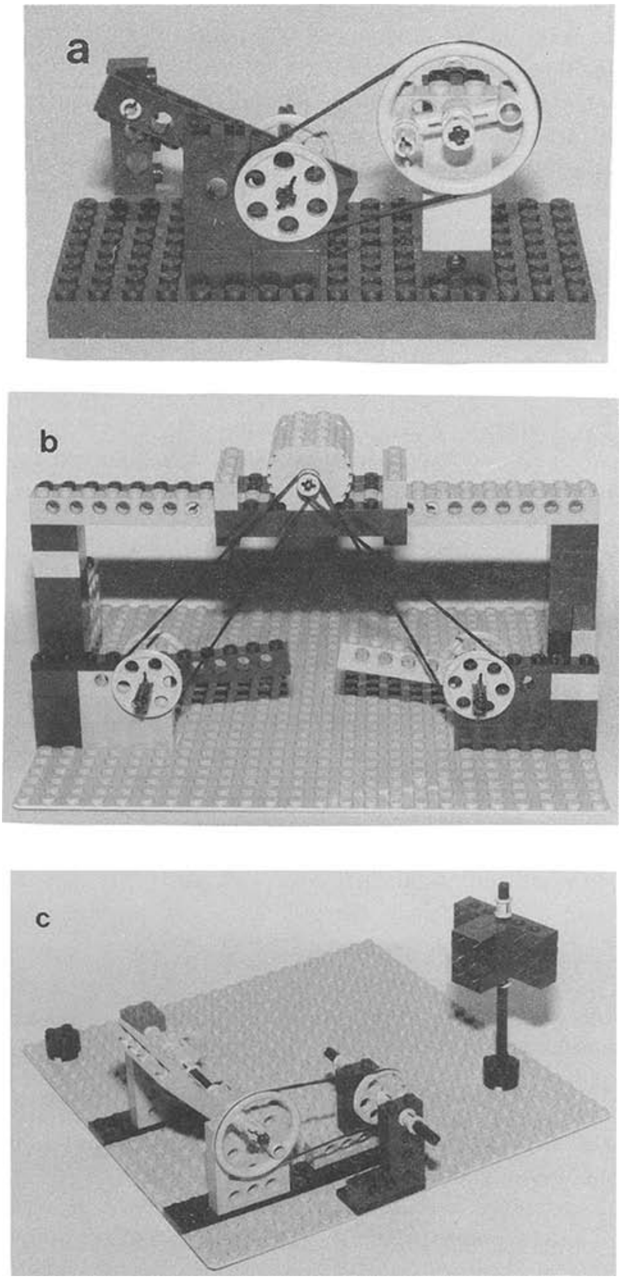


Fig. 1. Student projects using Lego-Logo: a, b, and c: students' "automatic door" projects constructed of Lego-Logo. In many constructive projects using technology, solutions can show multiple levels and multiple types of "correctness." Here three models are shown of an "automatic door," constructed by 6th grade students in the *Control Systems* unit using Lego-Logo.

tion patterns. Just as instructional means are rich and varied, so instructional assessments draw upon a range of formats, administrative organizations, lengths of time, and media of expression. The instructional assessment activities span many modalities, take many forms, and vary in scope from a few minutes to large projects that extend over months. Instructional assessment activities developed thus far include activities based on work sheets, written journals, oral presentations, written reports, dramatic presentations, creative writing, drawing, field trips, laboratory experiments, extended projects, class discussions, and quizzes devised by teachers and students. They include student-constructed and student-selected work and graphical modes of expression. They deemphasize the use of multiple-choice, norm-referenced tests. The activities take place in the instructional sequence from before instruction begins to after the last instructional activity. Some of the instructional assessments are informal and less structured. Some are highly structured in administration and analysis. All are essentially work samples and portfolios drawn from classroom learning activities.

Instructional assessments are intimately connected with content and curriculum. The assessments sometimes focus on very specific content, other times on more general skills, higher order thinking, or explanatory models. Explicit models of content knowledge and skills are incorporated in instructional assessment. Traditional assessments have tended to emphasize the multiple-choice question format, pencil-and-paper mode of delivery, and a standardized administration. Traditional assessments reports deemphasize specific content; the interpretation has often been of a "score" in reference to a "norm."

Instructional assessments provide motivation and action-oriented information on how to improve, in forms that are useful to the people with responsibility to improve. This information helps teachers and students at critical junctures in the teaching/learning process. Importantly, in many cases the assessments are intended for the students, as well as the teachers, to administer, analyze, and act upon. The assessment criteria and procedures are usually made explicit to the students; indeed, some activities involve students in formulating and applying assessment criteria. Their involvement extends, where possible in the units, to generation and evolution of the evaluation criteria (e.g., "What makes a good explanation? What would be good criteria for a good experimental design?"), as well as actually applying the criteria to their own activities and performances of peers. An element of self-assessment extends to the entire learning experience, as students reflect on what contributed to their own learning, what was significant, and how they could improve. Suggestions to teachers are often couched in this community of reflective members thinking about how to improve learning and teaching.

One type of instructional assessment we have focused on is diagnostic (pre)assessment. Administered a few days prior to the beginning of a unit, diagnostic preassessments are intended to provide the teacher with information about student knowledge, skills, and attitudes before starting the instructional module or activity. Considerable research in science education suggests that students may bring with them inconsistent conceptual knowledge of science topics that interfere with learning (e.g., McCloskey, 1983). On the other hand, students may also have rich and deep knowledge, particularly in science, beyond the curriculum that would be useful for teachers to know about. In addition to their diagnostic functions, preassessments also provide motivation and serve as advance organizers for the students.

Our preassessment activities are centered around a set of questions designed to tap these areas:

1. Student attitude about the topic (e.g., "I think meteorology is *boring, interesting, etc.*");
2. Student school experience with the topic (e.g., "Have you ever studied meteorology or the weather? When?");
3. Student knowledge of an explanatory model centrally important in the unit (e.g., "Explain what makes it rain. Include a diagram if you wish.");
4. Student knowledge of common knowledge associated with the topic (e.g., "Imagine you are a TV or radio weather announcer. Write a forecast for what the weather will be tomorrow.");
5. Student knowledge of technical terms associated with the topic (e.g., "Describe what each of these instruments does or is used for: barometer, weather vane, and thermometer.");
6. Student personal experience with some aspect of the topic (e.g., "Describe your most

unusual or scary experience involving weather.").

The questions may be the basis for an introductory class discussion of the unit or a written preassessment. A written preassessment has the advantage of allowing the teacher more time to analyze and digest the assessment information and to make appropriate decisions based on that information. In fact, written responses can be incorporated into classroom discussions.

Record of Performance

The experiences of the learners must be recorded to provide an evidential basis for the assessment. In many cases, production of the performance record serves other educational purposes, such as communicating worthwhile information to others, and engaging students in additional learning and practice.

In the project, almost all of the records of performances come from work samples, that is, from work produced by the students as part of their ongoing classroom or instructional activity. The instructional assessment activities include records of not only product, but also process (e.g., planning, decisions) and reflection (e.g., what went well, what could be improved, criteria for evaluating). Much of the work shows developmental and longitudinal progress, since it derives from the repeated developmental practice built into the units.

Interpretive Schemes

Two types of interpretive schemes, or learning progress maps, are associated with the instructional assessment activities. The instructional map provides an overview or instructional context of the particular activity in relation to other activities, goals, and experiences. Since the assessment takes place throughout the unit and not just at the beginning or end, it is important to have a sense of the place of each activity provided by the instructional map. A concept map gives an overall sense of the logical content organization at a high standard of expertise. Both of the maps also embody assessment criteria for the developing expertise addressed in the unit. Importantly, these learning progress maps provide the structure for articulated continuity between subject matter areas and grade levels. Instructional assessment incorporates elements of cognitive science-based diagnostic assessment, self-monitoring strategy learning, and portfolio and performance assessment.

One instructional map we have developed takes the form of a bulletin board used in a fourth grade class studying meteorology (Fig. 2). The board's major sections summarize the main points of what has been covered and show how they relate to the major topics to come. The instructional map clearly shows the students "why" the topic is important and interesting, and the "technology" of its study and applications are as integral to the unit as are the "science concepts." In addition, the bulletin board encourages the students to interact with each other, the teacher, and the board by providing a "question box" where students can pose questions,

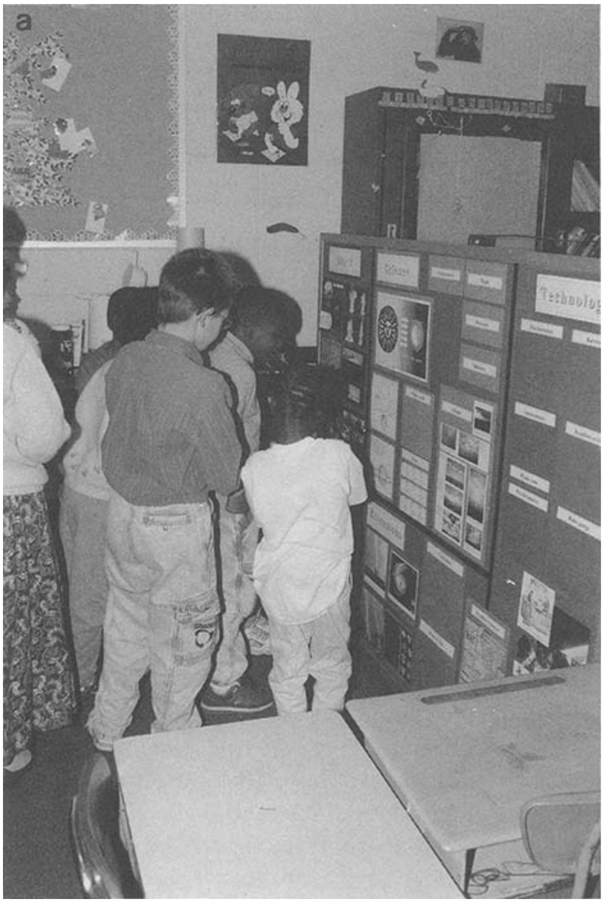


Fig. 2. Instructional map bulletin board: a: children looking at board; b: close-up of section of board. An instructional map on the bulletin board provides students and teachers an open framework for the *Weather* unit activities.

Fig. 2. Continued.

a display area for students' contributions, and hands-on extension activities. Unlike bulletin boards that are static, decorative wall displays, this instructional map is on eye-level for the fourth graders to touch and interact with.

The other type of interpretive scheme provides an analytical guide for interpreting knowledge, skill, and attitude components of the students' performances. These assessment guidelines help teachers and students analyze actual performances and relate them to appropriate instructional follow-up actions. The assessment guidelines provide the analytical and prescriptive bridges between the student performances and interpretation of the performances and guidance for subsequent instructional action. Assessment guidelines incorporate cognitive science analyses of cognitive development and knowledge representation, as well as instructional design principles of curricular sequencing. They include standards, criteria, rubrics, and procedures for relating them to the performance. Where appropriate, assessment guidelines include information for grading or other summative evaluations. The validity of these assessment guidelines and the reliability with which they are applied should be monitored by teachers and students, as well as by others interested in this critical interpretive aspect of instructional assessment.

Management and Reporting Aids

The information from instructional assessments is to be used by teachers and students, so it is important the information is reported clearly and usefully and that the information is managed in practical ways. The assessment guidelines always include provisions for reporting the results to students and teachers, with provisions for further instructional and/or curricular action. Our units include models for reporting to other audiences, such as administrators and parents.

Ties to Educational Follow-up Action

Instructional assessments provide a powerful feedback mechanism for improving learning and teaching. While the information provided through instructional assessments may inform changed perspectives and knowledge, it is usually important to have additional action taken. Each instructional assessment activity includes ties back to action that is educational or that is directly involved in additional learning and teaching. The information sometimes also is used to inform grades, program evaluation, etc.

Tools and Technology

Although the collaborating teachers in the science project have welcomed the new roles and responsibilities for assessment, they have needed practical aids to manage the increased information instructional assessment activities can provide and to structure instructional experiences. We have developed several computer-based tools that help extend the powers and motivation of learners and teachers to generate, understand, and use information from instructional assessments. The tools also help clarify conceptual issues and provide a platform for various research efforts.

The tools emphasize provision for the user to manipulate and control the tool to achieve a purpose, rather than using the computer as a teacher surrogate or an unfocused repository of information. These tools provide information useful for improving learning and teaching. They aid in structuring the tasks, applying the interpretive scheme, organizing information, or communicating and applying the information. These tools include: information management tools (*Video Analyst*), intelligent tutoring system (*The Weather Machine*), computer-assisted instructional courseware containing simulations and exploration environments (*Case of the Fallen Dough*), management tools (*Behavioral Banking*), instrumentation tools, and analytical tools for analyzing data bases of information (*Response Checker*). Each tool illustrates a different combination of instructional assessment components: the extent to which it assists the user to apply the interpretive scheme, supplies expert tutoring or instruction, helps define an assessment task, incorporates a model of an expert or the student, or organizes information for human analysis and action. Together they illustrate a range of technological aids, consistent with our emphasis on encouraging students and teachers to be responsible for and able to provide their own instructional assessment information. While the tools are simple, their practical contribution may be significant in terms of time saved, data gathered, etc.

Example of Instructional Assessment Analysis Using *Response Checker*

In this section we illustrate how a diagnostic preassessment task, interpretive analysis, and a computer-based analytical tool can provide quick, general group information, by analyzing a set of data generated by diagnostic preassessments. The data consisted of responses from five class periods of 7th graders ($n = 148$). These students were beginning study of a unit on plant biology and ecology.

Purpose

The purpose of the diagnostic preassessment question was to inform the teacher of student knowledge in a general way. In particular, the teacher's goal was to move the students towards being able to draw together factors from the environment, plant biology, and evolution to generate an explanation of some phenomenon in the natural world. The preassessment was to indicate how much students did such drawing together into multiple-cause explanations, in particular, the factors of environment, plant biology, and evolution.

Task

The question analyzed for the 7th graders was "Why do leaves change color and fall off the trees in the fall?" This question was chosen because it involves the three factors that the unit is concerned with: environment, plant biology and ecology, and evolution. In addition, the question is motivating for many students, since it leads them to wonder about a commonly observed phenomenon. The task also is used by the teacher to illustrate other aspects of the unit, including the on-going nature of scientific inquiry—the central mechanisms of plant hormonal regulation involved in leaf color change and falling were elucidated fewer than 50 years ago.

Interpretive Scheme

The interpretive scheme for this preassessment task is based upon comparison of the student explanations to an explanation generated to represent an "expert" level (actually, about a freshman college-student level). This explanation is represented schematically as a conceptual learning progress map in Figs. 3a and 3b. The teacher is provided an accompanying prose version of the conceptual learning progress map. (Due to space considerations, the prose version is not reproduced in this article.)

Two views of the conceptual learning progress map are shown in Figs. 3a and 3b. Just as teachers may need to understand more than they teach—and student understanding may be unevenly developed—so the map can be "zoomed in" and "out" to reveal successively more detailed explanations. Figure 3a shows a "top-level" view, while Fig. 3b shows a more detailed version of the same map. The explanations in these maps consist of factors and events (shown in boxes) and relationships (shown as arrows). In these conceptual learning progress maps, many "levels" of specificity may be invoked both for events and for relationships. The teacher and students may choose their own appropriate level of understanding and see their standard in the context of other possibilities.

While a detailed explanation of why leaves change color and fall is available to the teacher, for this analysis the teacher focused on the "big picture," concerning whether the students pulled together explanations involving environmental factors, biological factors, and evolutionary/adaptation factors. These sets of factors are grouped together in the learning progress map.

A number of key-word lists were generated that the teacher thought might reflect knowledge of each of these sets of factors. This was done both through conceptual analysis of the learning progress map and through exploratory analysis of the data base of responses. The first key-word list set deals with environmental factors: temperature, cold, light, day. The second set of key words deals with biological factors

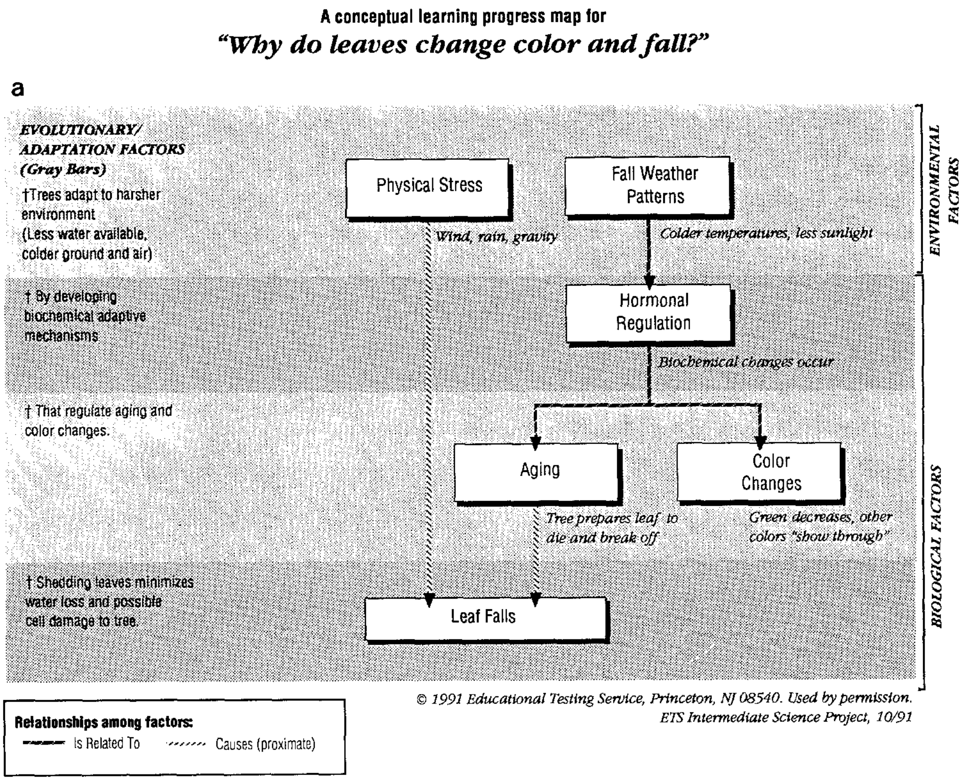


Fig. 3. Concept maps for "why leaves change and fall." a: A schematic conceptual learning progress map for explanation of "why leaves change color and fall" provides an interpretive guideline for analyzing students' knowledge. b: A conceptual learning progress map can represent various types of explanatory mechanisms and relationships in varying degrees of detail.

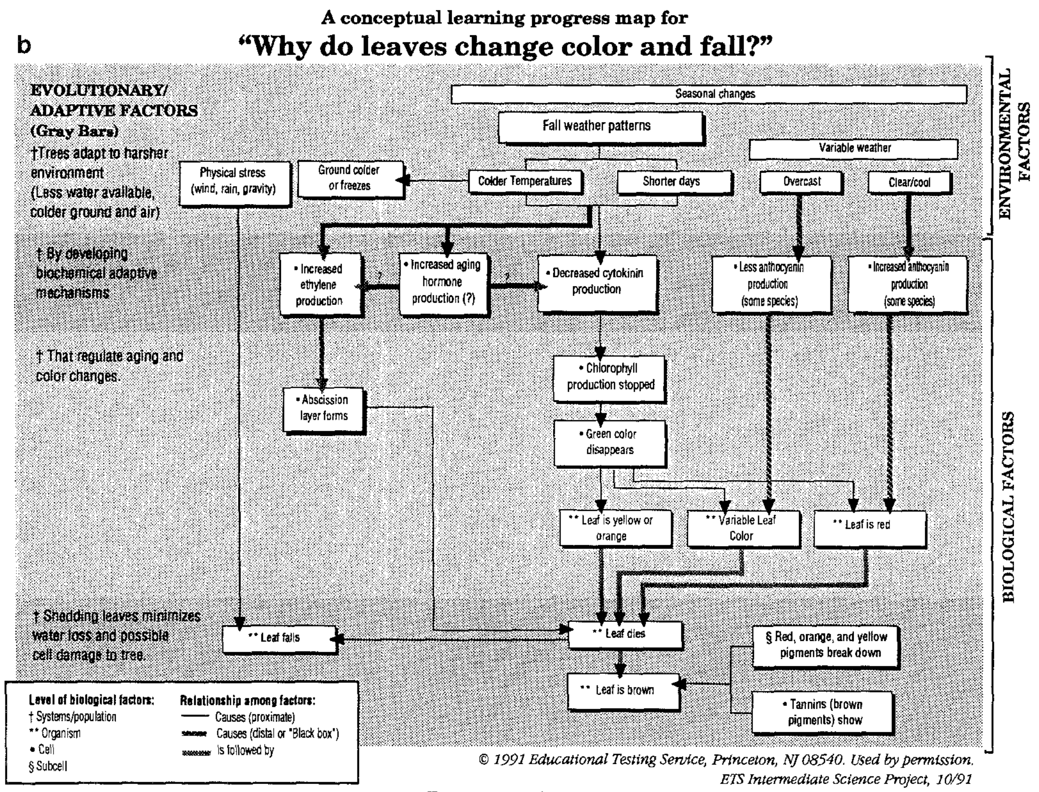


Fig. 3. Continued

or processes: die, old, chlorophyll. The third set deals with adaptation: adapt, evolve. This key-word list brought together some interesting results.

Results and Reports

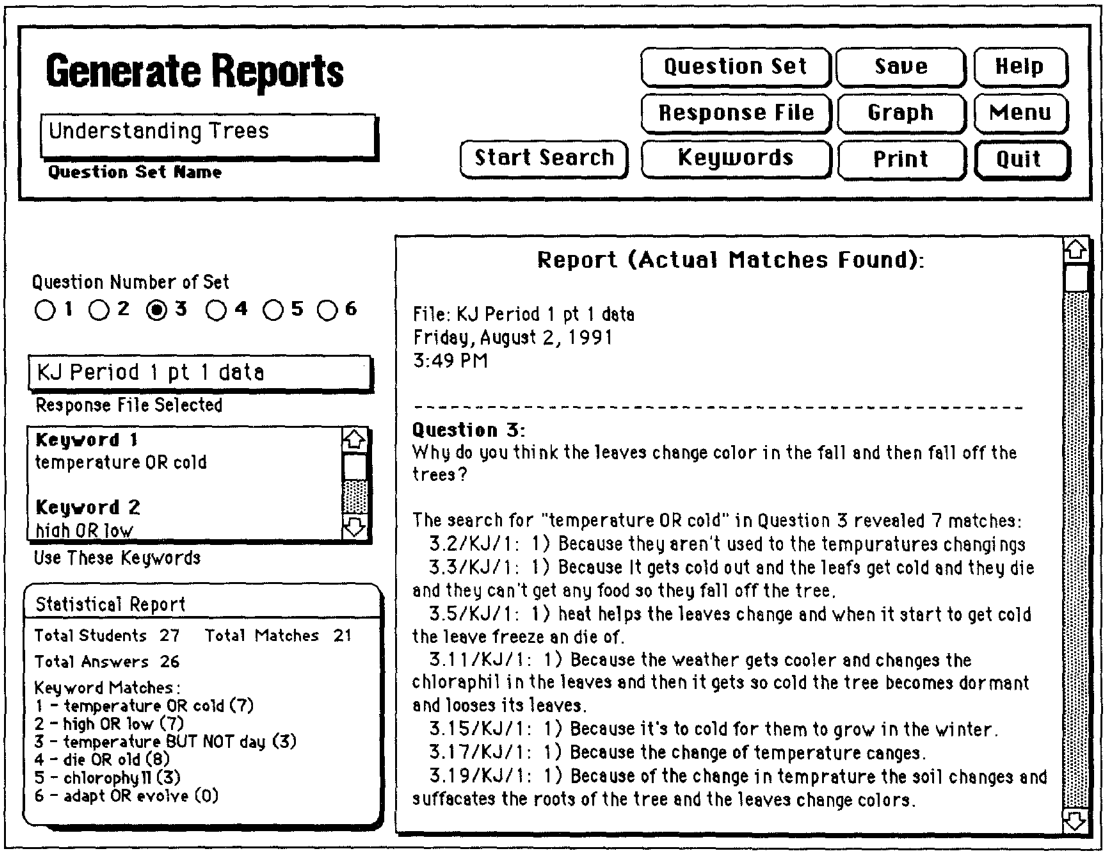
The results of using the tools were promising: the tools contributed to analyses that are feasible, interesting, and valuable to pursue. The tools were practical to use in the classroom and inviting for teachers to use.

A report that was printed out is shown in Fig. 4. Note that the software is able to deal with variant spellings of key words, such as "temperature."

The application of the analytical tool, the *Response Checker,* to the data base on "why leaves change color and fall" indicated that approximately 60% of the students drew upon notions of environmental factors to explain the changes, 30% of the students cited biological factors, while only 10% used a combination of factors. None of the students explained the leaf changes from an evolutionary/adaptation perspective. Since adaptation is a central concept in ecology and biology, the teacher has identified a major gap. Second, while several students attribute something to dying or growing old (31%), few mention a critical factor (chlorophyll, 12%). Even more striking, a secondary analysis shows that the students tended to mention either environmental or biological factors, but not combinations. Thus, the teacher knows that the class has not generated multiple-cause explanations and has not included explanations from an adaptation perspective. These tentative conclusions can be checked and refined by scanning the student responses and following up with classroom discussion. This information was available with only a few minutes work with the software. A parallel reading of the set of student responses by the classroom teacher took over four hours.

Instructional Follow-up

The teacher plans and carries out the appropriate instructional follow-up based on his or her analysis. Note that these instructional follow-up actions may take place over longer time periods. Direct remedial instruction is only one of the possible models for using the information provided by the preassessments. This might include a number of facets:



 ETS Intermediate Science Project, 10/91

Fig. 4. *Response Checker* screen for analysis of "why leaves change and fall." The *Response Checker* reports student responses that matched key words.

- curricular selection and instructional planning to fit the knowledge state of the children;
- drawing on strengths of individual children revealed by the preassessment;
- planning group or individual instructional experiences for weaknesses revealed by the preassessment;
- using the preassessment as a "baseline" for encouraging student self-assessment and marking progress;
- using the results of the preassessment to raise issues among the class, e.g., alternative naive conceptions, criteria of explanations.

TEACHER DEVELOPMENT

We have been especially interested in teachers' professional development as a key component in systemic change. One reason for this interest is that teachers are a major, essential part of the educational system, and sustained, systemic change will require teachers working together towards that change (AAAS, 1989; NSBC, 1989). For the foreseeable future teachers will mediate enhanced classroom practice and efforts to make the system more adaptive at the classroom level. The need to improve a district's science program often extends through all the system components. Importantly, teachers are one of the few agents that can influence each of the system components. Teachers must be able to monitor and improve their own learning and teaching.

A more important reason hearkens back to our focus on educational change: to the extent that we desire students who become life-long, independent, responsible learners and citizens, teachers must model effective learning processes in the classroom and involve students in them. Instructional assessments promote improved education to the extent that such assessments affect teachers' educational views, values, and practices. Even though teachers themselves often need strengthening, they become the center of the project efforts.

The instructional assessment project has been a promising catalyst for teacher professional development. We have worked extensively with a few collaborating teachers to help them understand the principles of instructional assessment, develop greater knowledge of science, implement the new instructional materials and tools, and work with other education professionals. An important component just initiated addresses the enhancements in infrastructure needed to support instructional assessments. These include team- and building-level organizational and policy issues. The roles of the principal and other administrators are centrally important to making adoption of instructional assessments successful.

We are interested in understanding and making more effective the interplay of instructional assessment and teacher development. We are interested in changes in practice, and especially in understanding the changes in strategies, perspectives, and values that underlie teachers' abilities to incorporate instructional assessments flexibly into their own teaching. Of central concern is the notion of transformation of teacher role towards more adaptive teaching, including skillful incorporation of assessment into teaching others to learn how to learn.

The use of instructional assessments have many benefits for teachers and students. We have observed several significant changes in the teachers that make us hopeful for large, deep changes. Teachers are encouraged to develop more detailed views of students' knowledge and more adaptive models of instruction. In their particular units, teachers quickly can analyze areas of relative strengths and weaknesses of student knowledge and plan curriculum and instructional choices accordingly. They change what they value or accept as evidence of good performances for grades. Some of these benefits result from enhanced components, some from systemic effects that arise from the combination of the whole. We begin with some observations of the types of changes seen in the teachers. We conclude with some tentative conclusions about the design of instructional assessments that contribute to such teacher development.

In the course of working with the teachers, we have collected data on teachers' professional behavior using observations, interview notes and transcripts, videotapes of the classrooms, and journals kept by the teachers. Data also included the revisions of the curriculum units and the teachers' written statements of their expectations and evaluations of the project gathered at various points. This section summarizes our observations of the impact of this partnership upon the teachers involved. These changes revolved around: (1) perceptions, (2) science knowledge, (3) awareness of students, (4) teaching methods and styles, and (5) professional interaction.

Perceptions of Curriculum, Instruction, and Assessment

The teachers have changed their perceptions, values, and practices about assessment to a greater or lesser degree. Instructional assessments change the assessment focus from grading or determining "who's better than whom" to content-based analyses focused on instructional action. Instructional assessments help teachers plan to be responsive to student conditions in systematic, informed ways. Both anecdotal accounts and observational records indicated that the experienced teachers we have been working with did not do preassessments before the project started but are all now actively involved in designing instructional assessments throughout at least one unit. The instructional assessments, then, provide a lever for moving teachers toward more flexible, student-centered teaching. Rather than relying exclusively on preset lesson plans, the teachers actively seek to know what their students are bringing and how they are developing. We have documented several instances where teachers did change their lesson plans and actual instruction based on the results of these instructional assessments.

Instructional assessments are foreign in practice to most of the teachers we have worked with or interviewed. Although formative evaluation is an old and oft-cited notion, in fact there are many barriers to actual use of instructional assessments in classrooms. Two key barriers the project is addressing are teacher beliefs and skill repertoires, which usually are more constrained than the instructional assessment information. If the teacher is not confident about his or her science knowledge and has had a difficult time presenting the information even initially, then the teacher has few options when the assessment shows that some students need additional or varied explanations. Adaptive teaching must be as rich as the assessment information. However, teachers are generally eager to expand their skill repertoires, within the boundaries of their general belief systems.

Science Knowledge

Another significant change observed through the project involved teacher knowledge in science. The instructional assessments require the teachers to focus on knowledge and skills that are central to building important conceptual models in science and to think through how these instructional objectives should develop and build together over the unit. Since the teachers need to analyze students' understanding and performances in terms of these models, the teachers themselves have been actively encouraged to deepen their science knowledge. The instructional assessment activities have encouraged the teachers to approach content areas they did not explore before and to involve students in science inquiry beyond cookbook experiments. Focusing assessment on understanding powerful themes and causal models, rather than on easily constructed and easily scored discrete bits of content knowledge, has been a powerful influence for upgrading the project teachers' knowledge of science. The resultant curricular and instructional deliberations indicated a change from a focus on "neat activities" and unconnected facts toward a concern for how the pieces fit together into a multirepresentational explanatory model.

Awareness of Students

Another teacher change that we have observed is a broadening of awareness of students' competence and performance. The instructional assessments emphasize varied representational forms that invite many types of communicative and performance competence. Teachers saw science in activities they had not seen before and are pursuing development of those cross-curricular connections. They also saw competencies in students that they were convinced would not have been manifest through traditional assessments.

For example, the many modalities of performance gave students who were less verbal or less aggressive in class opportunities to "shine." In the collaborating classes with significant proportions of special education students, the teachers were especially pleased with the changes in demonstrated competence; they felt that students both learn better and are able to communicate that learning through a range of assessments.

Teaching Methods and Styles

Several changes have been observed so far in the teaching methods employed by the project teachers. First, the teachers have encouraged more student inquiry than they did at the beginning of the project. Second, the teachers have been more willing than previously to explore with students areas that they do not fully command and to admit to their classes that they do not know the answer to a question. Third, they have been more willing than previously to allow students to explore ideas without a highly structured task to guide them. These indications of progress are small but promising.

Professional Interaction

Instructional assessment has the potential of providing a foundation for transforming science education at the classroom level into a self-improving system. However, we have encountered several challenging issues associated with implementation. We will briefly discuss some key issues concerning teacher development and school organization.

One of the clearest lessons from the study of teacher professionalization has been the organizational constraints on professional behavior. With nondifferentiated staffing and minimal aids for locating instructional materials, grading papers, and the like, teachers are pressed to fulfill their regular classroom obligations. Little discretionary time or energy is available during the school year for curriculum development, organizational planning, or even observing other teachers teaching. Instructional assessments generally result in more detailed knowledge of individual students, yet teachers have few resources and models for doing other than whole group instruction in science. Management challenges are often of foremost concern to the teachers. Impoverished resources sometimes lead teachers to depend on textbooks. Team, building, district, and state requirements or traditions sometimes constrain the curriculum by specifying what should be taught in particular grade levels. Parents may not expect much science to be learned in school.

We have tried to promote increased professional interactions as a solution to many of these problems. Some solutions are spurred by changes in building policy, but many of the problems are solved as teachers form a supportive professional community. They can discuss questions of pedagogy, consult about challenging situations, communicate about needed materials and resources, coordinate rooms, materials, and time, and share successes and failures. Establishing and extending such communities is essential, not only within classrooms, but between classrooms, schools, and the larger communities beyond school walls.

Initially all of the teachers indicated a lack of involvement in schoolwide professional activities for their schools. Principals were referred to as "others" and little interest was shown in attempts to coordinate science teaching with other school subjects. However, interest began to develop in coordinating the science curriculum with other areas, especially language arts and social studies. From this collaboration an integrated curriculum has emerged, combining math, science, social studies, and language arts, with science and social studies as the base. In one school two project teachers are working with the principal to bring together the language arts and science staffs to plan mutually reinforcing connections between the two subject areas in a schoolwide cross-curricular effort.

CONCLUSIONS

Motivating Change

We cannot claim ideal science teachers have been miraculously created in a few months through the partnerships that we have created. On the other hand, obvious changes have occurred as outlined above. The changes, significantly, are not just in one or two areas, but tend towards deep and systemic changes of practices and conceptions. Moreover, there are indications in each area of internalization of principles and responsibilities along with skills and knowledge so that the teachers may be able to sustain the changes when the interaction with the research project staff is decreased. The changes in the various areas of content, pedagogy, professional interaction, and so on, are made coherent through the attention to instructional assessments. The focus on instructional assessments provides a context within which to work on these systemic problems, and a mechanism for helping teachers develop and change their educational system.

Through working on instructional assessment, teachers can develop their professional knowledge and expertise, while addressing major components for systemic change: curricula, standards, school organization, pedagogy, science content, and professional interactions. Instructional assessment provides a motivation for engagement and risk-taking. It provides a clear vision focused on the student; change is grounded in working for changes in learning and teaching. It makes the role of the teacher a participant in change, a change agent. It provides information and feedback directly connected with the enterprise of teaching, rather than money, status, or other indirect motivators or empowering factors.

Our experience thus far leads us to believe that instructional assessment can be a powerful leverage point for teacher change and educational reform. We are particularly interested in four aspects of the experience that provide some insight towards explaining how instructional assessments can have the transforming and renewing effects we have observed.

Congruence of Educational Purpose

Teachers readily accept instructional assessment because of a perceived congruence of purpose. The teachers see instructional assessments as being intended to help teachers do what they want to do, namely, to help students learn and help teachers teach better. The instructional assessments are perceived as teacher-promoting rather than teacher-proof.

Integration of Essential Concerns and Activities

Instructional assessments provide a powerful focus for systemic change. Curriculum, instructional practices and deep beliefs about teaching, learning, science, assessment, and education are naturally integrated in the concerns about what should be learned, how to determine students' progress, and what to do on the basis of on-going assessment information.

Natural Development of Deep Changes

Instructional assessment supports an incremental development for the teachers of deep changes. Learning and development take place within the context of applied environments, with close attention to how students will benefit. For example, the assumption that on-going assessment will entail change and that teachers should expect to make adjustments promotes a responsive pedagogy and a change away from the view that the science teacher should be a repository of "proved truths."

Support and Responsibility for Adaptive and Localized Communities

Instructional assessment engages the teachers in support systems that build up communities of efforts. These communal efforts provide support towards making changes self-sustaining, renewing, and propagating. Instructional assessment inherently involves teachers in discussion with students and with teachers within and across grade levels. It also involves building and district administrators in the discussion of substantive questions such as curriculum continuity and performance standards across classrooms and between grade levels and subject matter areas.

Alternate Views of Assessment's Role in Educational Reform

In instructional assessment we have emphasized the role of assessment in providing information at the classroom level. In doing so we assume that teachers and students need such information because they should be the ones who make many of the changes needed to improve learning and teaching. However, this view of assessment and responsibility to change at the classroom level may be contrasted with other models of how assessment might be used to drive change in the educational system. The argument essentially revolves around the structural relationship between assessment and the classroom. We acknowledge that there are many people who see assessment as having an important role to play in educational reform, but who place assessment external to the classroom or who downplay the importance of having assessment provide a feedback loop.

One prominent model of the role of assessment in educational reform might be called the carrot-and-stick approach. This view holds that people will not change unless threatened by sanctions or enticed by rewards. Assessment holds people accountable to certain standards and provides the basis for administering punishments and rewards by providing clear statements of discrepancies between goal and outcome. What happens between the setting of goals and measurement of outcomes is a "black box," with which assessment is not concerned. Stated simply and positively, the carrot-and-stick approach says that one has only to set the goals clearly and establish a strong incentive system, and the system will perform as desired. Since the rewards and punishments are strongly tied to performance on the assessment, the assessment becomes the definition of the goals, and the assessment's value becomes the positive or negative reward it brings. Many state accountability testing approaches implicitly embrace this behaviorist (and black box) view of educational reform. One objection to this carrot-and-stick view is that the combination of strong incentives and remote control often leads to subversion of the assessment; this is the basis for much of the criticism of current "high-stakes" assessments today.

Why not then create assessments that set worthy targets, but are by their nature such that they coopt subversion? The authentic target model holds that certain types of assessments are intrinsically difficult to subvert—that working towards them would constitute the desired educational experience and that doing well in them would embody the goals of education. The push to develop "authentic assessments," largely based on performance tasks, reflects this view. An emphasis on authentic assessments often assumes that the system will correct itself in pursuing the goal. One objection to this authentic target view of educational reform is that in science education, teachers, students, and administrators often need more than clear statements of the assessment tasks—they need training and support so they will be able to do well on the assessment tasks.

A third approach creates the specter of assessments to motivate the adoption of curricular and instructional practices along desired lines. This model, which might be called political message, differs from the preceding views in that assessments are seen primarily as tools for inducing change, not as providing information about performances to act upon. Underlying this model is a transformation of the role of assessment from providing feedback that a cybernetic system can use to correct itself, to assessment which sends a political (or perhaps educational) message, but is not expected to return instructional information to the teacher. Some rhetoric surrounding current state reform efforts in science cast the state assessment efforts in this role of political message sending. One objection to this view is that educators need assessment information so the system can be self-correcting; classrooms in particular should not rely on curriculum or instruction, however visionary, that is accompanied by political inducements posing as assessment.

These three views of assessment—as carrot-and-stick, authentic target, and political message—do not address the needs for sustained improvement, particularly at the classroom level. They do not address how teachers and learners can change. Indeed, these models may be viewed as noneducative, in that they treat change due to assessment as occurring primarily outside of a learning and teaching context.

THE REMAINING AGENDA

The implementation of instructional assessment and corresponding professional development of teachers will successfully take place within the context and infrastructure of institutional commitment to long-range usage of instructional assessment information and institutional development in science education beyond the sum of professional development of individual teachers and independent curricular units and activities. Teacher development for the near-term at least must be adaptive within these varied organizational structures and contexts. The challenge is to deal with systemic problems in a way that promotes the role of the teacher as a change agent, often when the teacher has significant needs for development. In addition, the heart of professional expectations and competence involves teachers' competence in assessing, judging, and acting upon their views of what students know, need, and can become. Perhaps more than any career ladder, teachers' professional standing will be recognized and ensured when classroom assessment is recognized and made a legitimate and respected part of our educational efforts to help students learn and to make systemic improvements in our educational system.

Assessment has a crucial role to play in educational improvement, particularly in areas such as science where high goals and major challenges exist. We hope that in bringing together research on learning, examination of subject matter, and technical tools, assessment may fulfill more of its promise to be a constructive force in improving education. We are persuaded that to improve science education, teachers must be given more support than "clear national objectives" and "mandated accountability standards." To achieve the national goals concerning science, much work needs to be done that will be educative in nature, supporting learning and teaching in the classrooms. Instructional assessment appears to be a promising, supportive lever for teacher development and curricular improvement. Instructional assessments can provide a model for local initiative and provide substantive support in integrating instruction, curriculum, and assessment. As the nation tries to upgrade its science education and searches for educationally valid conceptions of assessment, efforts such as this project will help promote a clearer view and give some empirical experience on profitable ways to approach those important tasks.

ACKNOWLEDGMENTS

The Intermediate Science Project has been funded by ETS since 1988 as part of ETS's corporate efforts to develop new ways to "put assessment in the service of instruction" (ETS, 1991). A goal of the Intermediate Science Project is to explore issues and models of instructional assessment in the context of middle school science. Brian Gong and Richard Venezky are codirectors of the project activities centered in Delaware. Brian Gong is a research scientist in ETS's Division of Cognitive and Instructional Science. Richard Venezky is Unidel Professor in the Department of Educational Studies at the University of Delaware. David Mioduser, a postdoctoral fellow sponsored by the project, graduated from Tel Aviv University, where he worked on the Israeli national science curriculum.

The work reported here has drawn on the talents of many people. We especially acknowledge the contributions of Susan Coffin, Wes Covalt, Mary

Gribbin, Laura Harley, Tammy Murdock, and John Sabatini, who have been associated with the project as Research Assistants and developers, and the students, teachers, and administrators of the collaborating schools.

REFERENCES

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

Archbald, D. A., and Newmann, F. M., (1988). *Beyond Standardized Testing: Assessing Authentic Academic Achievement in the Secondary School,* National Association of Secondary School Principals, Reston, Virginia.

California Department of Education, (1990). *Science Framework, 1990,* Bureau of Publications, Sacramento, California.

Department of Education, (1991). *America 2000: An Educational Strategy*, Washington, D.C.

ETS, (1991). (Educational Testing Service). *ETS 1991 Annual Report.* Princeton, New Jersey.

Frederiksen, J. R., and Collins, A., (1989). A systems approach to educational testing. *Educational Researcher* 18: 27–32.

Lesh, R., Lamon, S., Gong, B., and Post, T., (1992). Using learning progress maps to improve educational decision making. In Lesh, R., and Lamon, S. J. (Eds.), *Assessments of Authentic Performance in Elementary Mathematics,* AAAS Press, Washington, D.C. (in press).

McCloskey, M., (1983). Naive theories of motion. In Gentner, D., and Stevens, A. L. (Eds.), *Mental Models,* Lawrence Erlbaum Associates, Hillsdale, New Jersey, pp. 299–324.

NSBC (NSBC National Science Board Commission on Precollege Education in Mathematics, Science and Technology), (1983). *Educating Americans for the 21st Century: A Report to the American People and the National Science Board,* National Science Foundation, Washington, D.C.

NSF (National Science Foundation), (1990). *Statewide Systemic Initiative,* Program announcement, NSF, Washington, D.C.

Resnick, L., (1991). Examinations and learning. Presentation at the American Educational Research Association annual meeting, April 3–7, 1991, Chicago, Illinois.

Shepard, L., (1991). Interview on assessment issues with Lorrie Shepard. *Educational Researcher* 20: 21–23, 27.