C R E S S T

National Center for
Research on Evaluation,
Standards, and
Student Testing

Technical Report

You can view this document on your screen or print a copy.

UCLA Center for the Study of Evaluation

in collaboration with:

- University of Colorado
- NORC, University of Chicago
- LRDC, University of Pittsburgh
- The RAND Corporation

Cognitive Theory as the Basis for Design of Innovative Assessment: Design Characteristics of Science Assessments

CSE Technical Report 349

Robert Glaser and Kalyani Raghavan
LRDC, University of Pittsburgh

Gail P. Baxter
University of California, Santa Barbara

September 1992

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Graduate School of Education
University of California, Los Angeles
Los Angeles, CA 90024-1522
(310) 206-1532

The work reported herein was supported under the Educational Research and Development Center Program cooperative agreement R117G10027 and CFDA catalog number 84.117G as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

COGNITIVE THEORY AS THE BASIS FOR DESIGN OF INNOVATIVE ASSESSMENT: DESIGN CHARACTERISTICS OF SCIENCE ASSESSMENTS

Robert Glaser and Kalyani Raghavan
Learning Research and Development Center, University of Pittsburgh

Gail P. Baxter
University of California, Santa Barbara

Introduction

Efforts to develop more contextualized, direct measures of student achievement at national, state, and local levels are focusing on creating assessment situations that display how students use their knowledge to reason and solve problems. The assumption is that these complex, open-ended, performance assessments require students to engage in higher-order thinking processes in developing a solution to the problem, and that the scoring systems characterize these processes in such a way that differential levels of student competence can be ascertained. For the most part, these assessment situations were developed on the bases of rational and intuitive analyses of the processes that underlie performance rather than on empirical evidence of the kind of thinking that occurs. What is needed is an examination of the kinds of knowledge and cognitive processes that are actually being tapped by these performance assessments. Documentation of the match or mismatch between the skills and processes assessment developers intend to tap and those actually elicited provides empirical evidence bearing on the cognitive validity of these assessments.

The goal of this project is to investigate forms of reasoning and problem solving required of students by innovative assessment projects in science. A necessary first step is to describe various assessment tasks by the underlying processes which define the nature of task difficulty and the problem-solving activities that contribute to effective performance in these assessment situations. The long-range intent is to develop guidelines for designers of assessment situations about the ways in which student performance can be elicited and scored to ensure that appropriate cognitive skills are actually involved.

Survey of Assessment Programs

The initial phase of the project identified activities and school systems where new forms of science assessment were being piloted, collected the assessment materials being used, and gathered information including the rationales and frameworks for assessment development. Although we initially proposed to study science assessment at the middle school, innovative assessment projects at other grade levels have been collected and examined because they present interesting ways of analyzing the components of higher-order skills. To date, information has been obtained from a broad spectrum of projects including: a pilot study of higher-order thinking skills assessment techniques in science carried out by the National Assessment of Educational Progress at ETS (Blumberg, Epstein, MacDonald, & Mullis, 1986; NAE, 1987), ACT Science Reasoning Test (ACT, 1989), Connecticut's Common Core of Learning Assessment Project (Baron, 1991), California Assessment Program (New CAP Assessments, 1991), and the University of California, Santa Barbara/California Institute of Technology research project "Alternative Technologies for Assessing Science Understanding" (Shavelson, Baxter, Pine, & Yurè, 1990).

The statements of objectives of the aforementioned programs pay explicit attention to those performances that many educators believe are important aspects of reasoning in science, such as designing an experiment, analyzing and interpreting data, drawing inferences, and the like. These objectives served as guiding frameworks in the development and field testing of assessment situations. When viewed as "work samples" of scientific performance, these assessments have obvious face validity.

In science assessments, scoring categories of higher levels of performance, such as "analyzes scientific procedures and data" and "integrates specialized scientific information," are defined in terms of illustrative test items and not explicit description of the processes that underlie these performances. Although some analyses have been carried out in the course of item development, higher-level performances are defined primarily by difficulty in a psychometric sense, and less by underlying processes which define the nature of difficulty and the problem-solving activities that contribute to effective performance. To ensure adequate cognitive validity, an important question to analyze in studying these tests is: What kind of performance is actually elicited from students, and how does this performance differ among students at various levels of achievement?

Framework for Assessment Development

We have reviewed the kinds of innovative science assessments being developed and the rationales and frameworks behind these new achievement testing programs. By and large, there is careful delineation of important topics and concepts, the "big ideas," in various domains of science. Imposed on these topics is concern about processes of scientific reasoning, performance with understanding, application of knowledge to situations for further learning in school, and ability to understand and interpret events encountered in everyday life.

The task of this project is defined in this context—to carry out analyses that contribute to making the development of assessment procedures more targeted in tapping the kind of cognitive skills that underlie assessment objectives. Teachers and test developers would then have more guidance than is usually available about the details of the situations that they design and the ways in which students' performance can be elicited and scored to ensure that appropriate cognitive skills are actually involved. For example, at various levels of proficiency, students' performances display different forms of understanding. Students who have not yet acquired integrated knowledge of a concept will represent an assessment problem at a more surface-feature level than will a student with more advanced knowledge, who will perceive the problem in terms of underlying principles. How can assessment situations be designed to elicit such differences in performance?

Detailed investigations of a selection of assessment situations will be conducted using protocol analysis techniques that have become standard for studying the cognitive aspects of problem solving (Chi, Bassock, Lewis, Reimann, & Glaser, 1989; Chi, Glaser, & Farr, 1988; Chi, Glaser, & Rees, 1982; Ericsson & Simon, 1984). The match or discrepancy between descriptions of behavior and the actual cognitive processes that students carry out is an important issue in the development of assessment instruments that purport to be innovative in ways that tap higher-order thinking. This information should contribute to the design of assessment situations so that the translation of specifications into elicited performances can be more precisely accomplished.

Analysis of student protocols will be guided by a framework describing general dimensions of problem-solving performance along which individuals who are more or less proficient in a particular domain differ. These aspects of performance have been summarized as follows (Glaser, 1992):

1. **Structured, integrated knowledge**: Good problem solvers use organized information rather than isolated facts. They store coherent chunks of information in memory that enable them to access meaningful patterns and principles rapidly.

2. **Effective problem representation**: Good problem solvers qualitatively assess the nature of a problem and build a mental model or representation from which they can make inferences and add constraints to reduce the problem space.

3. **Proceduralized knowledge**: Good problem solvers know when to use what they know. Their knowledge is bound to conditions of applicability and procedures for use.

4. **Automaticity**: In proficient performance, component skills are rapidly executed, so that more processing can be devoted to decision-making with minimal interference in the overall performance.

5. **Self-regulatory skills**: Good problem solvers develop self-regulatory or executive skills, which they employ to monitor and control their performance.

The above general dimensions of performance will focus our investigations of the more specific cognitive skills of problem solving that students employ in assessment situations. The result of this work should provide information about the processes assessed or not assessed by current innovative assessment practices. It is anticipated that guidance for assessment development will consist not only of descriptions of cognitive aspects on which more or less proficient students vary, but also of the kinds of assessment situations in which performances of interest are likely to be elicited. Information of this kind would put test construction on a more efficient basis than the intuitions of good item design that currently are in place. The goal of this project is to assist these intuitions by further knowledge of the cognitive processes involved.

Study Sites

Assessment situations to be used for detailed study are drawn from programs reviewed in the initial phase of this project and include state level assessments for accountability and curriculum-embedded evaluations for monitoring instruction, as well as portfolio situations. We are working with individuals in the Connecticut program (Lomask, Baron, Greigh, & Harrison, 1992), with the California Assessment Program (CAP), with researchers and teachers involved in the "Transferring New Technologies to Teachers and Other Educators" project at the University of California, Santa Barbara, and with the "Portfolio Culture Project in Science Instruction" in the Pittsburgh schools (Duschl & Gitomer, 1991).

Connecticut Common Core of Learning

Two types of assessment tasks—Components I and II—have been designed to provide information on what students know and can do after 12 years of school (Baron, personal communication). Component I tasks integrate scientific methodology, use of models in science, and model-based reasoning in a challenging context. Component II tasks, on the other hand, deal with very basic concepts and their structural organization. The combination of the two tasks facilitates the assessment of students' content knowledge and understanding, problem-solving skills, and the use of the "scientific process." All tasks are administered and scored by teachers in their respective classrooms.

We have selected one Component I task, *Exploring the Maplecopter*, and three Component II tasks, *Growing plant*, *Digestion of a piece of bread* and *Blood transfusions*. These tasks were selected because they assess students' knowledge of some of the fundamental principles in physics and biology. Moreover, students revisit these topics several times during their schooling, typically beginning in fourth or fifth grade. It is expected then that various levels of competence will be observed.

Component I task. In general each Component I task consists of three or four parts, some requiring individual work and some, group activity. The first part introduces the task to the students, asking them to make some observations, provide a written description of the problem, make an initial hypothesis, and suggest possible ways of investigating the problem. The class is then divided into groups of three or four students. Each group pools the observations, ideas, etc. of each of its members. For example, students in a group may express differing opinions on which variables are salient in this particular task. As a group, they perform experiments to test their initial hypothesis, document the tests and observations they make, and provide written conclusions based on their experiments. The last part of the task is answered individually and consists of a set of follow-up questions related to the task, such as analyzing and critiquing a given set of data collected by an imaginary group on the same task, or performing a near-transfer task.

The *Exploring the Maplecopter* problem involves laws of motion, principles of aerodynamics, and the use of models in explaining scientific phenomenon (see Figure 1).

Students study the motion of maple seeds and design experiments to explain their spinning flight patterns over several (typically four or five) class periods. To encourage students to use models in their experimentation, directions for constructing a paper model of a helicopter are given to them. Students are then prompted to list the advantages and/or disadvantages of using models to explain the motion of maple seeds. The task does not have a clean, single solution. Rather students must rely on controlled experimentation and model-based reasoning to help them identify the causal variables involved so as to produce a convincing explanation of the “flight” of the maple seed.

Student performance on the maplecopter task is described with respect to one of four levels—Excellent, Good, Needs Improvement, or Unacceptable—on the basis of students’ records of their observations, statement of the relevant variables in this task, experimental design, data collection, presentation and interpretation, scientific explanation for the phenomena they observed, and conclusions (see Figures 2a and 2b). Within each of these general categories, teachers examine student responses for several critical aspects of the task. For example, a student’s initial individual report after observing the maple seed’s flight may include any of the following:

1. Two phases to the motion: free fall and spinning
2. Velocity of free fall phase is greater
3. Falls tilted with the seed end lower

EXPLORING THE MAPLECOPTER

Part I: Getting started by yourself
Throw a winged maple seed up in the air and watch it "float" down to the floor.
Describe as many aspects of the motion of the pod as you can (you may add a diagram if you wish).
1. Record all observations that you have made. Do not explain the winged maple seed's motion at this time.
2. Try to explain how and why the winged maple seed falls as it does.

Part II: Group Work
1. Discuss the motion of the winged maple seed with the members of your group. Write a complete description of the motion, using the observations of the entire group. (You may add a diagram if you wish.)
2. Write down all the factors that your group thinks might affect the motion of the winged maple seed.
3. Design a series of experiments to test the effects of each of these factors. Identify which of these experiments you could actually carry out.

Part III: Finishing by yourself
1. Suppose you want to explain the motion of a winged maple seed to a friend who has not yet studied high school physics. Write an explanation that is clear enough to enable your friend to understand the factors and forces which influence the motion of the winged maple seed. Specify the aspects about which you are more certain and those about which you are unsure.
2. In this activity you used simplified models to help explain a more complicated phenomenon. Explain all of the possible advantages and disadvantages of using models in studying the motion of a winged maple seed. Include specific examples from the models your group used.
Given a set of data generated by a group of students working on the maplecopter task, read the report and answer the questions:
3.a. Discuss the information given and how it is organized. Do you think it is complete enough for you to replicate the experiments? If not, what else do you need to know?
3.b. Can any valid conclusions be made regarding variables studied in this experiment? If so, explain fully what they are.

Figure 1. The maplecopter task. (Developed by the Connecticut Common Core of Learning Performance Assessment Project)

4. Rigid edge of the wing is the leading edge
5. Spins around an axis that is through a point in the seed part
6. Spins either wing side facing up
7. Spins either clockwise or counterclockwise
8. Motion different with different starting positions

Exploring The Maplecopter - Scoring: Individual Student Name ____________

Part I: Getting Started by Yourself

	Excellent	Good	Needs Improvement	Unacceptable
1. Make observations about the motion of the "maplecopter."	☐ 6 or more	☐ 4 - 5	☐ 2 - 3	☐ 0 - 1

1. Two phases: free fall & spinning.
2. Velocity of free fall phase is greater.
3. Tilted with seed lower.
4. Rigid edge of the wing is leading edge.
5. Spins around axis in seed.
6. Spins either side facing up.
7. Spins either clockwise or counterclockwise.
8. Motion different w/ diff. starting positions.

Others:

Part III: Finishing by Yourself

	Excellent	Good	Needs Improvement	Unacceptable
1. Explain the motion of the "maplecopter."	☐ E	☐ G	☐ NI	☐ U

Holistic judgement based on the following:
1. Reference to or consistency with conclusions from experiments.
2. Inclusion of the forces and factors studied.
3. Explanation of physics concepts is clear and appropriate to specified audience.
4. Lack of misconceptions.

	Excellent	Good	Needs Improvement	Unacceptable
2. Explain the advantages and disadvantages of models.	☐ 4	☐ 3	☐ 2	☐ 0 - 1

Explanation should be based on the following criteria:

Advantages:
1. Materials are cheaper or more readily available/nondestructive of original
2. Easier to control and manipulate variables/uniformity of models.

Disadvantages:
1. Parameters of model are not the same as the "maplecopter" (i.e. shape, materials, etc.)
2. Uncertainty about the generalizability of results from model to original.

	Excellent	Good	Needs Improvement	Unacceptable
3.a. Critique information given in sample group report.	☐ 5	☐ 4	☐ 3	☐ 0 - 2

The following deficiencies should be described:
1. No definition of dependent variable.
2. No description of method.
3. Poor description of independent variables.
4. No description of model.
5. Poor organization of data.

	Excellent	Good	Needs Improvement	Unacceptable
3.b. Drawing conclusions from the sample group report	☐ 4	☐ 2 - 3	☐ 1	☐ 0

Tentative conclusions can be made about the effects of the following:

1.a. Length of wing.
1.b. Added mass.
1.c. The stiffness of wing.
2. Conclusions are tentative due to uncertainty about accuracy of measurements.

Figure 2a. Scoring system (Individual) for the maplecopter task. (Developed by the Connecticut Common Core of Learning Performance Assessment Project)

Exploring The Maplecopter - Scoring: Group

Student Names ____________________

Part II: Group Work

Criterion	Excellent	Good	Needs Improvement	Unacceptable
2. Identify factors that might affect the motion of the "maplecopter"	☐ 6 or more	☐ 4 - 5	☐ 2 - 3	☐ 0 - 1
1. Mass of seed and wing. 2. Surface area of wing. 3. Distribution of mass between seed & wing. 4. Curvature of wing. Others: 5. Moisture level of seed and wing. 6. Initial dropping position. 7. Air (currents, pressure, humidity.) 8. Gravity.				
3. Design complete experiments for the "maplecopter."	☐ 4 Expts.	☐ 3 Expts.	☐ 2 Expts.	☐ 0 - 1 Expts.
The experiments designed should match the factor to be studied, independent and dependent variables should be defined, variables should be controlled and tested separately				
5.b. Gather and organize data from experiments with models.	☐ All 6	☐ 4 - 5	☐ 2 - 3	☐ 0 - 1
Students' work should be reasonable and appropriate on the following criteria: *Quality of measurements:* 1. Accuracy of data. 2. Repitition of experiment (until data are replicated.) *Manipulation and presentation of data:* 3. Clarity and organization (e.g. proper labels, units, scaling, etc.) 4. Appropriate symbolic representation (e.g. use of bars graphs vs. Cartesian coordinate graphs.) *Use of mathematics:* 5. Making calculations (e.g. taking averages) 6. Correct use of formulas to define new terms (e.g. velocity, forces, surface area.)				
5.c. Draw conclusions from experiments.		☐ Yes	☐ No	
Yes - Conclusions made are consistent with and supported by the data collected. No - Conclusions made are not consistent with or supported by the data collected.				
6. Control variables during experiments with models.		☐ Yes	☐ No	
Yes - Indication that students have either attempted to control variables or have considered how this might affect their results. No - No indication of the above.				

Figure 2b. Scoring system (Group) for the maplecopter task. (Developed by the Connecticut Common Core of Learning Performance Assessment Project)

If the student had 6 or more of these observations, he/she will be classified as excellent; between 4-5 of these earned a rating of good, whereas a student is rated as fair for listing 2 or 3 of these items. An overall individual grade and a group grade are then assigned by averaging the levels of performance on the different categories.

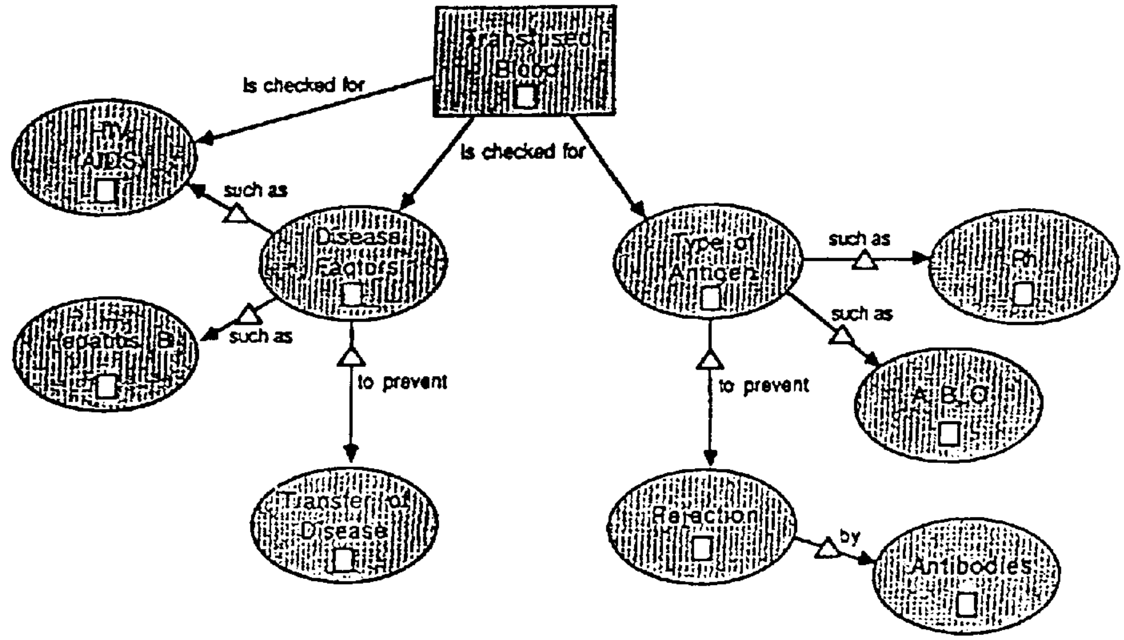
Component II tasks. The purpose of these tasks is to assess whether students possess a deep understanding of particular concepts as evidenced by a coherent and cohesive narrative or whether they possess fragmented pieces of knowledge as evidenced by a set of unconnected statements. Students respond individually to several open-ended questions or interpret a science passage in free format. Three tasks serve as examples: *Growing plants*—describe the types of energies and materials involved in the process of a growing plant and explain how these energies and materials are related; *Digestion of a piece of bread*—describe the possible forms of energy and types of materials involved in the digestion of a piece of bread and explain fully how they are related; *Blood transfusions*—state what you would want blood to be checked for and explain why the blood should be checked for each of these if the blood is to be used for a transfusion.

Concept maps are the basic evaluation tool for these Component II tasks. These concept maps provide a pictorial representation of concepts involved in a phenomenon and how these concepts are interrelated. Teachers construct a concept map for each student based on written responses to each question. An expert's (teacher's) concept map serves as a "template" against which student performance is evaluated For example, Figure 3 is the expert map for the blood transfusion task. Scoring focuses on two structural dimensions—size and strength. Size is defined as the number of core concepts included in the student's concept map over the total number of core concepts in the expert's concept map. Strength is defined as the number of valid connections in the student's concept map over the number of possible connections. The number of possible connections is defined in terms of the number of concepts included in the student's answer and not the total number in the expert's concept map. In other words, if the student did not mention particular concepts, then he/she would not be penalized for failing to provide information about the connections among these concepts. The strength thus indicates if the student knows the

Type 1 Scoring Form

Rater Initials:____________ Student Id. Number:____________

Concept Map:



Strength

Size	Strong	Medium	Weak	None
Complete	5	4	3	2
Substantial	4	3	2	1
Partial	3	2	1	1
Fragmentary	2	1	1	1
Irrelevant	1	1	1	1

X = —/9

Y =

Z =

Final Score ——

Figure 3. Concept map and scoring system for the blood transfusion task. (Developed by the Connecticut Common Core of Learning Performance Assessment Project)

"full story" at least about the concepts mentioned in the response. Size and strength scores are combined to reflect level of student understanding.

California Schools

Three assessments will be evaluated, each with a distinct purpose and focus: (a) California Assessment Program's (CAP) pilot of a statewide performance assessment at the fifth-grade level. (b) An electric circuits task developed as part of a study of the psychometric properties of hands-on investigations and alternatives that may serve as surrogates (Shavelson, Baxter, Pine, & Yurè, 1991). Results from this study suggest that student performance varies with method of presentation (Baxter, 1991). It is important therefore to examine the behavior of students conducting the same investigation under varying methods of task presentation. An electric circuits hands-on investigation being used in the classroom has been simulated on a Macintosh computer. And (c) a mystery powders assessment developed as part of a research project "Transferring New Technologies to Teachers and Other Educators" currently in progress at the University of California, Santa Barbara.

California Assessment Program. Because this test will be administered and scored by volunteer teachers in the state over the next month, details of the task cannot be provided here. Generally, however, the task is comprehensive in nature, requires both individual and group (pairs) work, and will require two class periods to administer. Teachers in the state volunteer to administer and score the assessments for students in their class.

UCSB/CalTech project. This project undertook to develop hands-on assessments and less costly surrogates. An electric circuits problem-solving task and a computer simulation surrogate were developed and evaluated. These assessments are now being used in one California school district as end-of-unit tests for a science unit on batteries and bulbs taught at the fifth-grade level. Teachers administer and score the assessments.

For the *hands-on* assessment, students are presented with six weighted boxes, each of which contains one of five possible circuit components. Using a collection of five wires, two batteries, and two bulbs, students have to determine the contents of each box from a list of five possible alternatives (two batteries, wire, battery and bulb, bulb, nothing). Two of the boxes have the same contents (a wire). All of the others have something different (see Figure 4).

Students record their answers, draw a picture of the circuit used to arrive at the answer, and provide a written explanation of how they knew what was in a given box. Performance is scored on the basis of student's written responses. One point is given for each box if the student provides the correct answer using the correct circuit or sequence of circuits to arrive at the answer for that particular box.

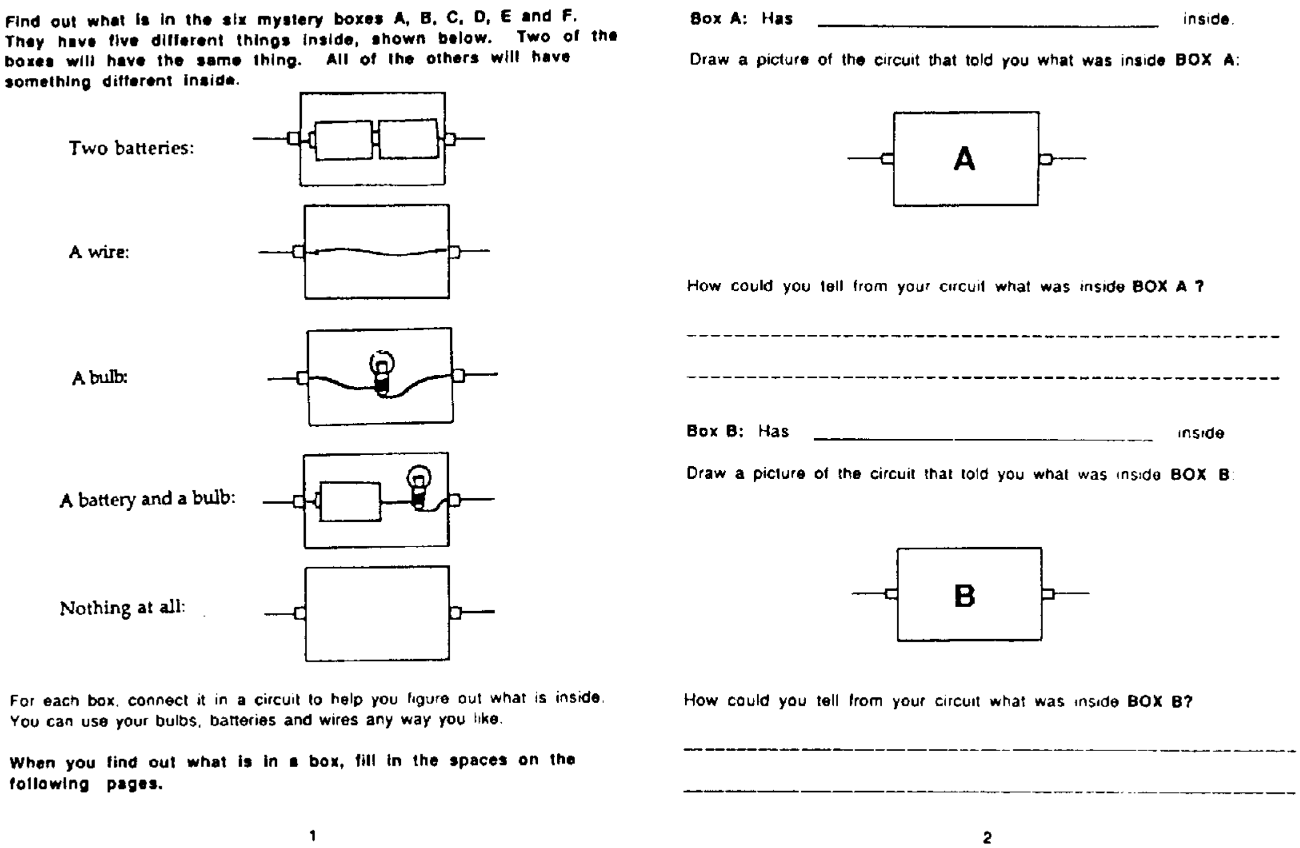


Figure 4. Hands-on electric mysteries investigation.

A *computer simulation* of the electric mysteries hands-on investigation described above was developed in a Macintosh environment so as to replicate as nearly as possible the hands-on investigation. Rather than manipulate batteries, bulbs, wires, and mystery boxes, students now manipulate icons on a computer screen. The display screen is divided into three sections: equipment, work space, and control panel. On the left side, a selection of equipment (battery, bulb, six mystery boxes) is presented. Students can drag the equipment they want into the work space area in the middle of the screen using the mouse. Clicking on one of the terminals displayed as black dots on a pair of equipment pieces produces a wire connecting them. Students can thus connect a number of circuits on the screen at once. Alternatively, they could leave one completed circuit on the screen for comparison. At the bottom of the screen, students can type in notes for themselves such as what they thought was in a box (see Figure 5).

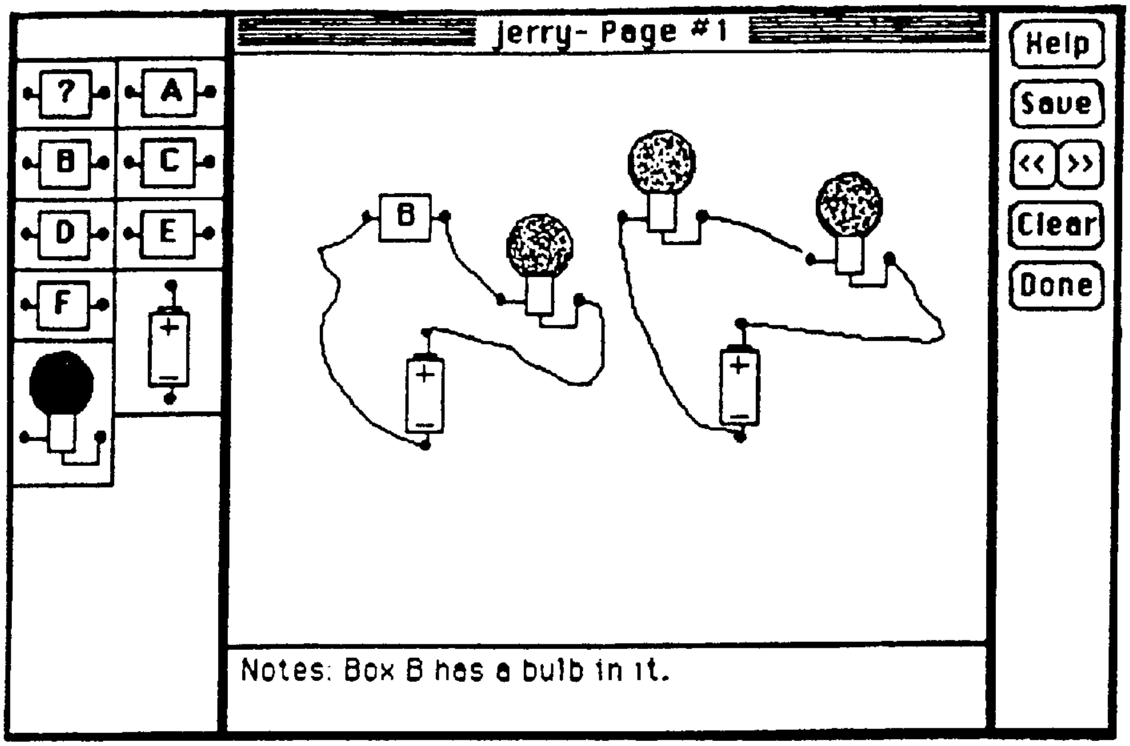


Figure 5. Electric mysteries —computer simulation.

The software emulates the behavior of a real circuit. For example, a bulb connected to a box that contained a battery and a bulb would appear dimmer than a bulb connected to just a battery. On the right are a set of control buttons to clear the screen, save, quit, or use arrow keys to scroll through the document for review.

As in the hands-on investigation, students are asked to determine the contents of six mystery boxes from a list of five alternatives. A computer record of student activities is maintained allowing a play-back facility for scoring purposes. The scoring is the same as that used for the hands-on activity, with one point for each box correctly identified with the help of one or more correct circuits.

Mystery powders. University of California, Santa Barbara (UCSB) is currently working with teachers in a large urban school district to conduct a study of curriculum embedded assessments for the purpose of monitoring instruction. Using a hands-on instructional approach, teachers use district-supplied kits to teach various scientific concepts and procedures. The UCSB project is developing assessments for teachers to use to evaluate whether students have learned specific concepts or procedures that a particular kit was designed to teach. For example, in the mystery powders unit, fifth-grade students work with five substances (sugar, salt, baking soda, cornstarch, and plaster of paris) over a period of six weeks. They observe each of the white substances under various conditions (e.g., one day after water has been added), systematically recording their observations in a lab notebook on a daily basis.

At the end of this six-week unit of study, students are presented with six bags and asked to conduct tests to determine the contents of each bag (see Figure 6). Some of the bags contain two substances (i.e., cornstarch and baking soda). Others contain a single substance (baking soda). Students work in pairs, conducting tests on each of the substances, recording their tests and observations as they proceed. When students feel that they have sufficient information to determine the contents of each of the bags, they are prompted to "use your lab notebook from class and the notes you took today to help you determine what each mystery powder is."

Scores are based on student observations, tests conducted and identification of the contents. For each of the six substances, students are given one point for correctly identifying the contents, and one to four points based on the completeness of the evidence provided (see Figure 7). In general, students must provide all the necessary evidence to distinguish one substance from each of the other substances. For example, to get four points for powder "A" (cornstarch and baking soda), a student must state that he/she added vinegar to the substance and it fizzed, and that he/she added iodine to the substance and it turned purple. Other combinations of tests and observations result in lower scores.

Data Collection and Analysis

Extended interviews with a sample of students following each of the assessment situations described above are in various stages of completion. Preliminary interviews have been conducted with students in Connecticut, and arrangements have been made with schools to conduct interviews with students taking the CAP assessment and each of the embedded assessments—electric mysteries and mystery powders.

MYSTERY POWDERS

Name: ____________________

You have six bags of powders in front of you and materials to do some tests. Find out what is in each of the bags A, B, C, D, E, and F.

Each bag has one of the "Mystery Powders" listed below:

- **Baking Soda**
- **Cornstarch**
- **Cornstarch and Baking Soda**
- **Sugar and Baking Soda**
- **Salt and Baking Soda**

Two of the bags will have the same thing. All the rest will be different. All of the mystery powders are things you might use for cooking.

Use any of the equipment on the table to help you determine what is in each bag.

Keep notes on what you did and what you found out on the following pages.

FINDINGS

Use your lab notebook from class and the notes you took today to help you determine what each mystery powder is. Fill in the table with your answers.

Mystery Powder	What's inside	What test(s) told you	How did you know (what happened)
A			
B			
C			
D			
E			
F			

Figure 6. Mystery powders.

MYSTERY POWDER SCORE FORM

Scorer ____________________ Student ____________________

Substance		+(X= essential set)	Indicator(s)	Observation(s)	
Cornstarch and Baking Soda (A)	1 0	(X)	iodine	turns purple, black,...	4 3 2 1 0
		(X)	vinegar	fizzes, bubbles	
			water	doesn't dissolve	
			touch	smooth, not grainy	
			sight	no crystals	
			taste	bitter	
Baking Soda (B)	1 0	X	iodine	turns yellow, not black	4 3 2 1 0
		(X)	vinegar	fizzes	
			water	dissolves	
		X	touch	no crystals, smooth	
		any of	sight	no crystals	
		these	taste	not sweet or salty	
Salt and Baking Soda (C)	1 0		iodine	turns yellow, not black	4 3 2 1 0
		(X)	vinegar	fizzes	
			water	dissolves	
			touch	grainy, not smooth	
			sight	has crystals, grainy	
		(X)	taste	salty, like salt	
Cornstarch (D)	1 0	(X)	iodine	turns purple, black	4 3 2 1 0
		X	vinegar	doesn't fizz	
			water	doesn't dissolve, turns gluey	
			touch	smooth, not grainy	
			sight	a powder, no crystals	
			taste	not sweet, not salty	
Sugar and BakingSoda (E)	1 0		iodine	turns yellow, not black	4 3 2 1 0
		(X)	vinegar	fizzes	
			water	dissolves	
			touch	grainy not smooth	
			sight	grainy	
		(X)	taste	sweet, sugary, like sugar	
Cornstarch and Baking Soda (F)	1 0	(X)	iodine	turns purple, black,...	4 3 2 1 0
		(X)	vinegar	fizzes, bubbles	
			water	doesn't dissolve	
			touch	smooth, not grainy	
			sight	no crystals	
			taste	bitter	

Total What's Inside: ____ Total What Test(s) Told You/ How Did You Know? ____

Figure 7. Scoring system for mystery powders task.

Connecticut

A group of researchers from this project visited a high school in Connecticut in March while the maplecopter task and the three Component II tasks were being administered to several sections of seniors. Nine students were interviewed and audio-taped each day as they progressed through the maplecopter task. Questions were guided by the students' answers to the part of the task they had completed that day. Questions focused on getting students to explain and elaborate on their written responses which included: initial observations and problem representation, list of causal variables, experimental procedures and their purpose, understanding of the use of models in explanations, and the final conclusions. In addition, they were asked to list and explain all the physics concepts learned in school that they thought were involved in the task and how they were related to the task at hand. Twelve students from three different science backgrounds (AP biology, human biology, and geology) were interviewed after they answered the Component II free-form response tasks described above. Questions were asked to try to discern what distinctions the student makes among the concepts that he/she has mentioned in the answer and how he/she thinks the different concepts are linked. On the basis of these in-depth interviews and student protocols, do we arrive at a "student" concept map that differs in any way from that constructed on the basis of the written responses only? The work on concept maps in learning and evaluation by Novak and Gowin (1984) will be helpful in our analysis.

California

Arrangements have been made to interview six students after they have completed the fifth-grade CAP assessment during the first week of May. For the CAP assessments, students conduct parts of the investigations in pairs and then, on the following day, write their own interpretations of the results. Consequently each member of the pair will be interviewed separately. The embedded assessments (electric mysteries hands-on and computer, and the mystery powders) will be administered by teachers at the end of the corresponding unit of study (June).

With respect to the electric mysteries assessments, particular attention will focus on the differential performance of the students with the two methods of presentation. Student performance on the computer will be played back so students can talk through their performance and the interviewer can question students on particular aspects of their performance. For example, “Can you tell me why the bulb did not light when you put two batteries in the circuit?”

For the mystery powders assessment, again students work in pairs. Does the performance reflect the understanding of both students or only the brighter student? Do students rely on their previous work with the substances to help them draw their conclusions on the assessment? Do students show that they understand the need to have conclusive evidence? Do students use all the information available to them when drawing their conclusions or, for example, do they just rely on tasting the powders?

Regardless of the particular assessment, protocols will be analyzed with respect to the following: (a) Student’s representation of the problem. Does the student understand the problem as the test developer intended? (b) Reasoned problem solving. Does the student use a trial and error approach or does the student recognize that he/she has particular knowledge and skills that are appropriate for solving the given problem? (c) Self-monitoring. Does the student check his/her thinking as problem solving progresses, or does he/she set a course in motion and pursue it to the end? How does the student know when the task is complete? (d) Relation between scores and understanding. Do the performance scores reflect level of student understanding, or can students get the correct answer with very little understanding of the underlying concepts? Questions such as these will be used to characterize differential performance levels and kinds of reasoning. The link between performance score and level and kind of reasoning and understanding can then be made.

Future Plans

Analyses of a few tasks are not sufficient to build a theory of cognitive performance that can inform assessment design. Rather, in-depth studies of many different tasks need to be undertaken to adequately characterize the knowledge structures and processes engaged by current assessment practices. During the next year, other suitable tasks will be identified as we begin developing a framework for the construction of performance assessments—a framework that assures a match between the cognitive skills and processes students engage in and those intended by test developers.

For example, contact has been made with the "Pittsburgh's Science Education Through Portfolio: Instruction and Assessment" project which is in the initial stages of designing and evaluating portfolios as a mechanism for assessing students' scientific knowledge (Duschl & Gitomer, 1991). Assessment in this context is viewed as a formative, instructional, and collaborative effort that occurs between student and teacher for the purpose of enhancing instruction (e.g., defining curricular objectives and lesson plans that will facilitate students' understanding of scientific explanations). Evidence of student learning, therefore, needs to be supported by data in the student's portfolio. Currently, work is centering around a sixth-grade instructional unit. This unit is an integrated activity that asks students to design and construct a vessel that can carry a maximum load. The principal learning objective is for students to construct an explanation for why things float and why some objects can carry more weight than others. It is anticipated that we will begin to work with these portfolio assessments in the coming year as this project progresses.

In subsequent years, we anticipate (a) the development of a beginning taxonomy of these processes to guide test design, and (b) descriptions of the ways in which assessment situations can either encompass the objectives of scientific reasoning or indicate how these objectives can be bypassed by situational design and scoring procedures. In general, based on its current work, the project plans to move more deeply into the development of a theory of proficiency in science achievement as it relates to the development of techniques for innovative assessment.

References

ACT. (1989). *The ACT assessment: Test preparation reference manual*. Iowa City, IA: American College Testing Program.

Baron, J.B. (1991, April). *Beyond the promise: Design, data and measurement in new forms of assessment*. Symposium conducted at the annual meeting of the National Council on Measurement in Education and American Educational Research Association, Chicago.

Baxter, G.P. (1991). *Exchangeability of science performance assessments*. Unpublished doctoral dissertation, University of California, Santa Barbara.

Blumberg, F., Epstein, M., MacDonald, W., & Mullis, I. (1986). *National Assessment of Educational Progress: A pilot study of higher-order thinking skills assessment techniques in science and mathematics* (Grant No. NIG-G-84-2006-P4). Princeton, NJ: National Assessment of Educational Progress.

California Department of Education (1990). Unpublished documentation of the California Assessment Program 1990 field trials. Sacramento: Author.

Chi, M.T.H., Bassock, M., Lewis, M.W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*, 145-182.

Chi, M.T.H., Glaser, R., & Farr, M. (Eds.). (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.

Chi, M.T.H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 7-70). Hillsdale, NJ: Erlbaum.

Duschl, R.A., & Gitomer, D.H. (1991). Epistemological perspectives on conceptual change: Implications for educational practice. *Journal of Research in Science Teaching*, *28*(9), 839-858.

Ericsson, A.K., & Simon, H.A. (1984). *Protocol analysis. Verbal reports as data*. Cambridge, MA: MIT Press.

Glaser, R. (1990, October). *Testing and assessment: O tempora! O mores!* Paper presented at the Thirty-first Horace Mann Lecture at the University of Pittsburgh, PA.

Glaser, R. (1992). Expert knowledge and processes of thinking. In D.F. Halpern (Ed.), *Enhancing thinking skills in the sciences and mathematics* (pp. 63-75). Hillsdale, NJ: Erlbaum.

Lomask, M., Baron, J., Greigh, J., & Harrison, C. (1992, March). *ConnMap: Connecticut's use of concept mapping to assess the structure of students' knowledge of science.* A symposium presented at the annual meeting of the National Association of Research in Science Teaching, Cambridge.

National Academy of Education (NAE). (1987). *Learning by doing: A manual for teaching and assessing higher-order thinking in science and mathematics.* Princeton, NJ: Educational Testing Service.

The new CAP assessments. (1991). *California Curriculum News Report*, *16*(3) 6-7.

Novak, J.D., & Gowin, D.B. (1984). *Learning how to learn.* New York: Cambridge University Press.

Shavelson, R.J., Baxter, G.P., Pine, J., & Yurè, J. (1990, April). *Alternative technologies for assessing science understanding*. Symposium presented at the annual meeting of the American Educational Research Association, Chicago.