# Using Computer Simulations to Assesss Hands-On Science Learning[1]

## Gail P. Baxter [2,3]

Two methods of assessing student learning of a hands-on instructional unit are compared. One method involves manipulation of concrete materials, and the other method involves manipulation of icons on a computer to solve an electric circuits problem. Sixth-grade students in an inquiry-based science program completed both assessments separated by three weeks. Results indicated that although mean performance was the same for each method, individual student performance varied considerably. Implications for use of computer simulations as an alternative to hands-on assessment are discussed.

KEY WORDS: Science assessment; performance assessment; computer assessment.

## INTRODUCTION

Computers offer a unique opportunity to link assessment with teaching and learning because they can provide an open-ended, unconstrained environment for conducting science investigations (Pine *et al.*, 1991). Although development costs are considerable, computer assessments offer a number of advantages over hands-on assessments (e.g., Helgeson and Kumar, 1993; Shavelson and Baxter, 1992). First, simulations are less costly and time consuming to administer than hands-on assessments, in part because large numbers of students can be tested simultaneously if sufficient computers are available. Second, computers maintain a full record of performance, so teachers and students can review problem-solving processes. Third, simulations offer the opportunity to teach and assess areas that are impractical to do hands-on such as genetics experiments or model-based reasoning tasks. Fourth, students experiment with the technology, discovering solutions to problems that they might not have found with other types of assessments or instructional activities.

In this paper, the use of computer simulations as an alternative to hands-on assessment is discussed.[4] An Electric Mysteries assessment developed by a team of researchers, scientists, and teachers is the focus of the discussion (for details, see Pine *et al.*, 1991; Shavelson *et al.*, 1991). Two methods of assessment were developed, one using concrete materials and the other a computer simulation. The former was developed under the assumption that the optimal way to evaluate hands-on learning is with hands-on assessment. In other words, the ideal assessment would provide students with equipment, pose a problem, and have an observer evaluate students' performances in real time. The latter (computer simulation) was developed in an attempt to find cost- and time-efficient alternatives to direct observation.

If simulations are used to assess hands-on learning, it becomes important to know to what extent student performance on a simulated investigation

[2]University of Michigan, School of Education, 610 East University, Ann Arbor, Michigan 48109.
[3]Correspondence should be directed to Gail P. Baxter, University of Michigan, School of Education, 610 East University, Ann Arbor, Michigan 48109.

Fig. 1. Electric Mysteries Part I (adapted from Shavelson *et al*, 1991).

matches that for the same investigation done with concrete materials. Any discrepancy in performance would call into question the appropriateness of using simulations as a viable alternative to direct observation of student performance. The paper begins with a description of the Electric Mysteries assessments. Next, a comparison of student performance with each method is presented. The paper concludes with a discussion of the strengths and limitations of using computer simulations for assessing hands-on learning.

## HANDS-ON ASSESSMENT

For this method of the assessment, students are presented with equipment and asked to conduct an investigation to determine the contents of six "mys-

tery" boxes. Students cannot see inside the boxes, nor does light shine through the boxes. The task has two parts (see Fig. 1). Part I serves to orient students to the equipment (batteries, bulbs, wires with clip leads, and mystery boxes) and the task. To this end, students are asked to create a circuit and draw a picture of it. Then, students are asked to determine whether a battery or a wire is in a box with a question mark on it. They are prompted to connect the question mark box in a circuit with a bulb to determine what is inside.

Part II of the task asks students to determine the contents of each of six mystery boxes A–F from a list of five possible alternatives (see Fig. 2). Students are presented with two batteries, two bulbs, and five wires to connect circuits to the six boxes. Two of the boxes have the same thing (box B and F

Find out what is in the six Mystery boxes A, B, C, D, E, and F. They have five different things inside, shown below. Two of the boxes have the same thing. All of the others have something different inside.

**Two Batteries**

**Wire**

**Bulb**

**Battery and Bulb**

**Nothing at all**

For each box, connect it in a circuit to help you figure out what is inside. You can use your bulbs, batteries, and wires any way you like.
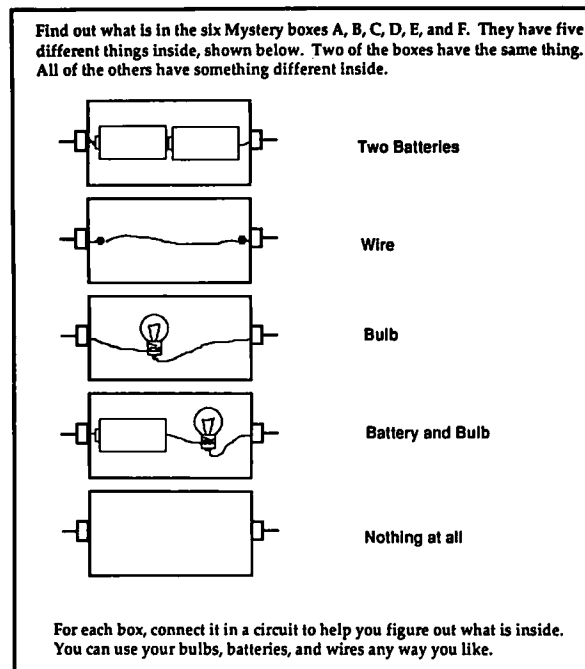
Fig. 2. Electric Mysteries Part II (adapted from Shavelson *et al.*, 1991).

each have a wire). All of the others have something different. The boxes are sequenced such that the first and fifth boxes can be solved by testing with a bulb in the circuit and the other boxes require a battery and bulb.

## COMPUTER SIMULATION

The Electric Mysteries computer simulation was developed in a Macintosh environment so as to replicate as nearly as possible the hands-on investigation (Fig. 3). The screen was divided into three sections: equipment pallet, work space, and control panel. To begin the simulation, there is a brief tutorial that illustrates how to drag simulated batteries, bulbs, and boxes from the pallet at left and connect them with wires. Wires are drawn by clicking at one terminal of a circuit element and dragging to another terminal. Wires are removed by double-clicking on them. Students could connect a multitude of circuits on the screen at once. Alternatively, they could leave one completed circuit on the screen for comparative purposes. In the tutorial, the student also learns how to save the screen as a "notebook page," with written

comments along with the circuit picture; to get a new blank page; and to go back to a previous page and edit it (using the buttons in the control panel at the right of Fig. 3).

Proceeding exactly as in the hands-on investigation, guided by similar written material, the student is first asked to light a bulb. Next, to establish the idea of using an external circuit to find out what is inside a closed box, the student does the "question mark" box. The result is written in the "notebook," and the circuit and the written answer are saved.

Finally, the student is shown on paper the five possible contents of the six boxes, as in the hands-on investigation, and asked to find out what is in each one. Box A can light a bulb and responds to the same strategy as the student learned to use with the question mark box. Box B is a box that will not light a bulb, and a new external test circuit needs to be used. The software emulates the behavior of a real circuit. For example, a bulb connected to a box that contains a battery and a bulb would appear dimmer than a bulb connected to a battery. A bulb connected to a box that contains two batteries would appear brighter than a bulb connected to one battery.
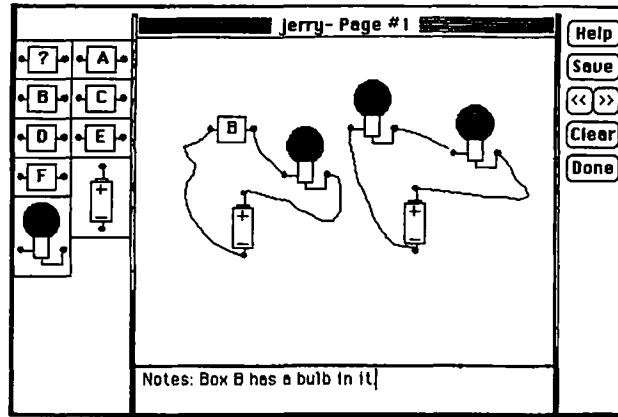
Fig. 3. Electric circuits computer simulation.



Fig. 4. Scoring system for Electric Mysteries (adapted from Shavelson *et al.*, 1992).

## SCORING

Student performance is scored on the basis of: (1) identification of the contents of each box and (2) the sequence of circuits used to arrive at the answer (Fig. 4). If the student correctly identifies the contents of a box and uses a legitimate circuit to test the box, he/she is credited with one point. If the stu-dent fails to correctly identify the contents and/or fails to demonstrate the use of an adequate circuit or sequence of circuits, no credit is assigned. The maximum possible score is 6 (one for each box). The initial part of the task (build a circuit and determine what is in the question mark box) are not scored, as these questions serve to orient the student to the task.
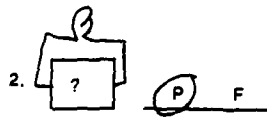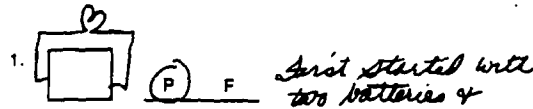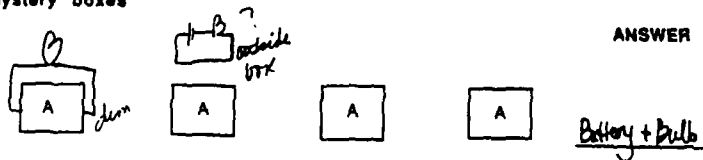
Fig. 5. Observer score form for hands-on investigation.

Hands-on performance was scored on the basis of rater observations. While the student conducted the investigation, the observer recorded the sequence of circuits the students used to test each box (Fig. 5).

For the simulation, when the student is finished and signs off, the computer writes a file in which his or her every move is saved for later replay with the evaluation program. First, the evaluation presents an overview of all the saved pages. Then, starting at a chosen page, the viewer can play back each move or just look at the final circuit of each page (Fig. 6). Performance is graded by observing each move in succession using the horizontal scroll bar. The vertical scroll bar scans pages, and a movie replay is also available. The viewer can add comments in the notes section, which are saved and could provide feedback to the student. Any screen can be printed, or a summary of all the saved pages.

## COMPARISON OF HANDS-ON AND COMPUTER SIMULATION

The computer simulation appears, at least on the surface, to be comparable to the hands-on investigation because each assessment method engages students in a cyclical process of hypothesis testing and refining to identify the circuit components enclosed in each of six boxes. Using their knowledge of what constitutes a circuit, and the impact of changing various components in a circuit (e.g., adding a second bulb), students test out their hypothesis by manipulating equipment (lab equipment or icons on the computer) and interpreting their observations. For example, if the bulb is dim when connected in a circuit to one of the boxes, students might reason that there is a battery and a bulb in the box. If the bulb is very bright, students might reason that there
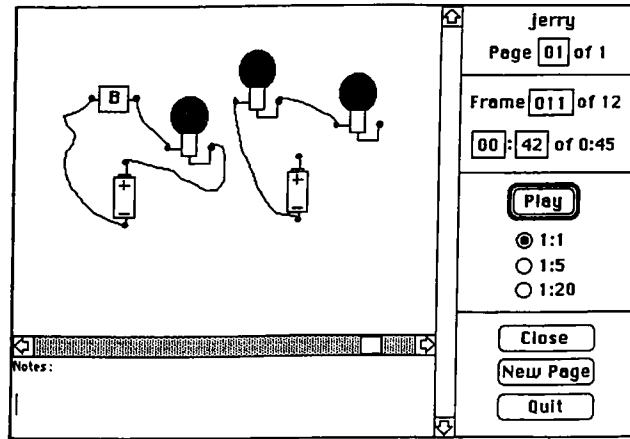
Fig. 6. Computer evaluation program.

are two batteries in the box. A fundamental question is: to what extent does student performance on the simulated investigation match that for the same investigation done with concrete materials?

To address this question, approximately 100 sixth-grade students enrolled in an inquiry science program completed both methods of assessment. For the hands-on investigation, students were tested individually by a trained observer who noted the sequence of circuits students constructed to test each box and the conclusions they made based on those tests (e.g., box B has a wire in it). Approximately three weeks later, students were tested on the simulation. The testing was done using four Macintosh computers in a single room, with two supervisors whose role was to schedule students, help them through the tutorial, and reboot the computers after crashes. Approximately 20–30 min were required for students to complete each method, although time limits were not imposed.

For the hands-on and computer simulation assessment methods mean performance across students, difficulty of each box, and individual student scores were compared. Results indicate that mean performance was approximately equal for both methods (3.76 and 4.05, for hands-on and computer, respectively). The score distributions were very similar, varying in only two respects. None of the students scored zero on the computer and almost twice as many students scored four on the computer ($N$ = 25) as on the hands-on ($N$ = 14) assessment.

Next, the difficulty of each box was examined. Recall that the ordering of the boxes varied by as-

sessment method. Nevertheless, for both methods the first and fourth boxes contained a battery. The second, third, and fifth were different nonbattery boxes, while the sixth in both cases was a second box with a wire in it. The intent was to keep the difficulty about the same for both orderings. Comparison of proportions of students answering correctly indicate that difficulty did not vary with assessment method. The box containing two batteries was always the easiest and the boxes containing a wire or a bulb were the most difficult.

Finally, individual student scores on the hands-on and simulated versions were compared. The differences were surprisingly large. At each score level for the hands-on assessment, every possible score was obtained on the computer simulation. In other words, students who scored 6 with the hands-on assessment, received scores of 1 through 6 for the computer simulation. The lack of relationship is reflected in the low correlation (0.35). Some students scored higher on the hands-on than on the computer; the opposite was true for other students. It is clear that for this sample of students the simulation and the hands-on methods provide different information about individual students' knowledge of circuits. What could account for this discrepancy?

To gain some insight into the performance discrepancy, a qualitative analysis of six students' performances on each method was carried out. Two of the students scored three points higher with hands on (6, 5) than computer (3, 2); two of the students scored three points higher with the computer (6, 4) than with hands-on (3, 1). Results indicate that all

six students displayed a good strategy for the method that resulted in the highest score. Typically, they tested each box with a bulb first and then, if necessary, a battery and bulb. They correctly interpreted the information from their tests and operated within the constraints of the task.

In contrast, students' performances with the method that resulted in the lowest score are best characterized by a combination of inadequate strategy (e.g., used battery and bulb to test every box), misinterpreted information (e.g., bulb dim therefore a wire in the box), and/or failure to monitor performance (e.g., did not compare conclusions to list of possible options). For example, students may have reported a wire three times. Twice the bulb was lit normally (correct interpretation) and once the bulb was lit dimly (incorrect interpretation). Had the student operated within the constraints of the task (two of the boxes have the same thing, all of the others have something different) this error could have been corrected.

In summary then, the computer simulation provides the same information as hands-on assessments averaging across all students. For individual students, there are large fluctuations in performance. Some students perform better with one method and some students perform better with the other. The simulations, then, are not strictly equivalent to hands-on investigations for individual students.

## IMPLICATIONS OF COMPUTER SIMULATIONS FOR ASSESSMENT

Three factors set the stage for research and development of computer simulations for use in elementary science classrooms: the increased availability of computers in the classroom, the increased use of hands-on instructional techniques for science teaching, and the demand for alternative, cost-efficient assessments symmetric with instructional practice. The research discussed here suggests a significant potential for computer simulations as a form of performance assessment given the desirable characteristics of this technology. The simulation offers a time- and cost-efficient alternative to direct observation of student performance. Materials do not need to be refurbished (e.g., batteries), formative feedback can be provided to the student via comments entered on the evaluation screen, and scoring is quick and easy. Perhaps most important is that the format is inherently motivating to students who feel quite free to experiment with the technology, posing questions and looking for solutions. The simulation then serves as an instructional tool and as an assessment tool (Shavelson and Baxter, 1992).

Heretofore, computer simulations have not received the attention they warrant given their unique potential to link testing with teaching and learning. The research reported here, and elsewhere (e.g., Helgeson and Kumar, 1993; Pine *et al.*, 1991) suggests the need for considerable empirical work beginning with the development and evaluation of appropriate simulations for teachers to use. Extensive development efforts permit an examination of the strengths and limitations of computer simulations for instruction and assessment. An understanding of the relative effectiveness of a diverse range of simulations in a variety of instructional contexts provides a basis for wide-scale implementation and use that maximizes the potential of simulations as an educational tool.

## REFERENCES

Baxter, G. P., and Shavelson, R. J. (1994). Performance assessments: Benchmarks and surrogates. *International Journal of Educational Research* 21(4): 279–298.

Helgeson, S. L., and Kumar, D. D. (1993). A review of educational technology in science assessment. *Journal of Computers in Mathematics and Science Teaching* 12(3/4): 227–243.

Pine, J., Baxter, G. P., and Shavelson, R. J. (1991). *Computer Simulations for Assessment.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

Shavelson, R. J., and Baxter, G. P. (1992). What we've learned about assessing hands-on science. *Educational Leadership* 49(8): 20–25.

Shavelson, R. J., Baxter, G. P., and Pine, J. (1991). Performance assessments in science. *Applied Measurement in Education* 4(4): 347–362.

Shavelson, R. J., Baxter, G. P., and Pine, J. (1992). Performance assessments: Political rhetoric and measurement reality. *Educational Researcher* 21(4): 22–27.