Benchmarking Science Education Software:
LESS THAN MEETS THE EYE

Abha Shrivastava

INTRODUCTION
Many countries are expanding the use of computers in schools in an effort to improve instruction. The hopes for benefits from computers for the teaching of science are perhaps greatest where a shortage of well-qualified teachers is often greatest and where the need for laboratories and equipment make the subject expensive to teach.

Instructional technology has evolved from simple CAI (computer-aided instruction) to interactive multimedia CD-ROMs. Currently, there are more than 10,000 English language software products intended for instructional or educational use with microcomputers in schools and in homes. Undoubtedly there are thousands more titles in other languages.

How good is this software? Does it cover well the most important content? Does it use pedagogical strategies that are likely to enhance real understanding? And do students and teachers easily use it? This article will briefly summarize the results of a study that examined how well the allegedly best English language science education software measures up to the national standards for the teaching of science as specified by the American Association for the Advancement of Science [AAAS].

Deciding which curricular materials to use is one of the most important professional judgments that educators make. In nations around the world, Ministries of Education, State departments, and local district advisory committees review and recommend textbooks. Their decisions influence instruction for years to come.

Some science education journals have published reviews of individual software packages and some science education resource centers have provided similar information. These reviews, however, have rarely judged the software by nationally established standards for science education.

THE AAAS BENCHMARKS
The American Association for the Advancement of Science (AAAS) is the premier association of American scientists. Its Science for All Americans indicates what all students should know and be able to do in science, mathematics, and technology by the time they graduate from high school, and its Benchmarks for Science Literacy specifies learning goals and desirable pedagogy for grades K–12. These documents are the result of years of research by a vast team of scientists and science educators.

The Organization of Economic Cooperation and Development (OECD) has these documents as key to "single most visible attempt at science education reform in American history" (OECD, 1996 as cited in AAAS, n.d.). AAAS’s science education efforts have attracted the interest of several other countries. For instance, fifteen mentor teachers from Panama attended a professional development workshop hosted by AAAS, as the first step in an ongoing relationship between Panama and the Association (AAAS, 1999).

The AAAS Benchmarks indicate that a coherent set of well-understood facts and concepts provides a solid base for further learning. It urges that science instruction focus be limited in number of facts and concepts that are of lasting significance. Quality, not quantity, is the priority. The premises of the AAAS science education reform efforts are as follows:

• AAAS promotes literacy in science, mathematics, and technology in order to help people live interesting, responsible, and productive lives. In a culture increasingly pervaded by science, mathematics, and technology, science literacy requires understandings and habits of mind that enable citizens to grasp what those enterprises are up to, to make some sense of how the natural and designed worlds work, to think critically and independently, to recognize and weigh alternative explanations of events and design trade-offs, and to deal sensibly with
problems that involve evidence, numbers, patterns, logical arguments, and uncertainties.

- Curriculum reform should be shaped by our vision of the lasting knowledge and skills we want students to acquire by the time they become adults. This ought to include both a common core of learning—the focus of Benchmarks—and learning that addresses the particular needs and interests of individual students.

- If we want students to learn science, mathematics, and technology well, we must radically reduce the sheer amount of material now being covered. The overstuffed curriculum places a premium on the ability to commit terms, algorithms, and generalizations to short-term memory and impedes the acquisition of understanding. Goals should be stated to reveal the intended character and sophistication of learning to be sought. Although goals for knowing and doing can be described separately, they should be learned together in many different contexts so that they can be used together in life outside of school. (AAAS, 1993, pp. xi-xii)

The Benchmarks are intended help educators decide what to include and exclude from a core curriculum, when to teach it, and why. The Benchmarks are divided into the following sections: the nature of science, the nature of mathematics, the nature of technology, the physical setting, the living environment, the human organism, human society, the designed world, the mathematical world, historical perspectives, common themes, and habits of mind. Within each section are separate benchmarks for grades K-2, 3-5, 6-8 and 9-12.

The sequence of benchmarks for any given topic reflects a logical progression of ideas, with benchmarks for earlier grades anticipating the more advanced benchmarks for later grades (Nelson, 1998). For instance, the benchmarks for “The Physical Setting” in the elementary grades (K-2, 3-5), include “the earth is one of several planets that orbit the sun, and the moon orbits the earth” and “… the rotation of the earth on its axis every 24 hours produces the night-and-day cycle…” Then the benchmarks for the middle grades (6-8) indicate that “because the earth turns daily on an axis that is tilted relative to the plane of the earth’s yearly orbit around the sun, sunlight falls more intensely on different parts of the earth…” Finally, the benchmarks for grades 9-12 include “weather and climate involve the transfer of energy in and out of the atmosphere…”

Many nations specify the educational content to be taught in the schools, in all discipline areas. The AAAS has gone beyond that and also defined seven criteria for the quality of instruction. These have been derived from research on learning and teaching and on the craft knowledge of experienced educators.

THE EVALUATION OF SCIENCE EDUCATION SOFTWARE

AAAS had previously developed a multi-step procedure for judging textbooks against its Benchmarks. That procedure, with a few small modifications, was applied to ten widely used science instruction CD-ROMs that were chosen from a sampling frame comprising 17 bestsellers and 30 highly recommended software packages. Two investigators worked independently applying the procedure to the same software, and then met to reconcile the few differences in their code.

Details on the study procedures and the adequacy of the content, pedagogy and usability of each of the ten software packages analyzed in this study are available in another source (Shrivastava, 2002). The general findings are summarized below:

Content Findings

The entire sample of ten software packages addressed only 53 of the approximately 270 benchmarks for the targeted grade level. On an average, each software title addressed 6.7 benchmarks. It is not desirable for a given software package to address many benchmarks, because that could require doing so superficially. However, the limited number of benchmarks covered by the sample of 10 software packages suggests that educators may have difficulty finding software that addresses some important science topics.

On an average, only 25 percent of the software content was found to align with the grade-appropriate AAAS benchmarks. The alignment for each of the ten software titles ranged from 0-65 percent. Additionally, five out of ten CD-ROMs were found to contain some benchmark-related content (12-45 percent) pertaining to lower grades than the ones mentioned in the titles. This suggests that widely used science education software is focusing largely on content that is considered unimportant by AAAS.

Even when the software did address a benchmark, it often covered only some of the important ideas within the benchmark. The average coverage of a given bench was rated 1.8 on a four point scale where 1 = minimal coverage, 2 = medium-well, 3 = moderately well and 4 = full coverage.

Pedagogy

The ratings of various pedagogical criteria varied considerably. High ratings were achieved by two of the criteria, “engaging students with relevant phenomena” and “developing and using scientific ideas.” Those ratings were 2.1 and 2.8 respectively on a four-point scale with 0 = none, 1 = poor, 2
educational reform (Galbreath, 1992; Jost & Schneberger, 1994), and indeed educational software has become a big and still growing industry.

The results of this study indicate that current widely available English-language science education CD-ROMs are by no means adequate as the primary source of science education. The software falls way short in its coverage of priority knowledge and skills, and it has some serious pedagogical shortcomings. Other evidence has also suggested that this software is not a substitute for conventional curricula (Fennimore, 1997; Harris, 1998).

This software, however, can be useful as a supplement to good textbooks and teachers. It is particularly adept at engaging students in science and helping them to develop and use scientific ideas.

The results of this study strongly suggest that prior means of identifying good science education software have been inadequate. Five of the CD-ROMs analyzed in this study were drawn from a small number that have been highly recommended, and the other five were drawn from among best sellers. Despite that, all had serious shortcomings in content coverage and pedagogy.

Education planners and administrators could make big mistakes in purchasing software if they rely on traditional reviews and market popularity. They would be much better served having some organization scrutinizing the software using procedures similar to those used in this study (Shrivastava, 2002). If their curriculum guides align well with the AAAS Benchmarks, the procedures could be used without modification. If their curriculum guides do not align well with the Benchmarks, then the procedures should be modified to address the priority content of the guides. A university group, an NGO, or a private contractor might conduct the reviews. Once prepared, the reviews could be shared widely with parents, libraries, and others who make independent decisions about software purchases.

Policy makers may also wish to consider spurring software developers to produce better science education software. That could be done in several ways. If a country or state were to publicize that future purchases of software would be based on a rigorous review of the content and pedagogy, and if they provided the developers with the evaluation criteria, the software is likely to improve. National organizations might bring together software developers, science educators, scientists, and policy makers for an exchange of ideas and to forge common understandings, much as has been done in the past with textbook publishers (AAAS, 2002).

Usability

The CD-ROMs were relatively problem-free in installation and running. The CD-ROMs rated “satisfactory” to “high” on interface and on creativity, reflecting the technological advances that have made software much more user-friendly and imaginative in the use of color and animation. They followed a logical and sequential format. Most of the CD-ROMs, however, did not allow students to save their work or to modify parameters for individual needs.

Implications

Use of computers in schools and homes has become commonplace in many industrialized countries and several developing countries have made large investments in computers for public schools. Many educational professionals have believed that multimedia software would play a large part in educational reform (Galbreath, 1992; Jost & Schneberger, 1994), and indeed educational software has become a big and still growing industry.

Satisfactory ratings were achieved by the criteria “introducing terms meaningfully.” This indicates that most of the software fairly often introduced technical terms in conjunction with relevant experience, rather than just having students memorize definitions of terms.

Low ratings were earned by the four other pedagogical criteria: “attending to prerequisite knowledge and skills,” “promoting student thinking about phenomena,” “assessing progress,” and “enhancing the learning environment.” Scores for all ten software packages indicate that none strove to help teachers or students identify prerequisite knowledge and skills needed by students to learn the content being addressed. Likewise, there was little scope for students to refine their understanding and interpretation—to do some thinking and wondering about the science content with which they just dealt. There was also very little assessment (or none, in some CD-ROMs) to test and gauge whether the students achieved the benchmark ideas. There was a severe lack of appropriate and adequate problems and exercises, either at the end of the lesson or embedded within the lesson.

Usability

The CD-ROMs were relatively problem-free in installation and running. The CD-ROMs rated “satisfactory” to “high” on interface and on creativity, reflecting the technological advances that have made software much more user-friendly and imaginative in the use of color and animation. They followed a logical and sequential format. Most of the CD-ROMs, however, did not allow students to save their work or to modify parameters for individual needs.
REFERENCES


---

Abha Shrivastava recently completed her doctoral studies in Education Policy at The George Washington University. She can be reached at abha@gwu.edu