

---

HENRY KELLY

# Games, Cookies, and the Future of Education

---

*Games, simulations, user models, and other information tools have revolutionized and personalized entertainment and services. What about education?*

---

In February 1990, President George H. W. Bush joined Governor Bill Clinton of Arkansas to embrace the goal that by the year 2000, “U.S. students will be first in the world in mathematics and science achievement.” The two leaders were reacting, in part, to a study stating that education in the United States was so bad that it would have been considered an “act of war” had it been imposed by a foreign power. The ensuing decade was marked by the release of a series of major studies by government or business highlighting the problem, with ever sharper rhetoric. Yet when the year 2000 rolled around, U.S. students ranked 22nd among 27 industrialized countries in math skills, according to a widely regarded international comparison. In 2003, a similar study ranked U.S. students 24th of 29 countries.

Today, even in the contentious atmosphere of political Washington, there is near-universal agreement that this situation must be remedied. No one doubts that a world-class U.S. workforce, skilled in math, science, and technology, is needed to maintain

or improve the competitiveness of U.S. companies, ensure national security, and meet critical needs in health care, energy, and the environment. There also is growing concern that U.S. wages and living standards are at risk as companies and investors must choose between training underprepared U.S. employees and finding a way to do the job using better-trained employees in other countries.

The dilemma, of course, is that consensus about the magnitude of the problem has not translated into agreement on what to do. Holding students and school systems to high standards is necessary, as called for by the federal No Child Left Behind Act, but there is widespread concern that this alone is not sufficient. It also will be necessary to understand better how students learn and to design and implement new tools to take advantage of this understanding. This job is doable, especially with the help of advanced information technologies. But meeting the challenge will take concerted efforts that begin at the national level and extend into state and local governments, school systems, and businesses.

## **Affordable solutions**

Recent advances in learning and cognitive research

---

Henry Kelly (hkelly@fas.org) is president of the Federation of American Scientists in Washington, D.C.

provide a solid basis for believing that real progress is possible in improving learning outcomes for anyone studying any subject. Among the recommendations for improving education are the use of individualized instruction, subject-matter experts, and rich curricular activities. But few of the proposals have been widely adopted, and in many quarters they seem hopelessly unaffordable within our traditional approach to teaching.

Advanced information technologies offer real hope that many of the recommendations can be implemented without an unrealistic increase in spending. We are all familiar with the unexpected ways in which information technology has improved our lives in other areas— instant messaging, sophisticated software that helps firms personalize online shopping, efficient systems for answering consumer questions, eye-popping simulations on inexpensive computer game consoles. These tools have the potential to reshape learning through interactive simulations, “question management” systems that combine automated and human responses, and powerful continuous assessments. Computer simulations could let learners tinker with chemical reactions in living cells, practice operating or repairing expensive equipment, or experiment with marketing techniques, making it easier to grasp complex concepts and transfer this understanding quickly to practical problems. New communication tools could enable learners to collaborate in complex projects and ask for help from teachers and experts from around the world. Learning systems could adapt to differences in student interests, backgrounds, learning styles, and aptitudes.

Despite huge investments in communications and computer hardware made by universities, schools, and training institutions, most formal teaching and learning still use methods familiar in the 19th century: reading texts, listening to lectures, and participating in infrequent—and usually highly scripted—laboratory experiences. The cookies on children’s computers might know more about what they like and do not like than do their teachers.

Given current conditions, it will take a significant and sustained investment in research to invent and test new approaches to learning. It took years of experiment and failure for other service businesses and the entertainment industry to find ways to improve the quality of their services and increase their

productivity through the effective use of information technology. The gains required decades of research, an unforgiving review of cherished management approaches, and a dramatic redefinition of many jobs. Education can benefit from what these companies learned if it is willing to undertake a serious process of research, evaluation, and redesign.

It is difficult to see how the kind of patient, long-term research and evaluation needed to bring these concepts alive can be done without a new, aggressive, large-scale federal program for research, development, demonstration, and testing. Once proven, of course, research results can be translated rapidly by individuals and companies into commercial products that can be used across the country by instructional institutions with innovative leaders. Although the implementation will be a bottom-up process, research should be conducted at the national level. A small fraction of total federal investment in education and training devoted to research, design, and development in this area would pay huge dividends. The research should not face political barriers, if only because it can design tools that would give more power to the nation’s diverse education and training institutions, enabling them to tailor instructional systems to unique local needs.

The federal research effort should be guided by an effective management plan that includes a clear definition of goals and ways to measure progress toward them. Progress can be measured in four different ways, based on the extent to which a new tool or approach 1) increases the speed at which expertise is acquired and depth of understanding achieved; 2) increases a learner’s ability to transfer expertise acquired to the solution of practical tasks; 3) decreases the range of outcomes among learners; and 4) makes learning more motivating (and more fun), if only to get more time on task.

The technology alone obviously is not sufficient to meet the goals and cannot substitute for talented teachers and experts. But taken together with skillful use of human instructors, the technology can do two dramatically new things. First, it can provide accurate, compelling simulations of physical phenomena and virtual environments for exploration and discovery. These can be used to illustrate complex concepts through the ancient art of talking and showing and can be used to build challenging assignments and games.

Second, it can combine artificial intelligence techniques and rapid connections to real experts that together can reproduce many of the benefits of one-on-one tutoring.

The essential first step is to define key research challenges and organize the research on learning technology into manageable components. The division must be somewhat arbitrary, because the pieces are obviously dependent, but some form of structure is essential to expand the research efforts beyond the current cottage-industry approach. An effective research structure also will link learning research with other information technology research working on similar problems.

The Learning Federation Learning Science and Technology R&D road map provides a well-defined structure for organizing the R&D around core research challenges. The road map was developed over a 3-year period using the methods pioneered by the SEMATECH Corporation, which built and revised a research plan that helped guide the revival of U.S. semiconductor manufacturing. More than 70 leading researchers from industry, academia, and government helped develop this road map through their participation in focused workshops, interviews, and preparation of technical plans. The road map organizes research into the following four topic areas: new approaches to teaching and learning enabled by new technologies, peer-reviewed simulations and virtual environments, systems that will make it easy for students to pose questions and receive answers, and assessment.

### **Improving teaching and learning**

It may seem obvious, but one of the most important lessons learned by commercial service companies trying to make effective use of information technology is that the place to start is not with the question “what can computers do?” but “what do we really want to accomplish?” It is essential to begin by understanding how learning can best be achieved and then ask technologists how much of the ideal can be achieved at an acceptable price. Fortunately, a now-classic 1999 report from the National Research Council (NRC), *How People Learn*, provided a superb answer to the “what

**Compared with the  
need, current  
federal research  
funding for learning  
technology is small  
and fragmented.**

do we want to accomplish” question in a comprehensive review of what is known to work in improving learning.

The report argued for approaches that give the learner lots of practical experience and opportunities to apply facts and theories in practical situations. It also cited the need to continue efforts as long as they remain challenging and reinforce expertise. Not surprisingly, some of the most powerful learning strategies also are the most ancient:

struggling to accomplish a difficult but highly motivating task that requires new knowledge; carefully scanning a complex, changing environment; and seeking individualized help from experts and friends.

At the time of the report, some critics worried that it would not be feasible to provide large numbers of students with the kinds of experiences and challenges suggested or to monitor each student to find out whether he or she was prepared for the next level of complexity. But the recent spectacular success of computer games provides a tantalizing example of what might now be accomplished. Well-designed, highly interactive simulations can provide a wide range of experiences, such as navigating difficult terrain, operating complex vehicles, and collaborating with colleagues to overcome obstacles. They have an almost frightening ability to capture and hold interest. Gamers will spend literally hundreds of hours mastering obscure details of new weapons systems in order to meet the motivating goals established by the artifice of the games.

Obviously, significant research is needed to find out how best to achieve the goals of *How People Learn* using new technologies. But where investments have been made, the results have been impressive. Training experts in the U.S. Department of Defense (DOD) are convinced that expertise gained through the use of flight simulators and large-scale military computer simulations has a high rate of transfer to practical skills in the field. They point to changes in the shape of the learning curve: phenomena well documented in training fighter pilots, surgeons, and algebra students. Novices make many more mistakes when they encounter their first prac-

tical applications than they do after a few dozen real experiences. Something in the experience reshapes the formal information and begins to approach real expertise. In many cases, simulated experiences can have much the same learning impact as real experience. Research can tell us when this is true and when it is not.

Even if we fully understand how best to use simulated environments, the challenge of actually building technically accurate and visually compelling simulated environments is enormous. Ideally, someone with a compelling idea for creating a simulation that required mastering a new set of concepts in biology, to choose one example, could draw on libraries of pre-tested software, such as simulations of biological systems, vehicles, and landscapes, to implement their ideas, instead of being forced to design their own new software. The task of building such a library and providing the needed peer review and updates is clearly enormous and must be the work of many hands. But for the system to be useful, the components built by different groups must be reusable and able to work together or interoperate; my simulated knee bone must recognize your simulated thigh bone. Most major computer-assisted design formats are interoperable, meaning that General Motors can build an engine design from software elements provided by the vendors supplying cylinder heads and fuel injectors. However, full interoperability of simulations built from robust peer-reviewed software is still an elusive goal in all domains.

The explosion of “in silico” experiments now under way in almost every branch of science should be a gold mine for developers of educational simulations. Several major federal research funding organizations have taken a critical first step by recognizing that a more systematic approach to software engineering is essential. The complexity and importance of software for academic research have outstripped the tradition of having self-taught graduate students build code with little thought to documentation, reusability, or interoperability with code developed by other organizations. The Defense Advanced Research Projects Agency (DARPA) has taken an early lead in this area with projects that are building a community of practice in software written by different teams for biological simulations. The National Institutes of Health’s (NIH’s) new National In-

stitute of Biomedical Imaging and Bioengineering also has begun to encourage the interoperability and development of new tools for peer review, error reporting, and managing intellectual property. This “digital human” movement can give rise to software models that can be used as the basis of powerful, accurate, and up-to-date instruction in biology. But the specialized research software will not move automatically into learning environments, if only because neither DARPA nor NIH has a mission in education.

### **Building better tutors**

Studies comparing individual tutoring with classroom instruction suggest that tutoring has spectacular impact. In a landmark series of studies, Benjamin S. Bloom and colleagues demonstrated that one-on-one tutoring improved student achievement by 2 standard deviations over group instruction. This is roughly equivalent to raising the achievement of the 50th-percentile students to the 98th-percentile level. In addition, the study found that the range of outcomes (the gap separating the best and worst students) was greatly reduced.

A combination of well-designed computer systems and careful use of human resources can approach the impact of good tutoring. Consider the success of two educational tools, Algebra Tutor and Geometry Tutor, designed by the company Carnegie Learning for use in a traditional teacher-led classroom. The computer tutors, which incorporate nearly two decades of research in artificial tutoring, augment the teacher with tutoring software that adjusts to the individual learner’s competency level. In a number of studies, these tools have produced an improvement of 1 standard deviation over conventional classroom instruction. One interpretation of these results is that the artificial tutor is twice as effective as typical classroom instruction (although it is only about half as effective as the best human tutors).

New instructional systems can have an enormous impact on this problem. First, they can create many situations where learners are highly motivated to ask questions—including deep questions about things they do not understand—instead of being embarrassed. And second, they can provide timely, accurate answers without the need for one tutor per learner, providing the best mix of automated answers and opportunities to talk with teachers and experts. If you keep crashing your airplane in a flight simulator, or if

your patient dies in a simulated surgery, then you are likely to be highly motivated to ask questions about how to improve your performance. Witness the sale of “hint books” for computer games sold to people who spend hours boning up on expertise valuable only for meeting the artificial objectives of the game.

Businesses and defense agencies have a large amount of work under way in the area of question management. The intelligence community, for example, has mounted a major research effort to create effective question-answering technologies. Existing systems have not fully succeeded, but the progress is significant. Search engines are becoming increasingly sophisticated and are by default the principal question-management tools for students. Business help desks provide as much automated advice as possible and connect clients to human experts only when the automated system is inadequate. In the case of systems developed for learning, of course, the answer should reflect the instructor’s pedagogical strategy. In many cases, the best answer is another question—instead of Ask Jeeves, Ask Socrates.

One important feature of the practical systems that are emerging is that most involve both automated systems and humans. Early “artificial intelligence” projects set out to meet the famous Turing Test: an automated system that is so good in conversation that the user cannot tell that it is not human. This is an interesting, but perhaps unattainable, goal. In the meantime, part of the challenge is designing a system that knows what it cannot do and then links the questioner quickly to the right human being.

### Sharpening assessments

The No Child Left Behind Act has put testing and assessment squarely at the center of national educational policy. There has been overwhelming bipartisan support for the idea that education, like all other endeavors, can succeed only if there are clear ways of measuring quality and holding students, teachers, and school systems accountable. The consensus fragments when it comes to the details on how to do this. In

---

*The evidence available from commercial products suggests that a rich new set of assessment tools is possible.*

---

particular, if the tests are not measuring the right skills and knowledge, then accountability is warped by rewarding the wrong behavior.

Ideally, the learning goals should make sense to the learner, to the instructors, and to an employer or another teacher interested in an accurate measure of the individual’s expertise. A good and timely assessment can be highly motivating. Prospective surgeons, for example, are presumably highly motivated to be able to perform surgeries correctly and are eager to get feedback on how well

they are doing. A series of NRC reports have offered a number of recommendations in this direction. The reports call for assessments that are integrated seamlessly into instruction and provide continuous and unobtrusive feedback. They also call for assessments that focus on complex aspects of expertise, not simply on short-term memory of facts. In such assessments, the learner’s thinking needs to be made visible in ways that can help the learner and the instructor make timely adjustment to the learning process.

But as in the case of so many other recommendations of cognitive scientists, this advice has been difficult to put into practice in general education because of limitations on teachers’ time. Although technology should be able to provide powerful help, examples are not found in education. On the other hand, designers of computer games have intuitively implemented assessment strategies that meet many of the NRC’s recommendations, albeit in highly limited domains. A good game continuously evaluates a player’s skill level, knowing that if players stay at a given level of expertise too long, they will become bored, and if they are allowed to advance too fast, they will become frustrated—both disasters for future sales. A good game keeps the player just at the edge of anxiety.

Many businesses also carry out sophisticated assessments whenever someone goes to their commercial Web sites. Each visitor’s action is carefully evaluated to ensure that the information presented on the screen is helpful and attractive to the individual. The evidence available from commercial products suggests

that a rich new set of assessment tools is possible. These tools can support learning environments that:

*Make adjustments and guide individual learning based on accurate models of what students have mastered (formative assessments) and on other student characteristics (including mood, level of interest, and learning styles) that are revealed by their behavior during the learning.* DOD conducted one of the most ambitious efforts of this type, in a system to help military personnel diagnose and repair the complex hydraulic systems on F-15 aircraft. The system continuously observed the decisions being made during a simulated repair session and used a sophisticated filter to develop theories about what the individual did and did not understand.

*Communicate what each individual has mastered at key milestones in the learning process (summative assessment) in ways that are understandable and credible, to the individual, future instructors (and automated instruction systems), employers, and others.* Sophisticated assessments can provide multidimensional records of the levels of a learner's expertise, including specific examples of how the person has performed in a simulation that illustrates this mastery in a practical way. In principle, records would be maintained in two forms: a public record available to employers and other interested parties, and a private set of records that would provide detailed information about the learner's background, strengths and weaknesses, interests, preferences, and other information needed by an automated or real tutor. These private records are the functional equivalent of personal medical records and should be carefully protected and available only at the learner's discretion.

*Assess the performance of the learning system itself in ways that permit comparison with alternative learning systems and provide information useful to perfecting system components.* Such assessments could tell designers whether learners are spending an unusually long time mastering a particular skill, or whether the instruction generates large numbers of bewildered questions.

### **Managing innovation in schools**

Research to design and test new approaches to learning will be pointless if educational institutions are not willing or able to use them. Education markets are notoriously difficult to enter; they are highly frag-

mented and often highly political. Investors lost considerable amounts of money during the 1990s as companies greatly exaggerated what could be done quickly and underestimated the effort needed to sell learning-technology products to these unique markets. Many companies now have simply abandoned the field, and no private firm is making an investment in learning-technology research that approaches the scale needed for a serious development effort.

Marketing novel products in education is difficult for a number of reasons. Even sophisticated instructional institutions, such as research universities or major government training operations, have no tradition of managing innovation. The new technologies cannot have a significant impact on learning outcomes unless they are accompanied by systematic changes in approach to instruction and new roles for faculty and staff. It is likely that new specialties will be created, such as instructors who spend more time as tutors or members of design teams that build and test simulations than as classroom instructors. But the culture of most learning institutions resists the exploration of such options.

The problem is highlighted by the huge difference between the way in which educational institutions use new technologies and the way in which successful service industries, such as banking and insurance, have adapted to new technologies. The first corporate attempts to use new technology were largely efforts to automate existing work without recognizing that dramatic improvements in the process were possible. During the early 1990s, the economic literature was replete with studies showing that the investment was a wasteful fad—and a lot of it was. But tough competitive pressure forced businesses to rebuild around the new tools and economics, and these companies are now showing substantial productivity gains.

Unfortunately, education seems stuck in the first phase of this process. Massive public investments over the past few years have succeeded largely in providing most students with access to computers and connectivity to the Internet. But for the most part, little has been done to capture the potential of technology. Progress is likely to be slow, because the mechanisms that drive innovation in business simply do not work well in education markets.

New information technologies designed for de-

livering tailored financial services, customer-friendly support, and spectacular computer games have demonstrated clearly that they can have a powerful influence on providing affordable education and training services. Conventional markets have failed to stimulate the research and testing needed to exploit the opportunities in education. Investors are doubly hesitant to enter the notoriously difficult market for educational products and are concerned that they will not be able to appropriate the benefits of basic research. This is a near-classic definition of a problem requiring public investment in research.

### **Federal help**

Compared with the need, current federal research funding for learning technology is small and fragmented. For many years the National Science Foundation (NSF) has managed a small but highly effective set of programs in the field and recently has funded three "science of learning centers." These new centers will study and model behavioral and brain processes; provide test beds for evaluating the use of computerized intelligent tutoring systems; and study neural processes and principles associated with the cognitive, linguistic, and social dimensions of learning. Appropriate to its mission, NSF concentrates its funding on cognitive science and other areas of basic research. No federal agency has a clear mission to support the applied research needed to move from theory to the development, testing, and implementation of innovations in learning technology. DOD has by far the best record in making effective use of learning technology, through support by DARPA and the Army's Institute for Creative Technology for several small but promising applied research programs in learning technology. The Army, Navy, and intelligence services all have ambitious programs specialized for their unique needs. But taken as a whole, the federal research programs are small and poorly coordinated.

Pressured by industry and academic groups, the Department of Education and NSF joined Microsoft, Hewlett Packard, and several major foundations to

---

*The task of creating practical markets for innovations in learning technology will fall primarily on state and corporate program managers.*

---

sponsor the development of a research plan detailing what would be needed to achieve ambitious long-term goals in enhancing education through technology. To build on this effort, the Departments of Education and Commerce held a major summit of corporate and academic leaders to identify ways to strengthen federal education-technology research programs. President Bush's National Science and Technology Council is starting to follow up on these recommendations by conducting a careful inventory of federal research already under way in relevant

areas. But even before the results are in, it is clear that there are major holes in the fabric.

The Digital Opportunity Investment Trust (DO IT) Act (S. 1023, H.R. 2512), introduced with bipartisan support in early 2005, proposes an entirely new approach. The DO IT bill would create an independent federal agency charged with managing an ambitious research program built around the research priorities identified by corporate and academic groups during the past few years. The program's primary focus would be on applied research and on the testing needed to ensure that the innovations actually translate into improved learning. Progress in this regard can be achieved by following the model of "spiral development" that has worked well in other fields of applied research, building pilot applications and thoroughly evaluating them. This effort will require close collaboration with educational institutions and the academic and commercial research community. The effort also will require confronting a host of thorny policy issues, including the management of intellectual property and development of technical standards for interoperability of records and software and reusable learning objects.

The task of creating practical markets for innovations in learning technology will fall primarily on state and corporate program managers. It is hoped that some of the federal demonstration programs can be conducted in ways that encourage participating organizations to explore basic changes in the way in which they approach education and training.

The most powerful tool available to the federal government is wise management of its own training programs. DOD invests at least \$50 billion annually in education and training. The overwhelming majority of this spending has nothing to do with skills needed in battle but supports training in areas such as financial management and engineering that are essentially identical to civilian equivalents. But DOD has not managed to design or implement a coherent plan to develop and deploy learning technology over the coming decade. Lacking a commitment to such a plan, DOD has not been able to use its market powers effectively to drive change in learning-technology products, and it has never organized an R&D effort commensurate with the need.

The Department of Homeland Security (DHS) is in even worse shape, despite the flexibility it enjoys by being a new organization. New security needs create enormous training challenges, because many dif-

ferent kinds of people need new skills and regular retraining. Simulations are particularly important for reinforcing skills that are seldom, if ever, required during routine public health and safety operations. Unfortunately, DHS has not yet acknowledged that research in improving training should be an integral part of its R&D mission.

Powerful economic forces are driving spectacular advances in computer processor power, mobile devices, and the software needed to deliver entertainment, answer consumer questions, and simulations for science and engineering. But these pieces will not self-assemble into the tools needed for education without an adequately funded, well-managed program of federal research, development, and demonstration in learning science and technology. The absence of a coherent national program to search for solutions in this area is, without question, the largest single gap in the nation's R&D program.