4. Opportunities for Action

The above strategies, while not a comprehensive set of possibilities, were chosen because they represent a core set of activities that NSF should pursue to make cyberlearning a reality. Some require *community workshops* to develop consensus; others are sufficiently well developed that *programmatic initiatives* can be launched soon. Here we articulate seven important opportunities for action that we feel have the greatest short-term payoff and long-term promise among the many that NSF might pursue: (1) advance seamless cyberlearning across formal and informal settings, (2) seize the opportunity for remote and virtual laboratories, (3) investigate virtual worlds and mixed-reality environments, (4) institute programs and policies to promote open educational resources, (5) harness the scientific-data deluge, (6) harness the learning-data deluge, and (7) recognize cyberlearning as a pervasive NSF-wide strategy.

4.1 Advance Seamless Cyberlearning

Education tends to be intentionally designed and provided either inside formal institutions like schools or as informal education inside science museums or afterschool centers. This can and should change, given the enormous changes in the digital resources, Web browsers, and other ICT platforms now available and increasingly used for learning outside formal designs. Advancing seamless cyberlearning across formal and informal settings is a large-scale opportunity where NSF investment could make a tremendous difference. Learners are in motion. But supports for their extended learning and education are not—to the detriment of the Nation and greater learning for all. Learning support systems can and will be organized along very different schemes than they are today, given the computational services made possible with cyberinfrastructure advances.

Seamless cyberlearning is learning supported by cyberinfrastructure so that it can be pursued productively either through learner intent, driven by interests or demands in the moment and regardless of location, or through intentionally designed educational activities, which learners can take advantage of as needed or when the situation requires (e.g., during schooling). This characterization indicates that seamless cyberlearning is about far more than just access to online courses anytime, anywhere—as important as these developments have been recently (Atkins et al., 2007). But what else does seamless cyberlearning entail? For example, youth today are extensively exploiting computing and mobile telephony outside of school to pursue their interest-driven learning through social networks. They use social network platforms like Facebook, MySpace, YouTube, blog sites, search engines, and instant messaging, not only for socializing, but to advance their learning and that of their peers about topics of personal consequence, such as hobbies, music, sports, games, fan culture, civic engagement, health, and nutrition, as described in the 2007 MacArthur Series on Digital Media and Learning.[20] Such interest-driven learning tends to be pursued outside school and often remains unconnected to school. At the same time, we know how vital the funds of knowledge and interests that learners develop in their everyday lives can be to promoting an integrated learning with formal education, in STEM domains, and for other life competencies (A New Day for Learning, 2007; Bransford et al., 2006). In fact, Estabrook, Witt, and Ramie (2007, p. iii) found that more youth as well as adults "turn to the internet…than any other source of information and support, including experts and family members" for help with many common problems.

Yet today learners for the most part have to exert their own efforts to coordinate the repertoires of knowledge and practices that they have developed through their experiences across many different settings, from classrooms to home, community to workplace. It is becoming increasingly evident that solving the problems

20 (Bennett, 2007; Buckingham, 2007; Everett, 2007; McPherson, 2007; Metzger & Flanagin, 2007; Salen, 2007).

affiliated with education and learning will require attending to design of the whole spectrum of experiences in which learning occurs, not only schooling and other formal educational institutions (Bransford et al., 2006). Creating environments for seamless learning requires vital cyberlearning infrastructure research and development (Ainsworth et al., 2005). Cole (1996) differentiates context as "that which surrounds us" and "that which weaves together." The latter definition makes clear how important cyberlearning infrastructure is likely to become, as it provides the technical substrate for weaving together in new designs the disparate learning and educational intentions and resources to make seamless cyberlearning a reality.

Seamless cyberlearning presents numerous "grand challenge problems" for research (Pea, 2007). Examples include (1) providing real-time access to developmentally relevant cyberinfrastructure learning support that will guide any learner toward meeting any learning standard, configuring requisite learning resources and human help from peers or mentors with verifiable reputations to help the learner attain such competencies in a certifiable manner; (2) creating "interest profiles" by inferring learners' interests from data-mining digital information on what they read, talk about, and attend to (with appropriate privacy safeguards), which can then be used for compiling engaging content and scenarios for their pursuit of enhanced skills and competencies using cyberinfrastructure; (3) providing Lifelong Digital Learning Portfolios for cyberinfrastructure management of all information media developed by a learner over a lifetime, in a manner usefully indexed for the learner's reflective learning and certification purposes.

While examples exist both in the United States and abroad of seamless cyberlearning starting to appear (e.g., Pea, 2006; Rogers, Price, Randell et al., 2005; Sharples, Taylor & Vavoula, 2007; Van't Hooft & Swan, 2007), without concerted and focused efforts to galvanize new developments using cyberinfrastructure, large opportunities for connecting learning experiences across settings are being lost. Achieving seamless cyberlearning will require advancing all of the strategies for building cyberlearning initiatives articulated in section 3 above, in particular, instilling a platform perspective, targeting new learning audiences with agile multipurposing of content, and instrumenting cyberlearning initiatives to capture metrics of learning experiences to improve them.

We recommend that NSF develop a program that will advance seamless cyberlearning across formal and informal settings by galvanizing public-private partnerships and creating a new interdisciplinary program focused on establishing seamless cyberlearning infrastructure and supports.

Research Questions:

1. How can cyberlearning infrastructure be used to mediate personalized learning across all the contexts in which it happens?
2. How can the "right" resources, from digital assets to human peers and mentors, be provided in any context to support learning needs in the moments in which they arise?
3. What different needs exist for different age populations and STEM learning domains?
4. What scaffolding systems are necessary to support learning in these distributed learning environments (Pea, 2004)?
5. How should theories of learning and instructional design be expanded to encompass learning across the boundaries of all the settings in which people learn?
6. What forms of digital portfolios will be necessary to manifest evidence demonstrating learning activities and performances?

4.2 Seize the Opportunity for Remote and Virtual Laboratories to Enhance STEM Education

Laboratories—both those focused on observation and experimentation and those focused on design, fabrication, and testing—are an essential part of the teaching and learning experience for many branches of science and engineering, from grade school through postdoctoral work. Ideally, laboratories provide a window on science in the making, showcase the ambiguity of empirical work, develop practical skills, and foster teamwork abilities (Singer, Hilton & Schweingruber, 2005). Achieving these goals has proven problematic. Laboratories are expensive to maintain, they require low student-faculty ratios, and they depend on both talented teachers and well-designed activity sequences. They raise safety and ethical issues around toxic substances and biological phenomena. Even when physical labs are available, they often are scarce resources, with student lab time tightly rationed and faculty guidance limited.

NSF has the opportunity to improve the impact of laboratories as well as reduce the cost of providing laboratory experiences, by promoting infrastructure for virtual laboratories and remote laboratories. Virtual laboratories include interactive simulations of laboratory equipment and experiments as well as interactive models or simulations of scientific phenomena that are too small, fast, or complex to explore in typical classrooms. These resources are completely scalable and can be embedded in powerful curriculum materials.

Remote laboratories allow students to access physical laboratory equipment via the Internet. Students at many different locations can share the use of physical equipment, as in the MIT iLabs project, which has microelectronics test equipment and other instruments on the Web. MIT is collaborating with universities in Nigeria, Uganda, Sweden, China, and Australia around its use (See inset on Inverted Pendulum). In addition, some labs are giving students access to unique, world-class observational resources such as major telescopes or electron microscopes. These programs offer promise but have scaling limitations.

We recommend that NSF mount a program to stimulate development of remote and virtual laboratories and to research effective ways to deliver this type of instruction. Many studies reveal the weaknesses of both hands-on and virtual laboratories (Singer et al., 2005). We recommend funding centers to identify effective ways to provide laboratory experiences given the power of cyberlearning technologies.

Creating effective cyber-enabled laboratory experiences requires an investment in curriculum design, experimental research, and infrastructure. Creating effective ways to use these new resources requires trial and refinement, ideally conducted in the range of contexts where they will be used. Even with a promising, tested starting point, faculty need time to reimagine and redesign their courses and determine the impact of the innovations.

For virtual labs, we need focused attention on pedagogy, scale, and interoperability. Much excellent work (some of it funded by NSF) has been done to build software chemistry sets, biology labs, and basic experimental physics labs. Research also suggests instructional conditions under which these resources succeed or fail (Clancy, Titterton, Ryan et al., 2003). We need to think about how to make materials extensive and customizable. We need to consider the role of standards or other types of interoperability frameworks, something that might be done in the context of the open cyberlearning platform initiative described above.

Schools and universities need to create and maintain the infrastructure investment to facilitate the deployment and use of remote labs. Faculty fear that they may be forced to extensively redesign courses every year or two because of the lab equipment available to their students. We need to explore financial and support arrangements for sharing experimental and fabrication equipment worldwide. We also need resource-sharing policies, and mechanisms for scheduling, and where necessary, rationing the use of such equipment. For sustainability, we may need recharge mechanisms or some other system of allocating equipment time among remote users.

The University of Queensland Inverted Pendulum Remote Laboratory

The University of Queensland (UQ) was struggling to provide adequate access for students required to take a control theory course in its undergraduate program. Physical space limited class size to 60 students. Additional laboratory space was unavailable. Course content was challenging and uninspiring. Take the classic inverted pendulum control experiment. The student attempts to balance a pendulum with the weighted arm pointing upright toward the ceiling rather than hanging toward the floor. Students first write a Simulink model for the experiment and then write a MatLabprogram to control the motor that swings the pendulum, giving feedback from two sensors in the pendulum arm. Students work in teams of four to iteratively attempt to balance the pendulum via their MatLab application. Prior to the introduction of the iLabs pendulum implementation, 5 percent of the teams balanced the pendulum by the end of the 5-week experiment, while spending 50 contact hours on the task.

The iLabs inverted pendulum made the experiment accessible beyond lab hours. The iLabs software interface presented the data graphically and via a mixed-media video cam overlaid with a data-driven animation of the pendulum arm. This let students see the results of one experimental run directly compared with another, thus visually showing the impact that their code revisions caused. But the learning story is more compelling still. Students ran 30 to 40 experiments per team in the 5-week period prior to iLabs. With the remote lab, students ran 3,210 experiments, or on average 39.1 experiments per student. Contact hours decreased to 4 per week in the iLabs implementation from 10 hours per week previously, and the success rate for students balancing the pendulum went from 5 percent to 69.5 percent. Class size was increased from 60 students to 84 students, and student ratings of the course rose from 4.19 to 4.78 on a 5-point scale.

Finally, continued investment and emphasis is needed to ensure that scientific and engineering research programs that rely on cyberinfrastructure support perform the often modest incremental work to make their data available for analysis and reuse in instructional settings.

4.3 Investigate Virtual Worlds and Mixed-Reality Environments

Students today spend large amounts of time interacting with content and communities that are located in digital—or virtual—environments. These environments are both motivating and engaging and have many of the qualities defining the innovative potential of cyberlearning: they are *networked*, which allows for exchange between a variety of learners who do not have to be in the same physical space; they are *customizable*, in that they can be tailored to fit the needs and interests of learners on demand; and they support *computationally rich* models and simulations, which offer learners access to rich STEM content. As a result, online digital environments hold promise for cyberlearning in both the short and long term.

These digital environments sometimes take the form of virtual worlds. In these worlds learners can create a digital character, or avatar, to represent themselves, which they use to move around inside a virtually rendered world space shared with thousands of other avatars. They can socialize with others; build objects and share them; customize parts of the world; and hold lectures, do experiments, or share data. Individuals of many different ages are currently members of virtual worlds like Second Life, Whyville, There, and Activeworlds.

While virtual worlds occupy purely digital space, another kind of digital environment has similar promise: mixed-reality environments that combine digital content and real-world spaces. Interaction goes beyond a simple face-to-screen exchange, as in the case of virtual worlds, instead incorporating surrounding spaces and objects. To picture this, imagine a group of students in a lab interacting with a physics simulation being projected on the floor below them. Through the use of a wireless controller and motion sensors, the students are physically immersed in the simulation. They can *hear* the sound of a spring picking up speed, *see* projected bodies moving across the floor, *feel* the controller in their own hands, and integrate how the projected image moves in accordance with their own body movements to construct a robust conceptual model of the system.

The benefits of utilizing mixed-reality environments and virtual environments for learning are many. In the case of virtual worlds, time and distance become irrelevant, allowing cyberlearning to occur any time, any place. The benefit to data collection practices cannot be overstated. Visualization is enhanced dramatically, which creates opportunities for new modes of interaction, new audiences, and new models of assessment. With mixed-reality environments emerging, sensing technologies can be used to diagnose a learner's interests and patterns of activity (i.e., what they search for, read, listen to, talk about, and attend to), allowing the system to learn about the kinds of choices students are apt to make. Consider the previous example. Within the physics simulation, projected mass and spring models move across the floor as dynamic sounds articulate the velocity and acceleration of the moving particles. Relational models allow the system to identify that the student would benefit from increased physical activity in the space to drive the system in ways that reveal important relationships that otherwise remain hidden. Models of a student's past actions reveal a hesitancy to explore movement but a heightened sensitivity to auditory feedback. Faced with these data, the system encourages the student by triggering an *adaptation*: small physical movements in the space give rise to amplified changes in the sound that encourage further embodied exploration between these attributes of the underlying system. The student is drawn out through the sonic experience in the environment, and correct intuitive understanding of the simulation is reinforced through actual experimentation.

The ability for mixed-reality environments and virtual worlds to provide real-time access for learners—whatever their developmental capabilities or interests—means such platforms can guide students toward meeting any learning standard. In addition, these spaces allow for a kind of collaboration and exchange that have been difficult to achieve in traditional learning environments.

NSF should begin investment in leveraging the use of virtual world and mixed-reality environments for STEM learning. This includes developing an infrastructure for assessment and support and requires connection to larger concerns around openness, virtual laboratories, and the data deluge.

4.4 Institute Programs and Policies to Promote Open Educational Resources

In education, as in so many areas of human activity, the burgeoning of the World Wide Web is the most significant cyberinfrastructure phenomenon to emerge in the past 50 years, and the one with perhaps the greatest transformative potential for education. In the 1990s, educators regarded the Web primarily as a low-cost distribution mechanism: author-educators (including NSF grantees) would create materials and make them available on Web sites for students and teachers to access.

Increasingly, however, the Web is being recognized as an enabler for collaborative creation of significant information resources that aggregate contributions from hundreds or thousands of individuals. Wikipedia is the most famous example of the collective intelligence of crowds. Collaborative creation is especially appropriate for educational materials, including text, video, simulations, games, and other content, because the effectiveness of educational materials often hinges on the ability to adapt them to fit the needs of particular cultures, school systems, and classrooms—even individual teachers and learners. In education, therefore, resources on the Web are especially valuable if they are *open educational resources* (Smith & Casserly, 2006).

Open educational resources (OER) are teaching, learning, and research resources that reside in the public domain or have been released under an intellectual property license that permits their free use or customization by others. Open content includes video, multimedia, cognitive tutoring courses, open textbooks, journals, books, data, laboratories, music, library collections, lesson plans, simulations, games, virtual worlds, and so on. Other OER include freely usable and reusable tools to support open content, including open-source content and learning management systems, search engines, communication systems, and intellectual property licenses. Major institutions such as the BBC, U.S. public television stations, and Harvard University are unlocking their resources from behind passwords, intranets, and archives and figuring out ways of making them available to everyone, everywhere. But it is the freedom to share, improve through rapid feedback loops from users and other experts, reprint, translate, combine, or adapt them that makes OER educationally different from resources that can merely be read online at no cost. The importance of openness is now being recognized worldwide, both in developed and developing nations (Atkins et al., 2007; Wilensky & Reisman, 2006).

There are noteworthy examples of OER in many STEM-related areas. Three examples of special interest are (1) the OpenCourseWare activities described earlier, which have spread across the world and receive multimillion visits monthly; (2) open textbooks to address the high price of texts, the lack of quality in many of them, and their scarcity in many developing nations; and (3) open full courses in mathematics, engineering, and science. The open textbook movement is picking up momentum at the secondary school and community college levels, especially in mathematics and science. The power of an open textbook is that it provides the opportunity for users to modify, adapt, and extend it to adapt to new knowledge developments in its field and to improve its usefulness in classrooms and for students in different cultures, at different levels of prior knowledge, and with different languages. Open textbooks can also be augmented with video, simulations, and assessments and may be modified to provide feedback loops for students to relearn material that the assessments indicate they have not mastered. In effect, the textbooks can facilitate individual tutoring, sometimes called personalization.

Some open full courses already have the capacity to provide such personalization. In particular, the Carnegie Mellon cognitive tutor courses have this capacity. In a recent experiment, a randomized group of Carnegie Mellon students used the cognitive tutor statistics course and, given only half the time (a 7-week semester), were more successful on the final examination than students who had taken the normal full-semester lecture course (Lovett, Meyer & Thille, in press). The power of technology to provide personalization will become greater and greater as we improve the quality of the instructional aspect of the Web-based courses. Ensuring the quality of the content requires considerable one-time costs as it is built into modular courses and from then on some modest updating costs. However, the net costs over even a short period and the guarantee of high-quality content for our students suggests that we can imagine a substantial increase in access and learning.

In open access of research results and scholarly journals and other publications, there are extensive and complex developments both in the United States and elsewhere. A move toward open access is happening rapidly both at the level of funding agencies (notably the recent requirement for the deposit of articles describing research funded by the National Institutes of Health [NIH] within a year of publication, so that these articles are available for public access) and at the level of individual institutions of higher education, such as Harvard. In addition there are literally hundreds of open academic journals. There are also movements to encourage the sharing and reuse of data (although this is subject to constraints such as human subject regulations, of course). All these developments will help to contribute to the effectiveness of cyberlearning initiatives. We would certainly welcome, for example, a NSF policy on articles reporting research results that was at least as strong in encouraging open access as that adopted by NIH (and would also urge that some thought be given to consistency of policy across Government agencies).

With regard to materials developed with NSF funding that are primarily educational in nature (rather than reports of research results or research data)—and recognizing that this is a slippery distinction—we believe that an NSF-wide policy that encourages principal investigators to make their materials open and to be concerned about their sustainability is essential.

1. NSF should require clear intellectual property and sustainability plans as part of grant proposals for educational materials it supports. The default expectation around intellectual property is that the materials should be released on the Web as open educational resources under a license provided by Creative Commons, where appropriate (perhaps with attribution only), at some identified point within the term of the grant. This will facilitate machine searching and processing of the material and also help with reuse and recombination of materials. As part of the evaluation of proposals, grant reviewers should give careful attention to these plans, and also to any arguments advanced for more restrictive conditions on NSF-funded educational materials.

One challenge of open educational resources is that of creating sustaining funding models so that materials remain available and improve over the long term. For some OER materials there may be long-term public or university funding. Other materials might be maintained by subsequent reusers. The cost of storing vast quantities of content drops dramatically every year, and organizations such as the Internet Archives make storage and bandwidth free. There could be commercial models as well. That might seem contradictory from the perspective of businesses built on charging for access to material. And yet, there are significant viable businesses on the Web that are based on open resources (e.g., in providing support services for Linux open operating systems). In education, one could imagine commercially successful service opportunities involving personalization, tutoring, and certification, built on a base of open resources. Other potential sustainability models include using advertisements and corporate sponsorships. It is important to explore these and other models as a step toward realizing the potential of educational collaboration and continuous improvements on a wide scale.

2. NSF should launch a program to identify and demonstrate sustainable models for providing open educational resources, whose goal is to create mechanisms whereby educational materials developed by grantees will continue to have impact long after NSF support has ended. All materials development grants should include required discussions of sustainability, and this should be an important criterion in proposal evaluation.

In some cases, it may be appropriate for NSF itself to provide infrastructure for sustainability, similar to how the National Library of Medicine and its National Center for Biotechnology Information help ensure continued open access to the results of biomedical and life-sciences research, particularly that funded by NIH. In other cases, services within the data management and stewardship components of the cyberinfrastructure being developed under initiatives such as DataNet may meet these needs.

4.5 Harness the Deluge of Scientific Data

The amount of data in the world is at least doubling every year. As a result, it is becoming increasingly difficult to find relevant information and to extract meaningful knowledge from the ever-larger quantities of raw data. This challenge is present at every level of society, and it creates a huge demand for people with appropriate skills in navigating and processing information.

With the steady improvement of digital sensors and of the computers processing their data, every step of the scientific process is also changing. A new scientific paradigm, e-Science, is emerging, where computing techniques are an essential part of every step in the scientific workflow (Borgman, 2007; Hey & Trefethen, 2005). Inexpensive imaging sensors (CCDs) first revolutionized astronomy, and a decade later they fundamentally altered photography, turning cameras into inexpensive, portable, and embedded devices. Gene sequencers and gene chips have enabled the assembly of the first human genome; in 10 years this technology will become part of everyday medical diagnostics. It is expected that in a few years the number of online sensors accessible through the Internet will exceed the number of computers today.

Modern scientific experiments use computers as an integral part of the data collection process. There is an emerging trend of large collaborations that are formed to undertake massive data collection efforts (virtual observatories) on every imaginable scale of the physical world, from elementary particle physics to material science, biological systems, environmental observatories, remote sensing of our planet, and observing the universe. These collaborations generate enormous datasets that are (or soon will be) made available online, for public use. Scientists will have this vast repository of data available to test their hypotheses and to combine data in novel ways to make new discoveries.

These virtual laboratories/observatories represent an entirely new way of approaching scientific problems. To be successful in these new approaches, scientists need to acquire a multitude of skills and ways of thinking. They have to be equally at home in their narrowly defined disciplines, in data management skills, and in computational skills that might require statistical analyses over billions of data points. These shifts also underscore the fact that the nature of scientific computing has become more data-centric than computing-centric (Bell, Gray & Szalay, 2006; Gray, Liu, Nieto-Santisteban et al., 2005).

We need novel ways of harnessing this data deluge and to turn it into new opportunities for learning, either involving new groups of people or engaging students in a totally different fashion. A recent trend to capitalize on people's interest in gaming has been to involve them in science-related activities that resemble gaming activities while delivering educational content (see inset on GalaxyZoo).

There is very little past expertise in this area, as the data deluge is a relatively recent phenomenon. We need to invent ways to teach and train the next generation as we go. Yet this is an area with enormous potential. The need for such skills cuts across all fields of science and much of society. This also raises an interesting question along a strategic continuum: where should our training focus? Should we solve these huge interdisciplinary problems by encouraging large interdisciplinary teams to work together, or should we increase the versatility of individuals and provide an ever-faster rate of retraining?

A paper on "antedisciplinary science," the science that precedes the emergence of new disciplines, argues rather forcefully that today we reward interdisciplinary teams, while the same cannot be said about individuals with a broad knowledge base (Eddy, 2005). In today's accelerating world, it is clear that we need to think more actively about how to provide a flexible enough scheme for science education to allow new, groundbreaking ideas to be translated into specific training programs for the next generation at an ever-faster rate. And little today in science is accelerating faster than computing, both using cyber-infrastructure and its methods for teaching about science.

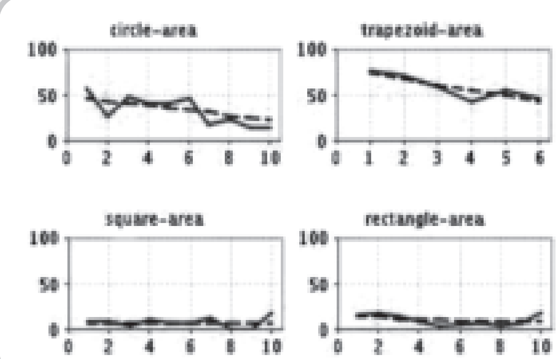
4.6 Harness the Deluge of Learning Data

Scenario: To prepare for her algebra class in fall 2015, Ms. Washington gets online to access digital portfolios of her incoming students, which include records of their past mathematical successes and challenges. From experience she feels that some students have trouble learning algebra because they do not see it as interesting or important, whereas others have weaknesses in crucial prerequisites, like negative numbers. Because a good share of students' prior mathematics work has been done online in cyberlearning systems (simulations, virtual labs, math games, homework, tutors, online assessments, etc.), Ms. Washington has access to a rich set of quantitative and qualitative information about her students. The digital student portfolios provide summary statistics and representative examples of student past performance, and she knows to heed the recorded levels of cognitive and psychometric reliability of the different kinds of data available. Using these data, she identifies two risk groups among her incoming students: the disengaged and the unprepared. She then begins to plan activities specifically targeted to these groups, including selecting cyberlearning resources from the Internet such as collaborative math games for the disengaged and intelligent tutors to adaptively help the unprepared in their specific areas of need.

A large number of cyberlearning projects have been accumulating vast amounts of student data in a variety of domains and grade levels. These include interaction data from online courses,[21] intelligent tutoring systems, virtual labs, and online assessments in subject matter, including elementary reading (Zhang, Mostow & Beck, 2007), middle school science (Buckley, Gobert & Horwitz, 2006), middle and high school mathematics (Koedinger & Alevan, 2007), and college-level science[22] (Van Lehn, Lynch, Schultz et al., 2005; Yaron, Cuadros & Karabinos, 2005). There are also a large number of projects collecting and analyzing video of classroom and informal learning interactions (Goldman, Pea, Barron et al., 2007). In the future, we expect increasing amounts of learner data available from formal and informal learning activities in the context of online chat, cell phones, games, and even toys. Further learning-relevant data from brain imaging and physiological sensors will also become increasingly available and useful, especially when coupled with other forms of behavioral data (Varma, McCandliss & Schwartz, 2008). Machine learning, psychometric, and cognitive modeling methods are increasingly being combined to discover improved cognitive-affective-psychometric models of student achievement and engagement through embedded assessment in cyberlearning systems. [23]

Open-learning data repositories are beginning to emerge,[24] along with new computational techniques for analyzing such data. A new field of educational data mining is emerging, as indicated by the first conference on the topic in 2008.[25] NSF should encourage data contributions, data use, new algorithm development, and, most important, common standards for data storage so both data and algorithm sharing are facilitated.

The figures below are learning curves generated from an open repository of learner data[26] collected during student use of an intelligent tutor for geometry.



These learning curves show a change in student error rates (the y-axis) over successive

21 http://www.cmu.edu/oli/
22 See also http://www.cmu.edu/oli/.
23 Machine learning has been employed to create automated embedded assessments systems that address both student cognitive achievement and affective engagement (Baker, Corbet, Roll et al., in press; Cen, Koedinger & Junker, 2006).
24 Large-scale educational data can be found at http://www.icpsr.umich.edu/IAED. Fine-grained student learning behavior data from recorded online, paper, and speech interactions in math, science, and language courses can be found at http://learnlab.web.cmu.edu/datashop (Koedinger, Cunningham & Skogsholm, 2008). Video data on classroom interactions can be found at http://talkbank.org/data, among other sites (Derry, 2007).
25 http://www.educationaldatamining.org/EDM2008
26 The Pittsburgh Science of Learning Center's DataShop can be found at http://learnlab.web.cmu.edu/datashop.

opportunities to practice and learn (the x-axis) in attempting to apply a geometry concept (e.g., circle-area) during problem-solving. The solid line shows average student data, and the dashed line shows predictions from a best-fitting cognitive-psychometric model. Notice how for the circle-area and trapezoid-area concepts, the student average error rate is initially quite high, but with practice and tutoring it improves. In contrast, the learning curves for square-area and rectangle-area indicate that students have little trouble right from that start (less than 10 percent error rate), but nevertheless get lots of practice (10 opportunities). Given such visualizations, it is not hard to conclude that a redesign is needed to reduce the unnecessary overpractice on some concepts and instead spend the valuable instructional time where it is needed. Just such a redesign was done and compared to the original version in a randomized controlled classroom study (Cen, Koedinger & Junker, 2007) that ran inside the technology and was essentially invisible to students and teachers. The results indicated a 20 percent savings of student time without any loss in learning, transfer, or retention outcomes.

4.7 Recognize Cyberlearning as a Pervasive NSF-wide Strategy

This may be the most important of all of the recommendations. All disciplinary directorates within NSF fund the development of resources—tools, software, learning objects, databases, etc.—that have either primary or secondary roles as educational materials. Just as cyberinfrastructure approaches now underpin disciplinary thinking about new research initiatives, we believe that cyberinfrastructure and cyberlearning ideas need to inform the work of the disciplinary directorates in shaping programs and evaluating proposals dealing with educational materials. One way to encourage this would be to conduct a series of workshops on cyberlearning in specific disciplines, much as NSF conducted an earlier series of workshops on the potential for cyberinfrastructure to advance research in specific disciplines.