# Online Classroom Assessment Project

# Evaluation Report #1

**July 1, 2002**

**Prepared for the**
**Bill & Melinda Gates Foundation**

*Fouts & Associates, L.L.C.*

# Online Classroom Assessment Project

# Evaluation Report #1

**July 1, 2002**

**Prepared for the**
**Bill & Melinda Gates Foundation**

**Carol J. Brown, Ed.D.**

*Fouts & Associates, L.L.C.*

# *Executive Summary*
# Online Classroom Assessment Project Evaluation Report #1

**Carol J. Brown**

Washington is one of several states exploring the benefits and challenges of using web-based assessments at the classroom level. The goal in developing an online assessment system is to provide teachers with a diagnostic and assessment tool they can use to make adjustments in instructional strategies to align more closely with the EALRs, benchmarks, and frameworks. Funded by the Bill & Melinda Gates Foundation, a two-year pilot test of the system was administered by the Office of the Superintendent of Public Instruction (OSPI). During the first year of the project, grades three, six, and nine were given training and materials to assist them in using Assess2Learn™, and grades four, five, seven, and eight were brought into the program during the second year. Administration of the program was problematic and may have limited progress during the first two years. Approximately 60 teachers were involved in the training, although fewer than ten actually appeared to be using it at the end of the second year.

### Evaluation Question 1: *What impact does a system of online assessments have on instructional planning and teaching strategies?*

Teachers noted several benefits to using the online assessments. First, the immediacy of results was viewed as an important aspect of the tests. Traditionally, results from standardized tests are not available for several months, meaning that their usefulness for error analysis and remediation are lost. In the case of the online assessments, teachers and students were able to review test results together to understand errors and set realistic and meaningful learning goals. Teachers also found the tests useful in identifying strengths and weaknesses of individual students and an entire class. This allowed them to form small need-based groups and provide additional instruction as needed.

### Evaluation Question 2: *What practical issues need to be considered when assessing students in an online format?*

Several interesting issues were uncovered in terms of the practicality of assessing students in an online format. First, technical glitches were common. At the very least these were frustrating and confusing, and in some cases they were significant enough to cause users to quit the program. Platform compatibility, inconsistent servers, and formatting issues were among the problems encountered. In addition, teachers were in general agreement that the writing component of the assessments was time consuming and not worth the information gained. Keyboarding was a factor as well. Teachers found student writing to have less depth and creativity when done on the computer.

### Other Findings.

Teachers were enthusiastic about the training they received in preparation for administering the online assessments. Being able to participate in the development of test items, including how to score various types of questions, was invaluable. It helped them understand assessment in a larger context, and gave greater meaning to the WASL (Washington Assessment

of Student Learning). In addition, teachers believed their participation in the online training would be valuable in providing similar training to their building colleagues in those cases where the program was to be expanded. Teachers need not be technology experts to use the system, they said, and could be brought on board with relative ease.

## Conclusion and Recommendations

Evaluation results revealed several benefits of the assessment system, according to teachers. Information gathered from the assessments was used to identify learning strengths and weaknesses, and allowed teachers to modify and adjust instruction to meet specific individual and classroom needs. In addition, immediate feedback proved helpful in the analysis of errors and academic goal-setting. Teachers also predict that the system will be valuable in looking at student performance over time to see learning trends. Challenges were numerous, and included technical glitches, compatibility issues, inconsistent servers, and formatting problems. These challenges are similar to those experienced by teachers in other states using different assessment systems, which is not surprising given the fact that the development of electronic assessments is still in its early stages. However the potential of online assessment outweighs the problems, and it seems likely that work on these systems will continue. Although the first two years of piloting Assess2Learn™ in Washington have been difficult for a number of reasons, lessons have been learned that will significantly improve future implementations. Early results suggest that the Assess2Learn™ system can be, in time, a valuable diagnostic tool for teachers in the state.

The following recommendations are based on the evaluation of the program over the last two years.

- Consider limiting participation to a small number of teachers and classrooms until all of the technical problems are resolved.
- If the project is to be expanded to a large number of classrooms, considerable technical support for those classrooms should be provided until the major technical problems have been resolved.
- Ensure that whoever is selected to administer the project has the time, interest, and organizational skills necessary to direct and support an innovative program such as Assess2Learn™.
- Resolve Macintosh compatibility issues.
- Re-evaluate the writing component. As it stands, teachers do not see it as an efficient component of the system.
- Make design changes in the system that will allow even young children to use their Office skills.
- Continue the training sessions, including the opportunities for teachers to participate in item selection and item analysis.
- Once the system is operating dependably, plan an agenda of presentations to education groups across the state.

# TABLE OF CONTENTS

# Online Classroom Assessment Project
# Evaluation Report #1

## INTRODUCTION

The Online Classroom Assessment Project, funded by the Bill & Melinda Gates Foundation, is a program designed to provide a series of web-based assessments for educators in Washington state. Developed by Riverside Publishing and administered by the Office of the Superintendent of Public Instruction (OSPI), Assess2Learn™ was pilot-tested by third through ninth grade teachers during the 2000-2001 and 2001-2002 school years. The evaluation of the Online Assessment Project focused on two evaluation questions, and data were gathered from several different sources, including written reflections, teleconferences, and phone interviews. This evaluation report describes the first two years of the Online Classroom Assessment Project in Washington State.

### Description of the Online Classroom Assessment Project

To explore the potential of online assessment, the Bill & Melinda Gates Foundation awarded a grant to the Office of the Superintendent of Public Instruction (OSPI) to work with Riverside Publishing in developing and implementing an online assessment system. The goal of the assessments is to "provide teachers with a diagnostic and assessment tool they can use to make adjustments in instructional strategies to align more closely with EALRs, benchmarks, and frameworks" related to Washington's statewide WASL test (Washington Assessment of Student Learning). The Assess2Learn™ system was designed to be used three times a year (Fall, Winter, and Spring) by students in grades three through nine. Tests are progressively more difficult, and provide informal, low-stakes "WASL-like" opportunities to help students prepare for the state test, while also measuring growth. Teachers benefit by getting immediate feedback regarding their students' abilities, which in turn allows them to modify instruction as needed.

Teachers of grades three through nine spent a minimum of six days in training seminars helping in the development of test items, and were also trained in how to administer and score the assessments. Costs of the training days, with the exception of substitute pay, were covered by the grant. Participating teachers were expected to implement the assessments in their classrooms over the course of the school year and to share their perceptions of the assessments with Riverside, OSPI, and the Gates Foundation.

Participating teachers were from school districts across the state that had been previously awarded Gates District Grants. Approximately 60 teachers took part in the training, although their implementation efforts were not consistent. Training during the

2000-2001 school year was directed at teachers of grades three, six, and nine, and teachers of grades four, five, seven, and eight were trained during the second year. Representatives of Riverside Publishing provided the training and worked at different times with both grade-level groups and content-area groups. A schedule of training sessions is included in Appendix A.

## Background

Traditional paper and pencil tests continue to serve as the primary means of assessing students in elementary and secondary schools across the United States. However, as the demand for school and student accountability increases, the limitations of traditional testing formats become more apparent. Results of standardized tests, for example, are reported too slowly to be of use in modifying classroom instruction. In addition, as technological literacy becomes a necessity in both academic settings and the corporate world, students are increasingly using computer technologies as learning tools. This raises a concern about the lack of alignment in learning and assessment and has led a number of states across the country, including Washington, to explore the potential of online assessment.

### Research on Past and Present Use of Computers and Related Technology[1]

The computer was introduced into education in the 1970s and its first use had teachers and students learning to program. Since that time there has been an evolution of best practices. As software gained in sophistication, the computer became the tutor or surrogate teacher. Students followed the commands on the computer screen and received rewards for correct answers. They also began to learn through playing games and simple simulations. Teachers of writing discovered the value of using a word processor and soon students were writing more and revising with ease. Other teachers saw the value of the computer in creating rich learning environments and had students using databases, spreadsheets, presentation and research tools across all subject areas. Next, the Internet impacted technology use. Suddenly there was a volume of knowledge available to students with access and a network of people throughout the world that enhanced communication and the exchange of ideas. Real problem solving in collaborative groups became the norm in some classrooms. Online courses were available and students in rural areas had expanded learning opportunities in a variety of subject areas. Previously abstract concepts could now be illustrated and manipulated because of technology advancements. A whole new learning environment became possible and along with it, the need for appropriate assessment tools.

**Technology as a Transformational Agent and Learning Tool.** Of particular importance is the potential for computers and related technological tools to be used in transforming the classroom, such that students' educational experience is qualitatively improved. In the past decade, the use of computers has expanded from use primarily as an instructional delivery medium to use as a transformational too and integral part of the learning environment. In fact, many proponents of the current reform efforts see

[1] This section is taken from Fouts, J.T. (2000), *Research on Computers in Education: Past, Present and Future.*

technology as a vital component of a new educational paradigm in which the curriculum, teaching methods, and student outcomes are re-conceptualized (Means, 1994). This view was adopted by the U.S. Department of Education at least as early as 1993. In "*Using Technology to Support Education Reform"* (United States Department of Education, 1993) it was stated that "technology supports exactly the kinds of changes in content, roles, organizational climate, and affect that are at the heart of the reform movement."

In these settings the computer and related technologies are serving at least four distinct purposes: (1) they are used as previously, to teach, drill, and practice using increasingly sophisticated digital content; (2) they are used to provide simulations and real world experiences to develop cognitive thinking and to extend learning; (3) they are used to provide access to a wealth of information and enhanced communications through the Internet and other related information technologies; and (4) they are used as productivity tools employing application software such as spreadsheets, data bases, and word processors, to manage information, to solve problems, and to produce sophisticated products.

One of the central components of school reform is the desire for higher academic standards and a stronger focus on higher order thinking, problem solving skills, and learning associated with "real world" applications. To accomplish these ends, a new learning environment for schools is necessary. Proponents of school technology assert that it is just that type of environment and those types of learning that are facilitated by the new technology. At the same time, there is a predominant belief that the traditional standardized tests are inadequate to measure the types of learning teachers are now being asked to teach. This has resulted in a demand for new assessment procedures for the learning outcomes. Those new assessments are taking the forms of projects, portfolios, demonstrations, and new standards-based tests.

The changing use of technology reflects the changes in understanding over the last two decades about how the mind works and how children actually learn. There is a strong base of basic research that supports these ideas. This research is derived from the findings of researchers in developmental psychology, cognitive psychology, linguistics, and neuroscience and coupled with the philosophical ideas of constructivism (Duffy & Cunningham, 1996). Taken together they serve as the basis for many of the current beliefs about what and how children should learn in school. "Our understanding of human learning has . . . evolved based on a wealth of evidence collected over a wide range of different domains and media from which a process based on the passive assimilation of isolated facts to one in which the learner actively formulates and tests hypotheses about the world, adapting, elaborating, and refining internal models that are often highly procedural in nature (Shaw & President's Committee of Advisors on Science and Technology, 1998). The National Research Council's Committee on Developments in the Science of Learning articulated an idea central to this new understanding of human learning: "A fundamental tenet of modern learning theory is that different kinds of learning goals require different approaches to instruction; new goals for education require changes in opportunities to learn" (Bransford, et al., p. xvi). "These new learning

opportunities should take place in learning environments that are student centered, knowledge centered, assessment centered, and community centered…" Key conclusions:

- Because many new technologies are interactive, it is now easier to create environments in which students can learn by doing, receive feedback, and continually refine their understanding and build new knowledge.
- Technologies can help people visualize difficult-to-understand concepts, such as differentiating heat from temperature. Students are able to work with visualization and modeling software similar to the tools used in non-school environments to increase their conceptual understanding and the likelihood of transfer from school to non-school settings.
- New technologies provide access to a vast array of information, including digital libraries, real-world data for analysis, and connections to other people who provide information, feedback, and inspiration, all of which can enhance the learning of teachers and administrators as well as students (Bransford, et al. p. xviii-xix).

If indeed learning is being transformed through the use of computers and related technology, so too must the means by which this learning is assessed.

## Rationale for Online Assessment

Researchers, technology experts, educators, and policymakers give several reasons for pursuing online assessments to measure student achievement. These include the increased demand for accountability in schools and classrooms, the limitations of traditional standardized paper tests, and the lack of alignment in classroom practice and assessment. As well, there is a growing use of electronic training and assessment in business and industry that necessarily has an impact on K-12 schooling.

### Greater Accountability

Reauthorization of the Elementary and Secondary Education Act calls for yearly testing in reading and math for students in grades 3-8. Tests must be aligned with state content standards and must include a diagnostic reporting mechanism for individual students. This demand for greater accountability has prompted many state policymakers to look for strategies that more efficiently measure student performance. To this end, the potential of electronic assessment is being explored in several states, including South Dakota, Oregon, Virginia, Georgia, Idaho, and Washington. Online assessments offer the benefit of immediate feedback, which provides teacher and students a means of identify areas of weakness and analyze errors. Electronic assessment also provides a method for tracking academic growth over time. Together, these features could facilitate continuous, formative assessment of student learning and provide the accountability necessitated by ESEA and state legislation. As an example, the Northwest Evaluation Associate has devised a scale for Idaho's online test that will "allow scores to be reported consistently across grades and years. That permits educators to determine what value they are adding to a youngster's performance from grade to grade," noted company president Allan Olson. "It also makes it easier to target instruction" (Olson, 2002, p. 7).

### Limitations of Traditional Test Results

For test results to be useful and to drive instruction, they need to be delivered in a timely manner. This is seldom the case with traditional standardized tests, however. Results from fall assessments are often not available until late winter or spring, too late to be used in identifying learning gaps or modifying instruction. Results from spring assessments, on the other hand, are not typically seen by teachers until the following school year. Given this reporting gap and the academic time lost to summer break, the formative value of these tests is minimal. Online assessment offers the potential for immediate feedback. There is the hope that getting scores right away would allow teachers to identify learning gaps, areas of weakness, or deficient test-taking strategies. Teachers could then individualize instruction for students or for the class, thereby making better use of instructional time. In addition, advantages are seen for schools with highly mobile populations of students. Incoming students could be tested immediately on entering a school and with instant feedback could be appropriately placed.

Several states are also exploring the value of adaptive testing, which is designed to provide individualized assessments for every student at every testing opportunity. Adaptive online assessments modify the difficulty level of a test based on a student's performance. The computer would switch a student who had missed several items to a series of questions that were less advanced and which, it is hoped, would isolate weaknesses in skills. A student who completed a string of questions without errors would be switched to a more advanced track of questions (Trotter, 2001, p. 35). "It's designed specifically to provide every child a test that is challenging from beginning to end," according to Olson (2002, p. 6).

### Alignment of Learning and Assessment

As technology becomes an integral part of classroom teaching and learning and students do more of their work on the computer, a "widely unrecognized but critical limitation" is emerging (Russell & Haney, 2000, p.3). Research suggests that students who use computers on a regular basis to write, to analyze data, and to present information are at a disadvantage when they are assessed in a different mode. Several studies have been conducted which appear to support this claim. One study, for example, compared student test performance based on whether a computer or pencil and paper was used. While scores on the multiple-choice section of the tests (math, science and language arts) differed little, scores on the open-ended section of the tests were significantly different (Russell & Haney, 2000). Another study of student performance on a state-mandated writing assessment again compared students who wrote using the computer and students who used paper-pencil. Results indicated that "open-ended language arts items that require students to generate responses using paper and pencil severely underestimate the achievement of students accustomed to writing using a computer" (Russell & Plati, 2001, p. 1). "As in the two previous studies, this experiment found that students who wrote their compositions on computer produced longer responses that received higher scores. This effect was larger in grade eight than in grade ten, but was statistically and practically

significant in both grade levels" (p. 27). "Educators and policymakers must revamp tests to reflect those changes. Tests that are delivered in a mode that's different from the one students are learning in will eventually become indefensible" according to Randy Bennett, a researcher with ETS (Trotter, 2001, p. 32). "They won't be credible to parents, teachers, and students. In the end, they won't be credible to the testing community either."

### Increased Use of Technology in Business, Industry and Education

One of the most frequently cited reasons for integrating technology, including electronic assessment, into the classroom is the fact that business and industry has come to depend on information technology. Jobs that require technological knowledge and skills are increasing, and it has been estimated that half of all workers will one day employed by industries that use or produce information technology (Henry, Buckley, Gill, Cooke, Dumagan, Pastore, & LaPorte, 1999). In addition, distance learning is becoming a viable option for those wishing to continue their education and training, and virtual high schools are being developed in a number of states across the country. Internet-based learning is an essential tool in businesses and schools, and this is reflected in the Web-based Education Commission's report to the President and Congress:

> The question is no longer *if* the Internet can be used to transform learning in new and powerful ways. The commission has found that it can. Nor is the question *should* we invest the time, the energy, and the money necessary to fulfill its promise in defining and shaping new learning opportunities. The commission believes that we should (Kerry & Isakson, 2000, p. 134).

If this is indeed true, it has important implications for education, and for assessment. According to assessment researchers, learning needs to be measured in the context in which it takes place. As the Web-based Education Commission has stated, "Perhaps the greatest barrier to innovative teaching is assessment that measures yesterday's learning goals…..Too often today's tests measure yesterday's skills with yesterday's testing technologies….paper and pencil" (Kerry & Isakson, 2000, p. 59).

Several other practical advantages to online assessment have been noted, including the following:

- *Develop tests.* Computer technology makes questions easier and cheaper to write. With inexpensive, high-quality test questions, teachers will have better resources for classroom assessment and students will have more opportunities to practice critical skills.
- *Deliver tests.* Computer delivery enables tests to be individually administered. Different tests can be administered simultaneously to different students in the same classroom. Special accommodations, such as extended time, can be made for those with disabilities.
- *Present new types of test questions that include audio, video, and animation.* This capability makes it possible to measure important skills that paper-and-

pencil tests simply cannot assess (e.g., skill in using computers to search for information). It also makes the need for specialized testing equipment, such as audiocassette recorders and VCRs, obsolete.

- *Transmit responses to essay and other open-ended questions to human graders for scoring on computer terminals or to programs that can score such questions automatically.* These capabilities make it practical to include more open-ended questions in state assessments because the cost and time required for grading are greatly reduced. Including more open-ended questions will encourage students and teachers to focus on problem-solving activities that are more like the ones required for success in work and advanced academic environments.
- *Distribute test results.* With electronic distribution, parents, students, and decision-makers can get results faster. In addition, parents can try sample questions and get immediate feedback, giving them a better understanding of the skills that state assessments measure (Linn, 2002, p. 3).

## Challenges of Online Assessment

While the benefits of online assessment are compelling, there are significant challenges in implementing such programs as well. These include cost, dependability of the technology, privacy and security concerns, test format, and measurement issues. Each needs to be considered when making decisions about an online assessment system.

### Cost

Certainly it is not inexpensive to develop and implement an online assessment program. The technology, including hardware, test software, and delivery mechanisms are all costs associated with starting an online program. So too are maintenance, training, and record keeping. Reducing the amount of paper testing may offset some of these costs, however the investment required to develop an online program is substantial

### Dependability

One of the most important elements of a successful online assessment system, and also one of the greatest challenges, is "ensuring dependability. Computers and the internet do not always function as desired" (Linn, 2002, p. 7). When service is interrupted or slowed down, the impact may interfere with a student's test-taking session. Results of pilot tests indicate that sections, and even entire tests, have been lost due to inconsistency of hardware and telecommunications.

### Security

Security of student demographic and test information must be assured before school districts will approve of an online system. In addition, test items must also be reasonably well protected from theft or tampering. While these concerns are no different than those for traditional pencil-and-paper tests, different approaches to resolve them may be required.

### Testing Conditions

As with other standardized assessments, efforts must be made to ensure similar test conditions for all students. It is easy to see how students in a school with outdated computers and a slow Internet connection might be hindered in their test-taking efforts when compared to students with access to new, powerful computers and a high speed connections. While there is no evidence yet on the extent to which these conditions might affect test performance, they will need to be addressed as online systems are implemented on a larger scale.

### Technology Expertise

Another challenge to the efficacy of online assessment is the degree to which a test-taker is comfortable with computers. Keyboarding and familiarity with using a mouse are basic skills that could influence a student's performance when testing in an online format. While students are being exposed to technology earlier and earlier, it will be some time before schools can expect all students to come with basic computer skills. A bigger issue, according to some, is the degree to which students have or have not experienced an "integrated" curriculum. Problem-solving and self-directed learning are important aspects of learning with technology, and until students are familiar with this type of learning environment, online assessment may not give a true picture of their skills.

### Test Format

Numerous problems emerged with the presentation and function of online assessments. For example, students may have to scroll up and down to see an entire test question or graphic, making it difficult to visualize and organize the problem. Likewise, on-screen calculators can be cumbersome to use, causing students who pilot-tested online assessments to resort to hand-held calculators and pencil/paper to complete their work. Save functions, scroll bars, and forward/backward arrows are a few of the features that seem to work inconsistently. While test developers should be able to resolve these issues easily, they are a reminder that online assessment is still a developing technology.

## The Future of Online Assessment

Washington is one of several states, including South Dakota, Oregon, Virginia, Georgia, Idaho, and North Carolina, currently piloting and implementing online assessment projects. Brief descriptions of each states' efforts are provided to give a context for the work underway in Washington.

### South Dakota

South Dakota worked with EdVISION to develop a low-stakes online assessment system for measuring progress in several subject areas. The Dakota Assessment of

Content Standards (DACS) is intended to measure not only student progress in meeting the state content standards, but also to assist schools in aligning their curricula and improving student achievement. The assessments are adaptive, such that early responses determine the progression of questions for each student. In this way, all students are tested at the appropriate level. The DACS will be mandatory in grades three, six and ten in reading and math. The state is doing extensive pilot testing and plans to implement the system statewide in the near future. "South Dakota has been able to move so quickly because state government has spent the past several years putting the infrastructure in place. Every classroom is wired with a high-speed Internet connection, and the ratio of students to computers is almost 2 to 1" (Linn, 2002, p. 4).

### Oregon

The state of Oregon chose Vantage Learning to help develop an online assessment system that uses adaptive tests to assess student progress related to state benchmarks and the Certificate of Initial Mastery. The Technology Enhanced Student Assessment System (TESA) is provided by the state, which also reports scores and trains staff in using the assessment. As in South Dakota, extensive pilot-testing is being done, and it is expected that the assessments will be available to all schools in math, reading, language arts, social studies and science by the 2003-2004 school year.

### Virginia

Virginia's efforts to develop an online assessment system are part of a larger Technology Initiative, which calls for a 1:5 computer to student ratio, Internet-ready local area network capability in every school, and high-bandwidth capabilities for delivering instructional, remedial, and testing services. The SOL (Standards of Learning) is comprised of multiple-choice tests for high school math, science, social studies, and English reading, literature, and research. The intent of the assessment is to report progress in meeting achievement standards, to provide information to improve instructional programs, and to document the quality of public education (Linn, 2002).

### Georgia

Georgia, working with NCS (Pearson) plans to offer two assessment programs via the Web. Reading, math, English and language arts will be assessed in first through eighth grades, and science and social studies will also be assessed in grades three through eight. The tests are criterion-referenced and results will be used to improve instruction and promote accountability. The second component of Georgia's assessment system consists of eight, criterion-referenced, end-of-course tests in four high-school subjects: math, science, English, and social studies. The EOCTs will be used to promote accountability and to diagnose students' academic needs. Results will also be used to evaluate schools and will be part of the high school graduation requirement.

### Idaho

Working with the Northwest Evaluation Association, Idaho is developing an adaptive online assessment system. Pilot-testing of the assessment, to be used in grades two through nine, will begin in September, 2003. It is hoped that the system will provide data to improve instruction in math, reading and language arts, which will in turn improve accountability. The system is being developed so that students will receive their scores within 24 hours, while schools and districts will receive scores within 72 hours.

### North Carolina

North Carolina is using web-based assessment as an accommodation for students with disabilities. Available in 2000-2001 as a pilot, the North Carolina Computerized Adaptive Testing System (NCCATS) will assess reading and math in grades three through eight, and will provide a comprehensive test at grade ten. The tests are multiple choice and follow an adaptive format, such that when a student answers a specific test question correctly, the computers move them to more difficult questions. If the student answers a specific test question incorrectly, the computer program provides less difficult questions. This process continues until the system is able to determine the student's actual level of performance in a given subject area.

### Minnesota

The Bloomington Public Schools began using a web-based testing system in 1999, in part to "increase the efficiency of traditional paper-pencil testing" (Bloomington Public Schools, 2002, p. 1). The CALT (Computerized Achievement Level Testing) is based on the levels test system developed by the Northwest Evaluation Association and is a four-stage adaptive model. As described in the district's website, "tests are constructed in blocks, each one representing one-third of the total test . . . after the completion of the first block, the computer branches to the adjacent high or lower block based on the student score from block one. Currently, a criterion of 65 percent or more correct is required for an 'upward branch', while a score lower than this will result in an easier set of items in the second block. When the student finishes the second block, the computer calculates the results from the combined blocks, and then branches a final time" (Bloomington Public Schools, 2002, p. 2). All test information, including the test items, difficulty level, and appropriate grade level are stored in a Microsoft Access database. Students can stop during the test and continue at a later time, and scores are available to teachers immediately, as are various test reports.

### Next Steps

While schools are still in the early stages of exploring and testing online assessment models, there is general agreement among assessment experts that electronic learning and assessment are "inevitable" (Linn, 2002, p. 7). Randy Bennett of the Educational Testing Service states that, "A continuing concern with such [assessment] reinvention is whether schools (and students) are ready technologically. . . ." (2001,

p.10). "Certainly, with few exceptions, it would be impossible to deliver large-scale assessments via the Internet today. But the trend is clear: the infrastructure is quickly falling into place for Internet delivery of assessment to schools. . . . " Likewise, an advisor to investors in education-related businesses stated that "We may not exactly be ready for [statewide online testing] now, but this is the way we're going to be doing this in the future" (Trotter, 2001, p. 6). He noted that online assessment has been "embraced by the corporate and professional worlds [and] has helped testing companies improve and standardize their technologies. And that has largely resolved questions about the validity of test items and concerns about the security of online tests" (p.6). Randy Bennett sums up the need for electronic assessment:

> Although there has been much recent intellectual ferment and experimentation in educational assessment, the practice of large-scale testing is much the same today as it was 20 years ago. Most large-scale tests still serve only institutional purposes, are administered to big groups in single sittings on a few dates per year, make little use of new technology, and are premised on a psychological model that probably owes more to the behaviorism of the first half of this century than to the cognitive science of the current half. There are good reasons to believe that this situation is about to change (Bennett, 2001, p. 1).

Bennett goes on to document the necessity of innovation in global competition, where businesses need to respond quickly to shifting markets, productivity needs, and customer service. Technology, he claims, is at the heart of these innovations and the influence on education is clear. Policymakers and educators "are beginning to expect the same things from testing agencies that they get in everyday commerce: innovation, productivity (as reflected in competitive pricing), and customer service. Increasingly, they are expecting adaptation to diversity too" (Bennett, 1997, p.1). Bennett has proposed three "generations" of response to these forces. The first generation "lays the basic infrastructure for electronic testing. In the second generation, large-scale tests undergo qualitative change, but their purposes remain essentially the same. The last generation bridges a rethinking of the purposes and mechanisms of large-scale assessment" (p.2). He believes that **"** New technology will permit this transformation [of large-scale assessment] by allowing us to create tests that are more firmly grounded in conceptualizations of what one needs to know and be able to do to succeed in a domain; by making performance assessment practical and routine through the use of computer-based simulation, automatic item generation, and automated essay scoring; and by changing the ways in which we deliver, and the purposes for which we use, large-scale tests" (1999, p. 11). His conclusion is that "this evolution is going to happen; it's going to take awhile, but it's going to happen . . . [however] there are going to be setbacks [and] it is going to be costly. Simply giving assessments on computers in a school setting is a big jump. . . . The way to do make it happen is to do it in relatively small steps, not bite it all at once" (Trotter, 2001, p. 32).

### Summary

There is a general perception among assessment, education, and technology experts that online assessment will be an essential teaching and learning tool in the near future. Online assessment will help meet the demand for greater school and student accountability. It will also provide a means of more closely aligning a technology-rich curriculum with assessment, and will give teachers a way to identify and target students' academic deficiencies. And although there are presently a number of challenges inherent in making it all happen effectively and efficiently, testing companies and a handful of states are making significant progress in the development of online assessment systems.

## EVALUATION DESIGN

The evaluation design for the Online Classroom Assessment project evolved over the course of the year due to changes in the number of participants in the pilot process. The original evaluation plan was designed to analyze and summarize the experiences of the first cohort of 30 participants. Each of these 30 teachers was asked to submit three reflective journals regarding their participation in the project, one after each administration of the assessment (Fall, Winter, and Spring). In addition, the evaluation called for phone interviews with 15 of the teachers, observations in 15 participant classrooms, an online survey of all participants' experiences, and interviews with a sample of participating students. The evaluation also called for an analysis of achievement test data in the event that student scores could be accessed. These measures would have provided a substantial amount of qualitative and quantitative data for the evaluation. However inconsistent administration of the program and a high drop-out rate on the part of participating teachers ultimately limited the efficacy of this evaluation plan.

Nevertheless, a letter was sent to all first-year participants prior to the start of the 2001-2002 school year explaining the purpose and requirements of the evaluation. Approximately half of the teachers responded, and in most cases their responses indicated that they were no longer participating in the project. Some teachers moved to counseling or administrative positions, others were in buildings where remodeling made the use of technology extremely difficult, a few did not have the technology to support the Assess2Learn™ system, some found that the program did not meet their needs or was too time-consuming to use, and some had not finished the training. A second letter was sent at the end of September 2001 again outlining the evaluation and requirements for participating teachers. Responses indicated that only two teachers from the first cohort were actively involved in using the assessments.

Given the limited number of first-year participants, a revised evaluation plan called for analysis of data from both first and second-year participants, recognizing that second year teachers would not have as much experience using the assessments. In the end, a reduced version of the evaluation plan was used in an attempt to gather in-depth information from a limited number of participants. The lack of a current database of

participating teachers slowed implementation of the evaluation, however, and given the late start, it was not possible to observe classroom use of the assessments. Instead, data were gathered from ten teachers through phone interviews and written reflections during April and May 2002. Seven of these were second-year participants who were using the assessments and three were first-year participants who had used the assessments for at least one year but were not using them anymore. Teachers responded to a series of evaluation questions (See Appendix B). While the sample size is small, these responses did provide interesting and useful information about Washington's efforts to implement an online assessment system.

For the purpose of this evaluation, two general questions were addressed:

**Evaluation Question 1:** ***What impact does a system of online assessments have on instructional planning and teaching strategies?***

**Evaluation Question #2:** ***What practical issues need to be considered when assessing students in an online format?***.

# RESULTS

Information gathered from teachers regarding their experiences with the online assessment system suggests benefits to both teachers and students. In addition, teachers were generally pleased with the alignment between the online assessment, the EALRs and the WASL.

## Benefits for Teachers

Teachers cited several benefits to using the online assessment system. Most important was the formative function of the assessments that allowed test results to be used to modify instruction. Most teachers indicated that they used the assessments to identify specific areas of weakness or learning gaps among their students. Following this they formed small groups and targeted instruction to address these specific needs. Several teachers also used the online test results to plan whole group instruction. For example, a class performing well in Geometric Sense could be moved ahead to another concept or skill. Likewise, if a class as a whole showed deficiencies in Number Sense, more time could be allotted to that area. And since questions are keyed to specific EALRs, "teachers can analyze and adjust instruction according to information revealed by the tests."

In addition to using assessment results to modify instruction for individual students or for the class as a whole, teachers believe that the tests will be useful in identifying trends in student learning. While these teachers have not used the system long enough to realize this benefit, they nevertheless expect it to be an important feature of the assessments. As several noted, the ability to see changes in student scores over the course

of a school year would give teachers valuable insight into their academic growth, as strengths and weaknesses become easily identifiable.

Teachers also used the assessments to develop learning goals with their students. Having immediate access to test scores allows teachers to actively involve students in identifying areas of weakness and planning for improvement, thereby establishing an important learning partnership. Teachers felt that students were empowered by their involvement in this process, and anticipated more would reach their goals because of this participation.

Nearly every teacher who responded to questions about the Assess2Learn™ system agreed that one of the most important features of the program was the variety of reports that can be generated. Both the format and content of the reports were useful to teachers, parents, and students. Particularly informative were those that allowed analysis of errors and those that helped teachers identify deficiencies around specific EALRs.

Finally, a few teachers discussed the benefits of having access to different levels of the online assessment. A fourth grade teacher, for example, might find that a third grade assessment to be more appropriate in measuring the abilities of a less capable student. In a sense, the assessments can be individualized to the ability levels of each student, providing important diagnostic feedback.

## Benefits for Students

Not only did teachers find the online assessment system useful for meeting their instructional needs, but they discussed benefits for students as well. First, teachers felt that it was important for students to participate in a "test-like" event without the anxiety inherent in the WASL. As one teacher remarked, "The assessments provide a chance for our students to experience a new testing situation without the high-stakes pressure of the formal tests." In addition, students are generally motivated to use computers, and although the online assessments were still tests, students approached the experience with less trepidation than they did typical pencil and paper tests.

Unlike the WASL and other pencil-paper standardized tests, results from the online assessments were available immediately, or as soon as a teacher was able to complete the grading process. This is invaluable for students, since they were able to see their results and do an analysis of their errors. In this regard, these teachers' comments supported claims made by testing companies and state policymakers that suggest that the immediacy of feedback will be one of the most important contributions of electronic testing. Similarly, some teachers suggested that the online assessments gave insight into effective test-taking strategies. For example, when students went back over the questions they saw more clearly *why* they had made some of their mistakes. The quick turnaround of test results facilitated an analysis of errors.

The online assessment also allowed teachers to share examples of exemplary responses. For example, when students were involved in the grading process and in the

analysis of their results, they had the opportunity to see "4" responses, a valuable learning opportunity.

## Relationship of Online Assessment System and EALRs and WASL

Teachers found the online assessment system to be reasonably well aligned with the state standards. While the online assessment was not as comprehensive nor as in-depth as the WASL, the various strands were representative, and provided students with an opportunity to participate in a "WASL-like" situation. For example, the third grade online reading assessment consists of Fiction, Non-fiction, and Poetry sections, much the same as the WASL. While most teachers felt the level of difficulty of online items was similar to the WASL, one teacher disagreed: "Frankly, the questions were nowhere as difficult (logical thinking) as the WASL...I was not always convinced that those [students] who passed a particular item would necessarily pass it on the WASL, where a skill is often embedded in an application that is much more difficult to figure out."

All in all, teachers who used the assessments were positive about the benefits, both for themselves and for their students. Online assessments are aligned with the EALRs and the WASL. In addition, they appear to provide useful diagnostic information for teachers in addressing student strengths and weaknesses, and should allow them to see changes in student learning over time.

While teachers who used the online assessments were generally positive about the benefits, they also discussed a number of practical issues that influenced the ease with which they were able to use the system. These are classified as 1) Technical issues 2) Student considerations and 3) Teacher considerations.

## Technical Issues

The most frequently mentioned challenge encountered by teachers attempting to use Assess2Learn™ was related to the technology itself. Teachers with Macintosh computers, either in the classroom or in the school computer lab, struggled to make the system work. Despite efforts on the part of the test developer, some teachers ultimately quit using the assessments because of continuing compatibility issues. The following example illustrates this challenge. "Last spring, the online tests worked well. The most current version of the test does not work reliably on our Macs. The Riverside team visited our school to see the problems first-hand, but we have not heard about a solution . . . there seem to be the same old problems and some new ones to boot! The Macs will work sometimes. If we put student B on the same computer, taking the same test right after student A has successfully completed a test, there is no guarantee it will work a second time. It is very quirky. We have one PC and haven't even had consistent luck with it. . . .It is evident that the bugs are alive and kicking. . . ."

Another teacher was more frustrated with the experience: "Trying to administer the online test was a disaster. Either the school sever was down, or Riverside's server was down, and then the students would lose their tests. Furthermore, our school doesn't

have an up-to-date computer lab. The last time I attempted to give the test, about half of the computers were not functioning. This project was an absolute nightmare and I did not want to attempt to do it again."

Smaller issues were troubling as well, and while many of these glitches were addressed during the year, some continued to frustrate teachers and students. In some cases the way in which a student used the system was noticeably different from their typical computer use. For example, scrolling with the roller ball on the mouse caused answers to be erased. While an older student might easily adjust to such a difference, it was more difficult for third and fourth graders to negotiate such idiosyncrasies.

## Student Considerations

Several factors influenced the ability of students to use the online assessment system effectively. As mentioned previously, students are usually highly motivated when computer technology is integrated into the classroom, and online assessments were no exception. Being able to take a test on the computer was less intimidating than taking a traditional paper and pencil test. On the other hand, teachers identified several factors that limited student performance when testing on the computer. First, not surprisingly, was a student's keyboarding skill. Students were more successful if they were proficient at keyboarding. While teachers are still gathering information, it appears at this stage that a lack of keyboarding skills could significantly limit the quantity, the creativity, and the flow of a student's online writing. Perhaps because so much of their concentration is focused on fine motor skills rather than on content, their responses generally did not seem to accurately reflect their abilities. Research on writing with computers suggests that the more students write with computers, the greater their need to be tested this way. On the other hand, students' efforts may be hampered if they have little skill or experience with the keyboarding.

## Teacher Considerations

When asked to consider the practical aspects of the online assessment system, teachers commented on the actual administration of the tests, the scoring procedures, the relevance of results, and the potential for school-wide use.

### Administration and Scoring of the Online Assessments

Once teachers worked through any compatibility issues and technical glitches, they were able to administer the assessments with relative ease. Student information needed to be entered into the system, and from there it was primarily a matter of getting the appropriate test ready for each student. Teachers' experiences in scoring the assessments varied. Most felt that the time necessary to score the writing component was overwhelming, and given the fact student writing was compromised by a lack of keyboarding skills, they generally agreed that this component of the test that was not as useful as the others. The following remarks were typical:

> We do not believe the writing portion of the test is practical for third graders. Their keyboarding skills are not yet at a level that allows them to quickly type or easily edit their stories. Writing and editing a story on the computer is an entirely different process than the one students experience taking the WASL.
>
> I had the most problems trying to get in and score the open-ended items. The time factor to score was too high . . . even though it is good for students to experience the need to write out their answers, I think that this may not be the format to do it in. Eliminating the open-ended questions might make the test quicker for the students to take and easier for the teachers to manage.
>
> Writing is easier to do with pencil and paper. I like the prompts and use them for the kids, but I have them do it with paper and pencil.

And from a high school teacher:

> I am concerned about the amount of time that it will take to administer this assessment three times during the year. As a writing teacher (primarily), I found last year that the online assessment didn't give me any additional feedback on a student's writing skills and progress than what I was already receiving by assessing their in-classroom work. Since the writing assessment is just that, writing, it doesn't enhance my knowledge of student performance any more than a non-electronic assessment, but takes up a significant amount of time that could be spend on instruction and coaching . . . if the purpose of this assessment is formative and is intended to inform classroom instruction, then I am not sure it is the most efficient means to do so.

Responses regarding the written response items in the reading component of the assessment were similar. On the other hand, teachers were enthusiastic about the math section of the assessment and the objective reading items. They felt that the content and scoring provided useful diagnostic information in a reasonable amount of time. As one teacher noted, "To get this kind of diagnostic information in a traditional testing format would take much, much longer."

When asked what changes and/or modifications would allow them to use the assessment results more effectively, teachers had a number of suggestions, which include:

- Designing the assessment system so that it mirrors Office applications as closely as possible. This way, students (especially younger students) would not have to learn different procedures.

- Fixing the glitches that cause student tests to disappear. While students are enthusiastic about taking tests on the computer, they nevertheless find it frustrating to put time and effort into the task only to have their answers lost because of a technical glitch.

- Designing or improving an "autosave" function, so that if a child leaves the test temporarily she/he does not lose everything.

- Developing a way for teachers to quickly correlate errors with the specific strand questions.

- Designing reports that allow teachers to automatically see a child's performance over time, both over the school year and across grades.

- Developing a way to easily load all students' names at one time.

- Designing a comprehensive spreadsheet that presents an item analysis for each individual student in the context of the overall class performance.

Practically, then, the most challenging aspect of the online assessment was the technology, specifically the compatibility of the Assess2Learn™ system and Macintosh computers. Older, slower computers were also problematic, as were inconsistent servers. These challenges are similar to those reported by users of online assessments in other states across the country. Washington teachers were not convinced of the value of the writing component of the test, but found the reading and math assessments to be user-friendly, both for themselves and for their students.

## Additional Findings

Analysis of qualitative data provided information not directly related to the evaluation questions, but that nevertheless is important in clarifying the efficacy of the online assessment system. For example, teachers were positive about the training they received to prepare them to use the online assessments. In addition to learning how to use the specific assessments, several felt the training had improved their understanding of the WASL and the way it is scored. Because of this they feel better equipped to prepare their students for the WASL experience, whether or not they continue using the online assessments. One teacher reflected that the training was " . . . incredible. I have a much better idea of how to prepare kids for the WASL. . . .it is good to involve them in the grading process."

A second finding concerns the ease with which the online assessment system might be expanded at the building level. Teachers were generally in agreement that once compatibility issues were resolved it would be a fairly simple matter to train other teachers to use the assessments. Teachers need not be highly skilled in using computers to learn how to use the system and incorporate the results into their classroom planning. In fact, one teacher noted that she had already trained two other teachers in the building

who "would admit themselves that they were not necessarily competent users of technology prior to this exposure, and they are using it effectively now." This is an important consideration given the expectation that use of the system will increase over the next several years.

# CONCLUSION AND RECOMMENDATIONS

Washington is one of several states exploring the benefits and challenges of using web-based assessments at the classroom level. The goal in developing an online assessment system is to provide teachers with a diagnostic and assessment tool they can use to make adjustments in instructional strategies to align more closely with the EALRs, benchmarks, and frameworks. Funded by the Bill & Melinda Gates Foundation, a two-year pilot test of the system was administered by the Office of the Superintendent of Public Instruction (OSPI). During the first year of the project, grades three, six, and nine were given training and materials to assist them in using Assess2Learn™, and grades four, five, seven, and eight were brought into the program during the second year. Administration of the program was problematic and may have limited progress during the first two years. Approximately 60 teachers were involved in the training, although fewer than ten actually appeared to be using it at the end of the second year.

Evaluation results revealed several benefits of the assessment system, according to teachers. Information gathered from the assessments was used to identify learning strengths and weaknesses, and allowed teachers to modify and adjust instruction to meet specific individual and classroom needs. In addition, immediate feedback proved helpful in the analysis of errors and academic goal-setting. Teachers also predict that the system will be valuable in looking at student performance over time to see learning trends.

Challenges were numerous, and included technical glitches, compatibility issues, inconsistent servers, and formatting problems. These challenges are similar to those experienced by teachers in other states using different assessment systems, which is not surprising since the development of electronic assessments is still in its early stages. However the potential of online assessment outweighs the problems, and it seems likely that work on these systems will continue. Although the first two years of piloting Assess2Learn™ in Washington have been difficult for a number of reasons, lessons have been learned that will significantly improve future implementations. Early results suggest that the Assess2Learn™ system can be, in time, a valuable diagnostic tool for teachers in the state.

## Recommendations

- Consider limiting participation to a small number of teachers and classrooms until all of the technical problems are resolved. As noted by one assessment professional, "Experimentation should occur in a few schools with the best principals and teachers, involving these individuals as collaborators. It they understand they are defining the future of the state's educational assessment system, they will work hard to help. If early ideas do not work under these highly

favorable conditions, they surely will not work statewide and should be reconsidered" (Linn, 2002, p.8).

- If the project is to be expanded to a large number of classrooms, considerable technical support for those classrooms should be provided until the major technical problems have been resolved.

- Ensure that whoever is selected to administer the project has the time, interest, and organizational skills necessary to direct and support an innovative program such as Assess2Learn™.

- Resolve compatibility issues. If Mac users cannot implement the system, it will be a difficult sell in Washington State given the high number of schools and districts that continue to operate Mac platforms.

- Re-evaluate the writing component. As it stands, teachers do not see it as an efficient component of the system.

- Make design changes in the system that will allow even young children to easily navigate it. Mirror what they already know (Office) as closely as possible.

- Continue the training sessions, including the opportunities for teachers to participate in item selection and item analysis. The TLP training model can serve as a design for using time efficiently.

- Once the system is operating dependably, plan an agenda of presentations to education groups across the state.

# REFERENCES


Bennett, R.E. (1998). *Reinventing assessment: Speculations on the future of large-scale educational testing.* New Jersey: Educational Testing Service Policy Information Center. Available at: www.ets.org/research/pic/bennett.html

Bennett, R.E. (1999). Using new technology to improve assessment. *Educational Measurement: Issues and Practice . 18* (3), 5-12.

Bennett, R.E. (2001). *How the internet will help large-scale assessment reinvent itself. Education Policy Analysis Archives, 9* (5). Available at: http://epaa.asu.edu/epaa/v9n5.html

Bloomington Public Schools (2002). *CALT description: Designing a computerized achievement level test (CALT) system*. http://bloomington.k12.mn.us

Bransford, J.D., Brown, A.L., & Cocking, R.R. (Eds) (1999). *How people learn: Brain, mind, experience, and school.* Washington D.C.: National Academy Press.

Linn, D. (2002). *Using electronic assessment to measure student performance.* NGA Center for Best Practice: Education Policy Studies Division

Means, B. (Ed.) (1994). *Technology and education reform. The reality behind the promise*. San Francisco: Jossey-Bass Publishers.

Meyer, L. (2001). New challenges. *Education Week*, *XX* (35), p. 49-54.

Olson, L. (2002). Idaho to adopt 'adaptive' online state testing. *Education Week*, January 23, 2002.

Oregon Department of Education. "TESA FAQ" available at: http://www.ode.state.or.us/asmt/aboutasmt/faqs/faqtesa.htm

Russell, M. & Haney, W. (2000). Bridging the gap between testing and technology in schools. *Education Policy Analysis Archives,* v8 (19). 1-10. Available at: http://epaa.asu.edu/epaa/v8n19.html

Russell, M & Plati, T. (2002). Effects of computer versus paper administration of a state-mandated writing assessment. *Teachers College Record.* Available at: http://www.tcrecord.org/content.asp?contentid=10709

North Carolina Public Schools (March 21, 2001). Preliminary information about computerized adaptive testing system. *Assessment Brief, 7*(7). Available at: http://www.ncpublicschools.org/accountability/testing/briefs/computerizedadaptive/

Shaw D.E. & the President's Committee of Advisors on Science and Technology Panel on Educational Technology (1998). Report to the President on the use of technology to strengthen K-12 education in the United States: Findings related to research and evaluation. *Journal of Science Education and Technology, 7*(2) 115-126.

South Dakota Department of Education and Cultural Affairs, "DACS [Dakota Assessment of content Standards]: DACS FAQs." Available at: http://www.state.sd.us/deca/DACS/faqs.htm

Trotter, A. (2001). Testing computerized exams. *Education Week*. *XX* (37), p. 30=35.

Trotter, A. (2001). Testing firms see future market in online assessment. *Education Week*, *XX* (40), p. 6.

United States Department of Education (1993). *Using technology to support technology reform.* Washington D.C.: Author. Available at: http://www.ed.gov/pubs/EdReformStudies/TechReforms/

Virginia Department of Education, "General Questions" available at: http://www.pen.k12.va.us/VDOE/Technology/soltech/FAQ/FAQ.html

# Appendix A Online Classroom Assessment Training Schedule

## Online Assessment Training Schedule

| | |
|---|---|
| October 6, 2000 | Nomination of teachers grades 3,6, and 9 |
| October 18-19 | Item review and selection (3,6, and 9) |
| November 14 | Online item review and finalization (3,6, and 9) |
| November 28-29 | open-ended item scoring, technology training and classroom deployment tips (3,6, and 9) |
| December 1-13 | Teachers pilot the assessment (3,6, and 9) |
| December 14 | Feedback session (Teleconference) |
| January 31, 2001 | Nomination of teachers grades 4,5,7, and 8 |
| February 27-28 | Item development and selection (3,6, and 9) |
| March 28-29 | Constructed response scoring training (3,6, and 9) |
| May 30-31 | Item development and selection (4,5,7 and 8) |
| September, 2001 | 3,6, and 9 ready to use Fall, winter and spring assessments |
| September 19-20 | Assessment review and constructed response scoring training (4,5,7 and 8) |
| November 6-7 | Item development training and item selection (4,5,7 and 8) |
| December 12-13 | Assessment review and constructed response scoring training (4,5,7 and 8) |
| December 2001-June 2002 | All pilot classrooms grades 3-9 using the assessments |

# Appendix B Reflective Questions for Online Classroom Assessment Participants

## Reflective Questions for Online Assessment Participants

1. Have you used the online assessments with your students? (When?)
   If so, did you find them useful?
   If not, why not?

2. What worked well in administering the most recent assessment? What challenges did you encounter? How might these challenges be addressed?

3. How relevant and useful are the online assessments in addressing the EALRs?

4. Are there any resources that would enable you to better utilize the information you get from the assessments?

5. How well did the training prepare you to use the Assess2Learn™ in your classroom?