Assessing ICT Literacy of Students

By Jon Haber, SkillCheck President and CEO

Abstract

Whether one is assessing Internet and Computing Technology (ICT) literacy as part of the No Child Left Behind (NCLB) 8th grade computer-literacy requirement, or assessing other grade-level students on a school, district or statewide basis, educators face similar challenges. This document discusses some of the issues surrounding the successful implementation of large-scale assessments for student ICT literacy. A more detailed discussion of these topics by the author will appear in ISTE's *NETS*S Resources for Assessment*, to be published in January 2005.

Introduction

Section 2402; Title II; Part (b) (2) (A) of the US Department of Education's No Child Left Behind Act (NCLB) of 2001 stipulates that by a specified date, every student shall demonstrate technology literacy by the end of the 8th grade:

> "To assist every student in crossing the digital divide by ensuring that every student is technologically literate by the time the student finishes the eighth grade, regardless of the student's race, ethnicity, gender, family income, geographic location, or disability."

In an effort to ensure that this requirement of NCLB is met, departments of education in most of the 50 US states have either published technology standards for 8th grade education (analogous to language and math standards that are being used to meet accountability requirements of NCLB for these academic areas), or have embraced emerging nation-wide standards, notably the National Educational Technology Standards for Students (NETS*S), published by the International Society for Technology in Education (ISTE).

Many of these standards (notably, the ISTE standards and US state standards influenced by ISTE's work on NETS*S) have published performance indicators for specific grade levels, several of which include general and specific indications of the knowledge, skills and abilities required for technology literacy at the 8th grade.

While educational standards exist within most of the 50 US states, and additional standards are available both nationally and internationally, the means to assess student performance based on these standards has only been implemented in a few locations.

Educators exploring the implementation of an Internet and Computing Technology literacy assessment program need to make a number of decisions on important standards, testing and implementation issues. While this paper makes reference to the 8th grade ICT literacy requirement of NCLB, the issues discussed in this document are relevant to any ICT literacy assessment project at any grade level.

Selecting a Standard: Local or National/International

Every state that has created an ICT literacy curriculum has taken the first step in creating a means of assessing student progress since the first question that emerges when contemplating assessing a standard is "upon what standard should an assessment be based?"

Basing a state assessment on a state's own standards seems a logical choice, one that will ensure that an assessment is built on the indicators already chosen as being relevant to a state teaching curriculum.

A state-level approach to such a project must deal with the fact that creating a high-quality, robust, valid, universally available assessment can be a time-consuming and costly process. While some states may have the resources to commit to such a development effort, another alternative is to turn to an existing assessment tool that is built on a *consensus standard*.

A consensus standard is an educational standard that may be based on, influenced or guided by the curriculum work done by one or more state, national or international organizations dedicated to creating ICT literacy standards. To a certain extent, the ISTE NETS standards represent high-level consensus standards for students, teachers and school administrators that have been influenced by and have served to influence the development of state ICT curricula in over 40 US states.

With the growing interest in 8th grade ICT assessment due to NCLB, commercial vendors (and, eventually, non-profits, consortia of states and other organizations) have created or are creating assessment solutions for measuring ICT literacy for 8th graders and students at other grade levels. Most of these products are created using a test development process that requires the creation of a consensus standard.

To create a consensus standard upon which an assessment can be built, test developers may synthesize existing standards such as the ISTE NETS*S standards and grade-level indicators, similar standard information from some or all of the US states, and potentially work outside of the US (for example, the UK's eSkills guidelines for ICT literacy).

In determining if the benefits of working with a consensus standard (the ability to take advantage of existing high-quality assessment products, rather than building one for each state) outweigh the disadvantages (notably, assessing students based on a standard that may not touch on every point covered in a state-level standard), the following factors should be considered:

- How closely does the consensus standard match the overall structure and goals, as well as the particular details of a state-level standard? Given that many state student standards are already built upon a common base (ISTE NETS*S), the difference between a consensus standard and a state's own standard may not be significant enough to rule out use of an assessment based on a consensus-based design.

- What type of assessment has been built on the consensus standard, and does it meet the requirements of educators (vs. a product custom created for a particular state-level ICT assessment project)?
- What is the cost of a consensus-standard based product vs. the cost of creating an assessment locally built on local standards?

Regardless of whether one chooses to base an assessment project on a home-grown local standard, or on a broader consensus standard that fits state requirements, the test development processes outlined below need to be followed. In the case of a local effort, those processes will be followed by the development organization working with a state to create a custom assessment. In the case of a consensus-standard product, the exam developers working on that assessment should perform their development based on the industry-standard processes outlined below.

Assessment Planning

Once a curriculum (state or consensus) has been chosen, professional assessment development goes through the following steps:

- Assessment Planning
- Item Writing/Assessment Automation
- Validation

Planning

Assessment planning normally starts with the creation of two critical documents:

- A test blueprint which organizes the information to be covered in the assessment
- A test construction plan that spells out all factors other than content that will go into the design of the assessment (such as length of time for the assessment, delivery and grading methods, etc.)

Assessment Blueprint

An assessment blueprint organizes the content to be covered by an assessment in a way that allows that content to be reviewed by subject matter experts and analyzed statistically during the process of test validation. An example of part of an assessment blueprint for an eighth grade IT literacy blueprint appears on the following page.

Domain 1.0 Basic Operations and Concepts

This domain includes the knowledge and skills required to demonstrate a sound understanding of the nature and operation of technology systems in order to be proficient users of information technology. Concepts covered in this domain include fundamentals of computer hardware, computer software and networks as well as how to identify and solve routine computer hardware and software problems. Terminology and concepts are grade appropriate for an 8th grade assessment.

Content Limits

1.1 Identify the purpose and function of computer hardware

Content may include the following:

1.1.1 Identify terminology related to computer hardware, including:
- Microprocessor/Central processing unit (CPU)
- Monitor
- Keyboard
- Mouse
- Printer
- Scanner
- Disk drive
- Hard drive
- CD ROM
- Memory (including RAM, ROM and disk storage)

1.1.2 Identify different types of computers and the purposes/best uses of each computer type, including:
- Mainframe computer
- Minicomputer
- Microcomputer/Personal computer
- Laptop/notebook computer
- Handheld computer/Personal Digital Assistant (PDA)

1.1.3 Identify parts of a computer by how they appear in a diagram of computer components, including:
- CPU box/case
- Mouse
- Keyboard
- Microphone
- Monitor
- Printer
- Speakers
- Floppy diskette drive
- CD ROM drive

1.1.4 Identify the important role of the central processing unit as the "brain" of the computer

While a test blueprint can be organized in a number of ways, an outline format is often the best means to present information that will map content be covered in an assessment. In the example on the previous page, the content is organized into:

- **Domains** - The high-level areas covered by the assessment. In the example, the first domain is Basic Operations and Concepts (of computers). In this case, the domains of the assessment are based on the six ISTE NETS*S standards, although other organizing frameworks can serve equally as well.

- **Sub-domains** - A breakdown of the domain into logical categories that further define the content covered in the domain. For example, a Basic Operations and Concepts domain (Domain 1) can include identifying the purpose of computer hardware (sub-domain 1.1), software (sub-domain 1.2) and networks (sub-domain 1.3).

- **Objectives** - One or more measurable elements that fall into the specified domain/sub-domain. Each item in the third level of the outline ("1.1.1 - Identify terminology related to computer hardware," "1.1.2 - Identify different types of computers and the purposes/best uses of each computer type") represents an objective.

In many ways, a test blueprint resembles a class curriculum or educational standard created for subjects such as English, mathematics or technology, and in many cases a test blueprint can serve as the basis for creating a curriculum or vice versa. However, there are certain elements unique to a test blueprint that makes it distinct from other types of educational standards documents, notably:

- Test objectives need to all be measurable. While educational curricula can include items to be taught that may never be assessed, an assessment blueprint needs to consist of objectives that can be measured in some way.

- The organization of content into domains, sub-domains and objectives needs to be arranged in such a way that a blueprint can be used for the basis of analysis for content validation (described on page 9). This may involve organizing objectives into groups differently than one would when organizing the same information as the basis of a class curriculum or academic teaching standard.

While the goals of a test blueprint may be different than those of a class curriculum or standard, the process of standards development is often improved by taking into account the need to both teach and assess to a certain standard. Thinking about teaching goals when creating a test blueprint, or understanding ultimate assessment goals when developing a curricula and academic standards, tends to improve the rigor and long-term usefulness of both development efforts.

Once completed, a test blueprint can be used as the basis for content validation described on page 9. This helps determine the balance of content from different domains required to create a valid assessment.

Test Construction Plan

A number of factors go into designing an assessment that go beyond content to be covered. For example:

- The length of time of the assessment
- The method of delivery (paper, computer, Internet, etc.)
- Testing methodology (fixed form, adaptive, etc.)
- Number of test forms
- The nature of test grading (human scored or machine scored)
- Scalability issues
- Commercial issues (price, delivery channel, etc.).

For example, a standardized college entrance exam, like the SAT or ACT, needs to fit a required test time, and needs to be delivered on a large scale in a wide variety of environments (classrooms, hotel conference rooms, etc.). Because of these constraints, a test construction plan for this type of exam might indicate that paper delivery is critical to ensure access to the test from any location. Such a plan would also specify the number of questions to be included in the test (to fit time requirements). Since these exams are often given simultaneously across the country, the use of a single exam form (as opposed to multiple versions of the exam) may be sufficient to minimize cheating.

One important area where test design must be balanced with the content requirements indicated in a test blueprint is the area of comprehensiveness. Most test blueprints cover all of the knowledge, skills and abilities specified as making up a specific content area (such as IT skills, English and mathematical ability). However, most test construction plans will indicate some kind of constraint (often time) that prevents a test from comprehensively covering all domains, sub-domains and objectives. A test construction plan should take into consideration the practical issues regarding how an assessment will be created, delivered to the desired audience and graded, and how scoring information will ultimately be used (in reports to the student or in consolidated data provided to the school, district and state).

Taken together, a test blueprint (in which a content analysis specifies the balance of items between different domains and sub-domains) and a test construction plan that specifies the practical assessment requirements (such as length of the exam which dictates the number of questions that can be included) help provide test developers with a map of how many items are needed to create the assessment. Performing this research before item writing begins helps minimize the creation of test items that may be unnecessary for a particular assessment. Similarly, understanding delivery requirements and other practical considerations may point to which kind of items are required in an assessment. For example, the need to deliver paper exams to huge numbers points to an assessment that will primarily consist of linear test items (multiple-choice, etc.) as opposed to other item types described in the following section.

Item Writing

The planning documents described in the previous section (a test blueprint analyzed for content balance and a test construction plan) provide test developers direction as to what objectives need to be covered and what type of items can be used to measure those objectives. For example, if the test construction plan indicates that a test will be in a written format, item writers will be required to create items that can be reproduced on the printed page (such as multiple-choice items) based on objectives set forth in the blueprint.

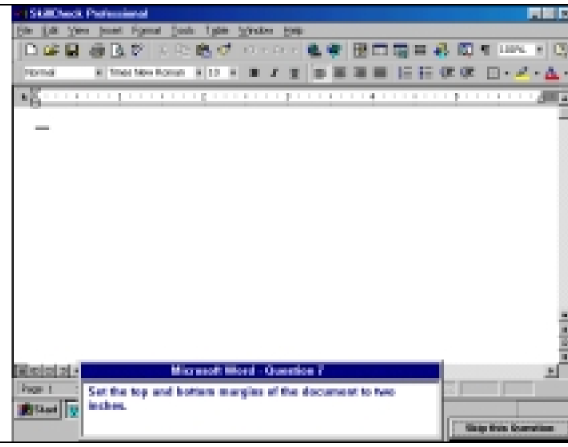
The types of test items that are available for producing an ICT literacy assessment include:

- Linear test items - This includes traditional item types such as multiple choice, multiple response (multiple choice with more than one correct answer), true-false or matching.
- Open ended items - This includes fill-in-the-blank, short-answer or essay
- Performance based items - Items which place a test taker in a real or simulated environment (such as a real or simulated version of an application such as Microsoft Word) and asks the test taker to perform a specific task

The strengths and weaknesses of each type of test item appears in the table below:

Item Type	Strengths	Weaknesses
Linear Test Items	• Can be delivered on paper, computer or over the Internet • Relatively easy to automate for software-based or Internet deployment • Objective items (each with correct or incorrect responses) are easy to grade and/or automate scoring	While good for testing knowledge, linear test items are less well suited for testing skills (see example below)
Open Ended Test Items	• Can be delivered on paper, computer or over the Internet • Relatively easy to automate for software-based or Internet deployment • Can cover both objective and subjective subject matter	While allowing for more versatility than "closed" linear items in which candidates can only select from provided responses, open-ended items are more difficult to score, often requiring human graders who must be trained to score open-ended items consistently.
Performance-Based Items	• Provide a more fair and accurate means of assessing skills • Objective performance-based items can be graded by computer • High reliability • High satisfaction level from test takers	Performance-based items are more difficult to automate than linear test items, requiring either significant work from software developers to simulate or control applications being tested, or significant manual process to grade work samples created using a mechanical type of performance-based assessment.

As a rule of thumb, linear items are best suited for assessing knowledge, while performance-based items are best for assessing skills. For example, consider an objective asking a test taker to demonstrate their ability to set margins using a word processor such as Microsoft Word. Such an objective can be measured using either a linear or performance-based test item:

Linear Item	**Performance-Based Item**
Which of the following commands will allow you to change margins? A. From the File menu, select Margins. B. From the File menu, select Page Setup C. From the Format menu, select Margins. D. From the Format menu, select Paragraph.	
This traditional linear test item (utilizing a multiple-choice question format) is testing just one component of the objective being measured, notably which menu choice contains an option to change margins. The question is not put into the context of the software program, and does not ask examinees to perform a complete software task.	This performance-based testing item, in contrast, places examinees in a perfect simulation of the software environment and asks them to perform the task from start to finish in any correct way the software allows. Only a fully interactive performance-based item allows one to test all components of the objective in a single test item.

A performance-based item can be created using different technologies. These include:

- Interactive simulations which emulates the features of a real software product such as Microsoft Word and monitors the test taker's progress through the simulation to determine if he/she has answered a question correctly.
- Concurrent assessments that place a candidate into the real application and asks him or her to perform a specific task, determining at the end of the task (often by checking the final state of the program after the task is completed) if the task has been performed correctly.
- Mechanical assessments in which a candidate is provided directions (usually in the form an exercise) at the end of which a human grader reviews a work product (such as a final document file) to ensure that all steps have been taken correctly (checking margins in the final document, etc.)

There are strengths and weaknesses associated with each approach to performance-based testing. Simulations, for example, provide more effective control over the testing environment and allow for easier deployment over networks and the Internet, but often lack access to the number of features available in a concurrent product. Concurrent tests can provide access to more working program elements than simulations, but face significant hurdles in deployment in complex environments (since they require the real application software be present on the each machine being used for testing). Mechanical assessments also require access to the real software being tested. In addition, the manual process of grading a mechanical test can be time consuming and costly when implemented on a large scale.

For purposes of creating an ICT literacy assessment for 8th graders (and other grade levels), the objectives of such a test will most likely cover both knowledge (such as an understanding of computer hardware, understanding of copyright and plagiary issues, etc.) and skills (such as how to use specific software such as an operating system, word processor and browser program). As such, an assessment covering this mix of knowledge and skill-based objectives would most likely include both linear and performance-based test items.

Validation

Validation is the collection and analysis of information to demonstrate that a test is successfully measuring what it purports to measure.

There are a variety of methods one can use to demonstrate test validity – the following being the most frequently used:

Content Validity

Proof, normally provided by subject-matter experts, that items in a test cover the most important and frequently used knowledge, skills and/or abilities needed to accomplish the job being measured by the test. For example, a test on Microsoft Word can be said to be content valid if it can be proven that the subjects covered in the test (setting margins, printing, etc.) represent the most important and frequently used functions critical to the success of a Microsoft Word user.

Construct Validity

Proof that the individual items in a test are accurate measurements of the subject being tested. For example, an item that asks an applicant to set margins within a performance-based simulation of Word or a concurrent test that uses the real Word program use a near-perfect construct to measure someone's ability to perform a task (in this case, set margins). A multiple-choice question that asks: "Which menu contains an option to change margins?" is a less exact construct for measuring the same function. In some cases, construct refers to the nature of the test itself. For example, a college entrance exam that tests vocabulary and mathematical knowledge and skills is built on the construct that these skills are an accurate predictor of success in higher education.

Criterion Validity
Proof that the overall test accurately correlates with some other independent measure. For example, assessing a test taker's on-the-job performance with Word one month after he or she has taken a Word test and seeing if test scores accurately predicted job performance is one type of criterion validation method known as a **predictive** study. Alternatively, a test can be administered to people whose job skill is already known and test scores can be correlated with an independent measure of those skills. This type of validation is known as a **concurrent** study. In both cases of criterion validation (predictive and concurrent), it is important that independent assessments be generated blind to test scores.

Reliability
Reliability, another important component of test validation, is the collection of evidence to demonstrate that a test provides consistent measurement of scores. For example, a scientific scale that gives different readings of an object's weight each time it is used is likely to have a low reliability (possibly due to a mechanical defect). Similarly, a hands-on driving test may be unreliable if all human evaluators that score candidate's driving ability are not trained to measure performance according to the same, objective standard. While tests can be evaluated for reliability by giving the same test to the same people at different times to determine if measurement is consistent (called *test-retest reliability*), there are a variety of statistical techniques (including *split-half reliability* and *internal consistency*) that measure test reliability by comparing the performance of test items within a test to one another.

Methods of Determining Test Validity
There are a variety of techniques that can be used to determine each type of test validity. For example, a focus group of subject matter experts can be convened to create and review a test blueprint and evaluate each domain, sub-domain and objective based on a quantitative scale, often using a carefully constructed survey tool that allows each blueprint component to be assigned a rating based on one or more five-point scales. Similarly, such a survey can be presented to a wide body of subject matter experts (possibly online) to collect more information than can be gathered within a small focus group.

Below is an example of the type of survey instrument that might be used to evaluate a test objective:

Domain 1.0 Computer Hardware

This domain includes the knowledge and skills required to identify different types of computers and computing devices, the components of a personal computer (including internal components such as microprocessors) and how these components function and interact. The domain also includes the knowledge and skills relating to computer storage, performance and maintenance procedures.

Content Limits

1.1 Identify types of computers, how they process information and how individual computers interact with other computing systems and devices

Content may include the following:

1.1.1 Categorize types of computers based on their size, power and purpose, including:

- Supercomputers
- Mainframe computers
- Minicomputers
- Microcomputers
- Laptop computers
- Handheld computers/Personal Digital Assistants (PDAs)

How important is this objective?

Importance
1 = Not important
2 = Of little importance
3 = Of modest importance
4 = Very important
5 = Critically important

How frequently do you make use of the knowledge or skills listed in this objective?

Frequency
1 = Never
2 = Rarely
3 = Often
4 = Very often
5 = Always

The results of such a survey are analyzed to assign specific weights to each domain, sub-domain or objective. For example, a test covering computer hardware and software, after being reviewed as part of a content study, might yield a result like the following:

Domain	Weight
Computer Hardware	40%
Computer Software	60%

In this simple example, a ten-question test that included four questions covering computer hardware subjects and six questions covering computer software subjects can be said to be sufficiently valid from a content standpoint.

Field Testing

Only by putting a test "to the test," by giving it to test takers, can the test's ability to measure what it claims to measure be evaluated. A field test provides test developers with important information to determine how well items in the test are performing, and how well the test is working overall. Outcomes can include:

- Information on the difficulty of each test item and other statistical information that can determine if test items are accurately measuring the objectives they are intended to assess.
- The test can be deployed to individuals whose skill level is also determined at the same time the test is given (a concurrent study) or after the test is taken (a predictive study). This independent evaluation can be performed using some kind of evaluation tool, such as a performance survey or grade at the end of a class on related subject matter. By correlating test scores with this independent performance measure, test developers can establish information for demonstrating the criterion validation of an assessment.
- The test as a whole can be analyzed for reliability using statistical techniques that determine the consistency of test items within the test.
- If a test fails to perform, especially if it does not correlate to an independent criterion measurement, that could lead to further analysis of the construct used in the test. For example, if a college entrance exam that used vocabulary and mathematical skill as a basis to predict success in college does not correlate well with some established measure of college success (such as graduation rates or grades), there may be a problem with the test questions used in a particular test. Alternatively, such results could lead test developers to review the assumption that the construct used as the basis of the test (vocabulary + math skill = college success) is an accurate construct.
- A field test of "beta" items can be used to gather data on a large number of test questions that can then be used to construct a final set of test forms to be deployed officially. By building a test that consists of items with good performance statistics based on the content balancing requirements of a content study, developers stand the best chance of having an assessment that is fair and reliable and can withstand further scrutiny utilizing additional validation techniques.
- A field test can also be used to collect data to determine scoring norms and a cut score (pass/fail mark) for a test. Analysis can also help developers to construct multiple test forms that, while containing different test items, are psychometrically equivalent.

There are a wide variety of methods for collect evidence of validity for a particular assessment. While these vary widely in complexity and not all are useful or practical in every circumstance, the more evidence collected concerning a test's performance, the greater the likelihood of success of a particular assessment project.

Implementation/Deployment

Once a test has been created, it needs to be deployed (made available to test takers) as part of an overall assessment program implementation. Deployment and other practical considerations need to be taken into account early in the process to assure that the final assessment product meets these practical considerations. Some issues related to implementation include:

Deployment

Traditionally, large-scale assessments have been delivered in paper-and-pencil (written) format, as software or over the Internet. The advantages and disadvantages of each delivery method is described in the table below:

Method	Description	Advantages	Disadvantages
Paper-and-Pencil	A written test, normally including linear and open-ended test questions	Can be delivered in any proctored environment, without requiring a computer or other equipment for each test taker	• Normally limited to linear items (limiting the usefulness of this format for effectively testing skills) • Logistics of distributing and collecting numerous paper exams and answer sheets • Manual data entry (or scanning) of answer sheets required for scoring

Mechanical	A set of instructions (exercises) a student must perform which are evaluated by a human grader reviewing an end product (such as the output file created by an application being tested)	Provides assessment of real world software without requiring development of custom testing applications	• Computer required for each test taker • Manual scoring requires trained graders who can score consistently, increasing the complexity and cost of widespread use of this test type • Exercises must be carefully constructed to ensure consistency
Software	Tests delivered in a software format on local computers or school networks	• Provides the ability to deliver both linear and performance-based test content • Automatic scoring of test items • Data processing capability of test scores • Can run on school computers and networks with minimum connectivity requirements	• Computer required for each test taker • Development of performance-based content normally requires significant investment • Software deployment localizes content and data on local machines, requiring additional effort to centralize content and scores
Internet	Test deployed from centralized servers to proctored environments with an Internet connection	• Centralized content and scores • New technologies allow performance-based content to be delivered over the Internet • Easier update of test content • More flexible score delivery and reporting options	• Computer required for each test taker • Internet connectivity required for each testing workstation • Security concerns regarding global access to Internet-based assessments

Deployment considerations will be based on the nature of the assessment (for example, if performance-based content is preferred then one of the automated assessment solutions described above is required) and the scaling issues discussed below.

Scaling

Most large-scale testing programs require testing of large numbers of students, often at the same time (to minimize exam overexposure which can lead to cheating). When one of the statewide English or mathematics exams designed to meet No Child Left Behind (NCLB) requirements is delivered to an entire grade level across a state, tens of thousands or even hundreds of thousands of students may be taking the same test simultaneously.

While many states have experimented with automating this process, to date the majority of academic or other large-scale testing (such as college entrance or AP exams) has been managed utilizing traditional paper-and-pencil testing. While this has solved certain logistical problems (notably, the ability to test in hundreds of locations across the state without the need for any equipment other than proper proctoring environments, desks and pencils), the logistics of managing hundreds of thousands of test books, answer sheets and other materials across many geographies is a significant burden, adding to the cost of implementing a testing program. And the limitation of written testing (which usually include just linear items that can be created on a printed page) makes this format less effective for testing skills, especially ICT literacy skills.

The good news for schools interested in implementing an IT literacy assessment is that a wide variety of powerful tools and technologies are available for creating and/or purchasing off-the-shelf high-quality, valid assessments that include sophisticated test items, such as performance-based test content for testing specific computer skills.

Scaling this technology into a school, district or even statewide program presents a number of challenges. Mechanical testing, which forms a component of at least one statewide IT exam program (in North Carolina) is extremely labor intensive on the part of human scorers. Software or Internet testing can eliminate the need for human scorers. However, implementing these solutions on a statewide basis requires that a test provider maintain enough capacity to deliver tens or even hundreds of exams simultaneously. Test centers must also maintain enough processing power and bandwidth to allow all students to take the test at the same time. And each automated or semi-automated solution (mechanical, software or Internet) requires that enough computer hardware be available to allow all test takers to take tests when necessary.

To minimize provider and school resource issues, some schools have utilized models that do not require all students to take the same test at the same time. This can include "on demand" testing in which students can take a test at any time, or scheduled programs in which different classrooms, schools or districts take tests at different scheduled times.

To a certain extent, a performance-based testing program lends itself to more flexible scheduling options. Unlike a linear exam that can be severely compromised if the questions in the exam are revealed before everyone has had the chance to take a test, performance-based exams require students to be able to perform specific functions correctly. Thus "cheating" on a performance-based exam would still require students to learn and understand the objectives of the exam in order to successfully answer test questions, even if those questions have been exposed to the public before a testing program is completed. Building multiple equivalent test forms for a testing program can also provide enough variety to ensure that test integrity is not compromised by exam overexposure.

Conclusions

As with any automated testing project, several factors (convenience of test delivery, convenience of scoring, availability of computer hardware and connectivity and test security considerations) have to be carefully weighed and balanced in order to ensure that an ICT testing program can be practically and successfully implemented.

Striking the right balance is the key to success in every element of the planning, development and deployment process for ICT literacy assessment. Educators must balance the need to meet state standards with the advantages of utilizing existing tools that may be built on a consensus standard. The benefits of performance-based testing must balance deployment considerations that limit this valuable format to automated test platforms that require each test taker to have access to a computer. And the need for widespread testing and security (which might encourage simultaneous testing across a grade level) must be matched with resources available to commit to testing, notably the availability of hardware and bandwidth within each school, and the available server resources of a test provider.

Success in an ICT literacy assessment project will require carefully weighing these factors in light of district and state-level ICT education goals, understanding what makes a quality assessment, and ensuring that all factors related to an assessment project (including deployment and scaling considerations) are taken into consideration before the first questions are written.