JRTE | Vol. 42, No. 4, pp. 361–389 | ©2010 ISTE | www.iste.org

Development and Validation of the Student Tool for Technology Literacy (ST²L)

Tina N. Hohlfeld
University of South Florida

Albert D. Ritzhaupt
University of Florida

Ann E. Barron
University of South Florida

Abstract

This article provides an overview of the development and validation of the Student Tool for Technology Literacy (ST²L). Developing valid and reliable objective performance measures for monitoring technology literacy is important to all organizations charged with equipping students with the technology skills needed to successfully participate in and contribute to a digital and global society. The purpose of the ST²L is to measure student technology literacy for low-stakes purposes of reporting aggregated school results, curricular planning, and students' self-assessment of technology skills. This article reports the development procedures and results of the pilot test conducted with eighth grade students (N = 1,561) to validate the functioning of this online, interactive tool. Analyses focused on item difficulty and discrimination by ability groups, completion time analysis, internal consistency reliability, and construct validity. ST²L was found to be a sound assessment tool for the intended purpose of low-stakes assessment of technology literacy. (Keywords: technology literacy, middle school, performance assessment, validation, student perceptions)

The Enhancing Education through Technology (EETT) section of the No Child Left Behind Act (NCLB) mandated that, beginning in the 2006–07 academic year, schools in the United States must document the technology literacy of their eighth grade students. In addition, international and national organizations (OECD, 2005; ITEA, 2007; ISTE, 1998, 2007; NAE and NRC in Gamire & Pearson, 2006), as well as many states in the U.S. (Metiri Group, 2009), have recommended including technology literacy skills in adopted standards. However, many important technology literacy skills cannot be objectively measured by traditional standardized assessment methods (Apple Computer, Inc., 1995; Lennon, Kirsch, Von Davier, Wagner, & Yamamoto, 2003; Quellmalz & Kozma, 2003; Russell & Higgins, 2003; Wenglinsky, 2005).

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

Achieving technology literacy for students is a challenging task requiring extensive planning, funding, training, monitoring, and evaluation. Currently, most states do not measure and monitor all of their students' technology literacy with consistent, valid, and reliable methods (U.S. Department of Education, 2009). Although a variety of strategies have been implemented, including achievement rubrics, such as the National Educational Technology Standards for Students: Achievement Rubric (Learning Point Associates, 2005); self-assessment surveys, such as Taking a Good Look at Instructional Technology (T.E.S.T., Inc., 2007); and skill/performance inventories (Georgia Department of Education, 2008; Stansbury, 2008), many educators and researchers agree that the best way to measure technology skills is through complex, real-world performance assessments (Axelson, 2005; Kay & Honey, 2005). Consequently, several performance-based, commercial tools have been developed, including the iSkills Assessment by Educational Testing Services (ETS) (Taylor, 2005). Other commercial organizations also provide assessment, training, and certification of technology skills (e.g., Computer Skills for Life by International Computer Driving License, The Internet and Computing Core Certification [IC^3] by Certiport [2009], and K to Eighth Power [2008]). Indeed, in November 2009, the ETS and Certiport released a new assessment, iCritical Thinking Certification, which is designed to use simulated and scenario-based assessment items with more traditional items to certify that individuals are able to critically apply digital technology skills in the workplace and academic environments. In addition, the National Assessment Governing Board (2009) has developed the draft of the Technological Literacy Framework, which will be used for the development of the 2012 NAEP computerized assessment for monitoring the progress of the students' technological literacy.

Australia and the United Kingdom also have supported the development of valid and reliable innovative assessments of information and communication technology (ICT) skills to use with their secondary students (MCEETYA, 2007; Qualifications and Curriculum Authority, 2008). Methods for developing valid, reliable, objective, and cost-effective assessments for measuring technology literacy skills are relevant and important for all organizations responsible for developing literate citizens who are prepared to responsibly participate in and successfully contribute to a digital and global society.

Background

Although commercial assessment tools are available, these tools generally require a substantial subscription fee or other cost for implementation. Consequently, the Office of Technology Learning and Innovation at the Florida Department of Education (FLDOE) funded a grant to develop an objective performance-based assessment tool for measuring student technology literacy without the price tag of a commercial product. Because the results

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191
(U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

of the assessments of student technology literacy for NCLB are aggregated and reported at the school level (Florida Department of Education, 2009), this assessment would have low-stakes consequences for the individual children who participate in the assessment. A specific goal of the project was to incorporate direct performance of technology-related tasks while staying within the boundaries of an objective, automatically scored assessment.

Conceptual Framework

The development and validation of the ST^2L was based on the union of two frameworks. The first is the Classical Test Theory (CTT), which provides the stages for test or measurement development (Crocker & Algina, 1986). The second framework is Design-Based Research (DBR), which emphasizes the iterative cycles used for developing educational innovations (Design-Based Research Collective, 2003). Crocker and Algina (1986) proposed 10 systematic steps for constructing tests:

1. Identify the primary purpose(s) for which the test scores will be used.
2. Identify the behaviors that represent the construct or define the domain.
3. Prepare a set of test specifications, delineating the proportion of items that should focus on each type of behavior identified in Step 2.
4. Construct an initial pool of items.
5. Have items reviewed (and revise as necessary).
6. Hold preliminary item tryouts (and revise as necessary).
7. Field-test the items on a large sample representative of the examinee population for whom the test is intended.
8. Determine statistical properties of item scores and, when appropriate, eliminate items that do not meet the pre-established criteria.
9. Design and conduct reliability and validity studies for the final form of the test.
10. Develop guidelines for administration, scoring, and interpretation of the test scores (e.g., prepare norm tables, suggest recommended cutting scores or standards for performance, etc.). (p. 66)

Using a computer for test administration provides the ability to include innovative methods and items. Parshall, Spray, Kalohn, and Davey (2002) propose that items can be innovative in five dimensions: (a) item format, (b) response action, (c) media inclusion, (d) level of interactivity, and (e) scoring method. The ST^2L is innovative in each of these five dimensions: (a) the items include simulations, (b) students must perform authentic technology tasks, (c) some tasks involve editing video and manipulating images, (d) the level of interactivity is greater than only clicking radio buttons, and (e) the ST^2L is automatically scored, the results are automatically recorded over the Internet, and a score report is automatically generated and delivered to the student. DBR focuses on continuous cycles of design, enactment, data collection, analysis,

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

and redesign (Design-Based Research Collaborative, 2003). DBR brings together the researcher, educators, and participants to improve an innovation (Design-Based Research Collaborative, 2003). This approach is especially relevant to the development of technology-based educational applications and assessments with innovative items. Thus, the iterative DBR process can support the 10 steps of classical test construction in the development of innovative, performance-based assessment technology literacy measures.

Development of the Student Tool for Technology Literacy (ST²L)

The FLDOE awarded a developmental grant to the Pinellas County School District, whose staff collaborated with the Florida Center for Interactive Media (FCIM) and measurement consultants. To minimize test anxiety among teachers and students using the low-stakes assessment, the development team used the word *tool* instead of *test*. The procedure for developing the tool was as follows: The development team (a) identified technology standards, (b) developed grade-level expectations/benchmarks for these standards, (c) outlined indicators for the benchmarks, (d) wrote knowledge assessment items, and (e) designed and programmed performance or skill assessment items.

Accomplishing all of the steps required to develop a valid and reliable performance assessment required an extended time span. Because the initial development of the ST²L began in 2005, the current standards that the International Society for Technology in Education (ISTE) published were the National Education Technology Standards for Students (NETS·S) (ISTE, 1998). Therefore, the development team used the NETS·S (1998), along with the standards published by other states and the standards used by individual school districts within Florida, to develop the benchmarks, indicators, and assessment items.

Identifying Technology-Related Indicators

Two groups were formed in 2005: the advisory committee, which was recruited from the Florida Council of Instructional Technology Leaders (FICTL) and from a cross-section of Florida's school districts, and the expert review panel, which was selected from experts with experience in instructional technology, student assessment, information science, or measurement. Approximately 20 people who were members of either the advisory group or the expert review panel drafted the indicators for the standards and benchmarks. The expert review panel then met with members of the FLDOE and FCIM to review the indicators and make decisions based on the measurability of the indicators and the types of items that could be developed in relation to each indicator.

The team developed a statewide survey, and each member of the advisory committee sent an e-mail with the link to the survey to 50 middle school educators and library media specialists to solicit feedback about the

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191
(U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

appropriateness of the indicators and the clarity of the wording. Because they used an informal snowball approach for the survey dissemination, response rates cannot be determined. Results obtained indicated that the majority of the indicators were appropriate expectations of technology-related knowledge for a middle school student. During their final analysis, the expert review panel decided that the indicators would be generic rather than hardware-specific in order for them to be relevant to all schools and settings. The final set of indicators is available in the Appendix (pp. 287–389).

Item Writing

Teachers and media specialists from Florida who were familiar with instructional technology and the capabilities of middle school students drafted the initial assessment items. The item-writing team sent any indicators that could not be assessed to the measurement team for review and/or deletion. A primary goal for the ST²L was the measurement of technology skills through direct examinee action rather than knowledge assessment. Thus, the item-writing team developed a variety of item types, including 67 performance-based tasks and 40 selected-response items, for a total of 107 items.

The selected-response item types include text-based multiple-choice and true/false items, as well as multiple-choice items with graphics and image map selections. The performance-based items require the examinee to complete tasks in simulated software environments. Decisions regarding the construction of the simulated tasks, including the choice of visual elements and their display on the screen, were guided by the Inventory of Teacher Technology Skills (Harmes, Barron, & Kemker, 2007; Parshall, Harmes, Rendina-Gobioff, & Jones, 2004), which the FLDOE had previously developed to assess technology literacy for teachers. School districts within Florida have diverse implementations of technology infrastructure; therefore, the decision to use a generic design for the software simulations was deliberate in order to make the tool relevant for all students. In addition, the FLDOE did not want to appear to endorse a particular platform, operating system, or software.

Each indicator formed the basic unit of assessment and was measured either by one task alone or by a series of steps, which were weighted proportionally. The completion of any given ST²L item was not dependent on the examinee's responses to previous items. This design decision was intended to ensure that an examinee would not be penalized on subsequent related items that followed a step that was completed incorrectly. The independence of each item is ensured by providing confirmation dialogue boxes for *Submit Step?* and *Skip Step?* buttons to allow the examinee to bypass any given performance-based item. With multi-step items, regardless of whether the examinee attempts an answer or chooses to skip the step, the program immediately scores the step and displays the screen needed to complete the subsequent task.

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

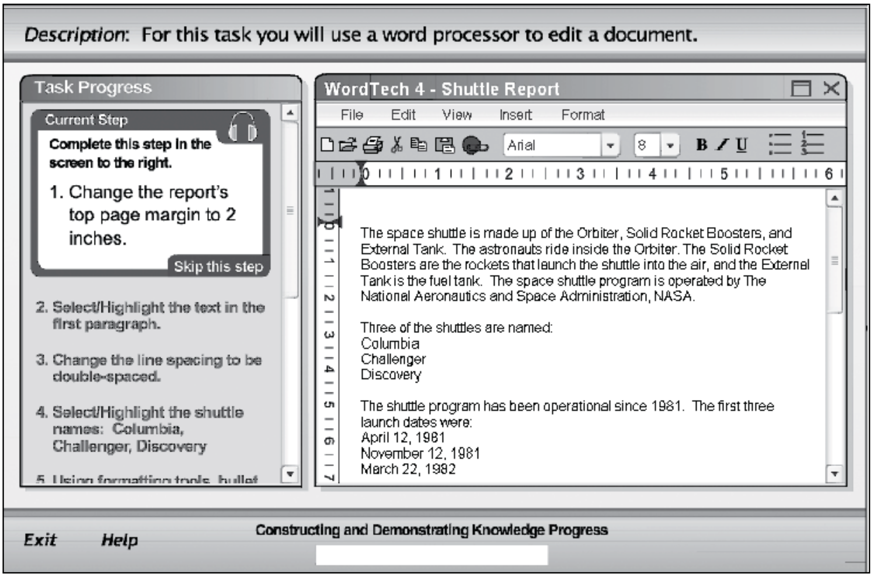

Figure 1. Example of a performance-based task item.

The ST²L was programmed in Adobe/Macromedia Flash and designed to be administered over the Internet. The design team elected to use Flash to create the interface for several reasons. First, Flash is easy for programmers, graphic artists, and instructional designers to use to create innovative interactive assessment items, which can include multimedia. Second, Flash can automate the process of delivering the assessment, collecting the data, scoring the assessment, and reporting the results to the student. Third, Flash applications delivered over the Internet are accessible on many different configurations of computer systems and browsers.

Figure 1 shows an example of the screen layout for a simulated task, and Figure 2 illustrates an example of a selected-response item. Brief instructions and a few practice items are provided as an introduction to the tool. The examinees are required to complete the practice items before beginning the actual assessment. The final version of the ST²L was composed of six different sections:

1. Software Use and File Manipulation
2. Ethics, Safety and Acceptable Use
3. Graphics, Presentation, and Video Editing
4. Spreadsheets
5. Browser Use and E-mail
6. Word Processing and Flowcharts

Usability Testing

After the development team completed approximately two thirds of the indicators for the Flash programming, the usability team conducted usability tests with 106 students in five middle schools, following the principles of DBR.

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191
(U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

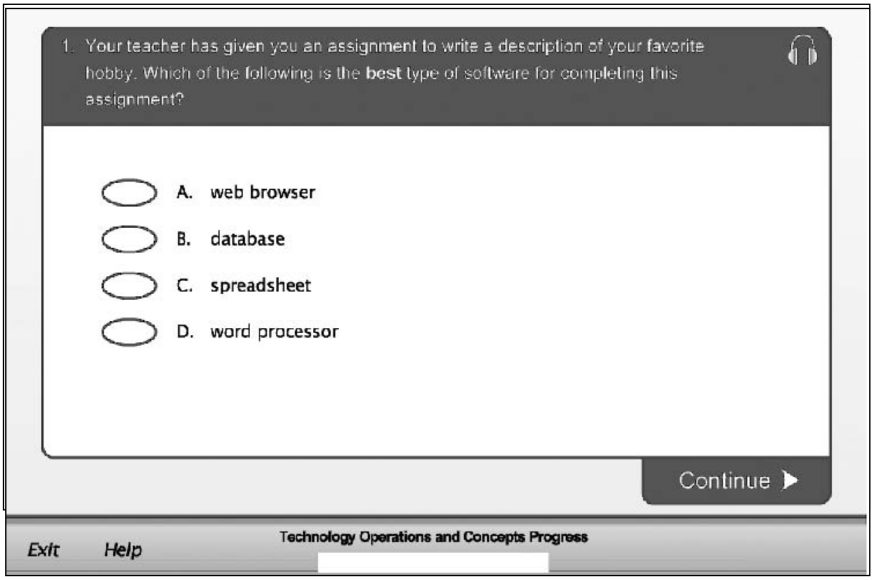

Figure2. Example of a multiple-selection item.

To obtain a representative sample of students, participating schools' technology instructors selected approximately seven students with basic to advanced levels of technological literacy. Each student completed the assessment in approximately 30 minutes, and then participated in a 40- to 65-minute follow-up discussion group session led by the project director.

During the usability test, the facilitators encouraged students to ask questions and express feedback. The discussion group followed a basic interview protocol that consisted of open-ended questions to stimulate conversation and collect feedback related to content and usability issues. An audio recording of the discussion sessions was transcribed as a reference for future revisions of the tool. A member of the usability test team also conducted interviews with at least one technology teacher/coordinator or principal at each school. These interviews, which ranged from 5 to 20 minutes, were conducted to provide potentially meaningful context to the usability testing process, to confirm the assignment of student technology skill levels, and to rate the level of access to technology at the school.

Results from Usability Testing

During the group discussions, every student who spoke had a favorable view of the tool as well as his or her experience being part of the usability testing. Students who contributed information found the tool's interface easy to use. The ST²L interface included a button for students to provide feedback about individual items during the administration of the usability test. The usability test team coded and analyzed this feedback to derive a set of themes that related to students' concerns about the tool. These data

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

indicated that most of the 106 students who participated found the ST²L to be satisfactory. Of the 465 comments provided, 72 (15%) were compliments or statements that the ST²L was just right. Some students (57 comments, 14%) even expressed the desire to use the ST²L for instructional purposes. Of a total of 125 comments related to the level of difficulty with using the ST²L, 57% of the comments stated that the tool was too easy; 31% of the comments stated the tool was just right; and 12% of the comments stated the tool was too difficult. Some students cited concerns about the ST²L in their feedback: 57 comments (14%) indicated confusion over the assigned tasks (mostly in the Browser section); 38 comments (8%) expressed concerns about the wording of some instructions; 40 comments (9%) expressed concerns about the user interface; and 23 comments (7%) questioned the authenticity of some tasks in the simulations.

The median score for the usability assessment was 27.9 out of 36 (SD = 4.0). Overall, 75% of the test takers received a score of 70% or greater, indicating a relatively easy set of items. The difficulty (p) of an item is measured by the proportion of examinees who responded correctly (Crocker & Algina, 1986). For the entire group of examinees, difficulty values ranged from .17 to .99, with most p-values between .70 and .99. This indicated that, for this group of examinees, the assessment was relatively easy.

Item discrimination is a measure of how well the item separates participants who perform well on the whole exam from participants who have the lowest performance. An analysis of the discrimination values of items answered incorrectly by more than half the examinees demonstrated that these items were measures of difficult advanced skills that high-level users answered correctly. Discrimination values (D-values) ranged from -.02 to .46. Most of the small D-values were accompanied by very high p-values, which is typical of mastery testing.

The internal consistency reliability estimate for scores on the entire tool was .74, which is considered acceptable (Nunnally, 1978). Because this assessment performs like a criterion-referenced test, there is a smaller degree of variability in scores across items and across examinees. Lack of variability results in lower measures of association, and thus, lower estimates of reliability than would be seen in a norm-referenced test (Crocker & Algina, 1986).

Pilot Study Design

A pilot study to validate the ST²L was conducted in the spring of 2008 with 1,561 eighth grade students from 40 middle/junior schools in 13 school districts in Florida. Figure 3 illustrates locations of the participating districts with stars. The pilot test team stratified the sample to ensure representation from small, medium, and large districts. Large school districts were defined as those with more than 100,000 students; medium school districts had more than 40,000 and fewer than 100,000 students; and small districts had fewer than 40,000 students. Five large school

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191
(U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

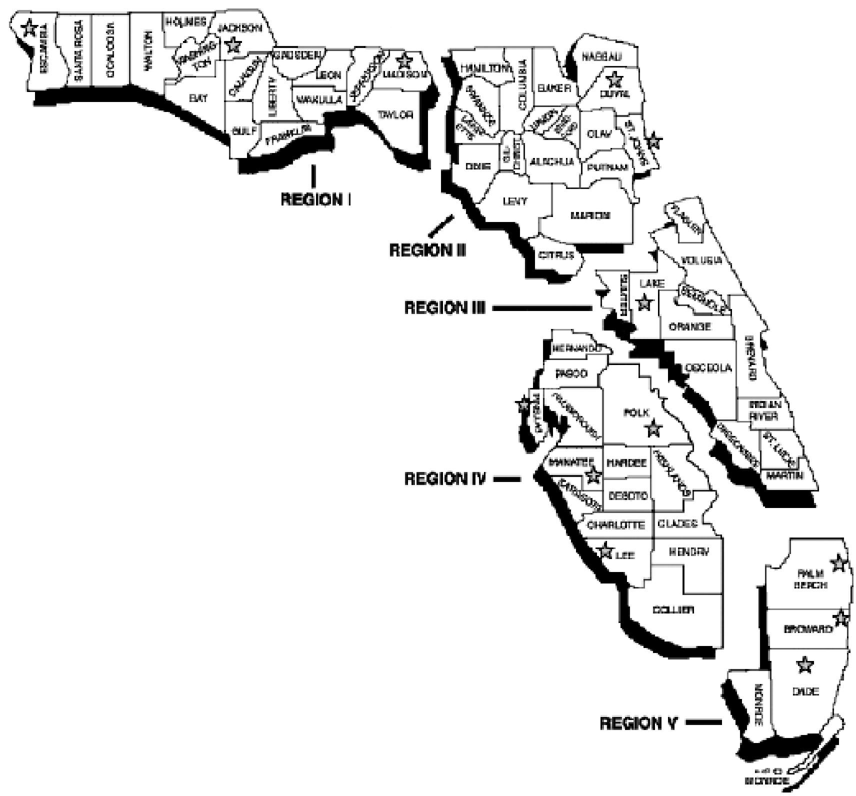

Figure 3. State map designating participating districts by region.

districts, four medium school districts, and four small school districts participated.

The pilot test team invited school districts that participated in the EETT Leveraging Laptops program (2006–07) in Florida to participate first, because they had been stratified across the state by size, rural, and urban areas and had access to technology (Kemker, 2007). Next, the team invited a second wave of school districts based on the sample stratification requirements and information collected about the schools (i.e., the percentage of students on free or reduced lunch, approximate number of student participants, the degree of technology exposure, headphone availability, and whether the administration of the tool would occur in a classroom or computer lab).

The following sequence was used for the pilot test: (a) a presurvey related to computer use and attitudes, (b) completion of the ST²L, and (c) a postsurvey. The presurvey collected basic demographic information, including gender, ethnicity, English as a primary language, age, and whether or not students were enrolled in free or reduced lunch programs as a proxy for socioeconomic status. The remaining items on the presurvey were adapted from items on the Programme for International Student Assessment

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

Table 1. Comparison of Population in Florida and Final Sample

	Students in Florida		Final Sample	
Category	Students	Percent	Students	Percent
District size				
Large	1,416,837	53.19	495	32.72
Medium	699,080	26.24	416	27.50
Small	548,056	20.57	602	39.79
Total	2,663,973	100.00	1,513	100.00
Region				
Panhandle	249,645	9.37	254	16.79
Crown	348,760	13.09	232	15.33
East Central	514,402	19.31	320	21.15
West Central	664,911	24.96	487	32.19
South	886,255	33.27	220	14.54
Total	2,663,973	100.00	1,513	100.00
Gender*				
Female	9,7846	49.05	766	50.63
Male	101,498	50.88	747	49.37
Total	199,344	100.00	1,513	100.00
Ethnicity*				
White	95,058	47.65	854	56.44
Black	45,837	22.98	318	21.02
Hispanic	47,877	24.00	218	14.41
Asian	4,456	2.23	43	2.84
Other	6,263	3.14	80	5.29
Total	199,491	100.00	1,513	100.00

*Note: * Eighth grade students only*

(OECD, 2003), which were related to students' use of and access to technology. The PISA questionnaire has been rigorously analyzed to demonstrate both reliability and validity across diverse international populations (OECD, 2003). The postsurvey included five items designed to assess the participants' satisfaction, their perception of how well the ST²L measured their ability, the perceived level of difficulty, and whether the tool reflected skills that they had learned at school.

To facilitate the administration of the instrument, the pilot test team created a proctor guide, which outlined administrative and technical requirements. The team also conducted a Web conference so that proctors and technology leaders could ask specific questions about the administration of a pilot study. After the administration of the ST²L, the team asked proctors to complete a survey to provide technical information (e.g., browsers used), the location of the administration, any technical difficulties (e.g., problems logging in), level of participant engagement, the type of computing resources

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191
(U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

available to student participants (e.g., headsets), and an overall evaluation of the proctors' perception of administrative issues.

Pilot Test Results

Participants

To determine the degree of success in obtaining a sample representative of the state (relative to the district size, number of students, and percent of students), we compared three categories: all students in Florida, the obtained sample, and the final sample. We dropped any student participant who did not complete at least one section of the ST^{2}L. This resulted in a final sample of N = 1,513 student participants, which is 96.9% of the obtained sample (N = 1,561) who began the assessment.

Table 1 compares classification categories of the school districts involved in the pilot program. Equally distributed school districts would have the same proportion of students in each category as the proportions of all students in Florida. Although the final sample does not match exactly, it was close enough so that we could use all of the data. Table 1 also shows the distribution of students across gender and ethnicity groups for the eighth grade students in Florida and the final sample. Initially, additional demographic information about the students' status for participation in free and reduced lunch (proxy for socioeconomic status), inclusion in special education, and having English as a second language was collected. However, due to a programming error, this valuable information was lost. The distribution of males to females was similar. Both the obtained sample and the final sample were similar in percentages for each ethnic group; however, when compared to the eighth grade population in Florida schools, the samples overrepresented white students and underrepresented Hispanic students.

To validate the function of the ST^{2}L tool for populations of students with diverse abilities, we grouped students into ability levels of technology literacy (beginner, intermediate, and advanced) based on their composite score created from two sections of the presurvey (Comfort with Computer Tasks and Frequency of Use). The composite measure was internally consistent with Cronbach's alpha calculated at $\alpha = .87$ for these data. We split the scores into approximate quartiles in which the top 22% were designated as advanced, the bottom 27% were designated as beginner, and the middle 51% were used to represent intermediate computing skills. Table 2 (p. 372) provides the distribution of students across these computer experience classifications for the final and obtained samples.

These analyses suggest that the final sample is similar to the demographics of students in Florida on some characteristics, and it is tenable to generalize these findings to the state of Florida in light of the stated limitations, as opposed to limiting the participants of the study by randomly sampling from the final sample to stratify across demographic characteristics.

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

Table 2. Distribution of Obtained and Final Sample by Computer Experience Classification

	Obtained Sample		Final Sample	
Computer Experience	Students	Percent	Students	Percent
Beginner	417	26.71	404	26.70
Intermediate	796	50.99	773	51.09
Advanced	348	22.29	336	22.21
Total	1,561	100.00	1,513	100.00

Psychometric Properties of the Presurvey (PISA)

The presurvey (PISA) included items that measure a participant's level of comfort with various computer tasks, frequency of computer use, and overall attitudes toward computer technology. Exploratory factor analyses (EFA) were conducted in each of the domains to explore the internal structure of the factors. The EFAs were executed with a random sample of 50% of the total participants or n = 780, which is approximately a 20:1 participant-to-item ratio and above the 10:1 threshold suggested by Kerlinger (1986). For each of the domains, we conducted EFAs using an orthogonal rotation (Promax), with the optimal model determined by using the proportion and Kaiser criteria. Although we ran additional models, the items did not meaningfully load or have simple structures, and thus, we decided to combine the individual items logically. The internal consistency reliability measured using Cronbach's alpha for the scores in each domain were Frequency of Computer Use (.79); Comfort with Computer Tasks (.87); and Attitudes toward Computers (.62).

ST^{2}L Test Quality and Characteristics

Time analysis. Time is an important consideration when assessing student skills. We included only those individuals who completed at least one section in the subsection analysis of time, and we included only individuals who completed all sections in the total time analysis (which could consist of several testing sessions). The median time for completing all sections of the tool was just over 37 minutes. However, the amount of time needed to complete all sections ranged from 13.72 to 183.35 minutes. Visual inspection of observations with the least amount of time to complete the tool revealed that participants were from different schools in different districts. We analyzed the data with the extended time outlier and without the outlier. The results changed very little; the median time decreased by .008 minutes, and the mean time decreased by .099 minutes. Because no proctors sent reports stating that there had been an irregularity, and the change in the mean and median times were so small, we retained the outlier in the data for our analyses.

Time might be a critical factor, because instructional time is valuable. Students with the least skills, who have the greatest need for instructional

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191
(U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

time, might require extended time to complete the assessment. Therefore, it is important to examine the time beginners need. When the time spent on each section is examined for the beginner group, the median amount of time for each section increases. However, the median time for completion of all sections of the ST²L tool increased by less than two minutes. The median time for beginners to complete all six sections of the tool was 39.24 minutes, whereas the average time was less than 41.54 minutes. These are reasonable time requirements for eighth grade students of all skill levels to spend in assessing their technology literacy skills. In addition, this time period fits within the typical school period of many schools.

Item analysis. Test quality is often evaluated by examining characteristics of the individual items through item analysis and by considering the overall test through estimates of reliability and validity. First, the difficulty (p) of an item is measured by the proportion of examinees able to respond correctly to the item (Crocker & Algina, 1986). Values range between 0 (most difficult) to 1 (least difficult). Ideal item difficulty is between .40 and .60 (Crocker & Algina, 1986). After calculating item difficulty, the discrimination (D) of the item can be determined by its point biserial correlation (Crocker & Algina, 1986). Item discrimination is a measure of how well the item separates participants who perform well on the whole exam from participants who have the lowest performance. Item discrimination values range between -1 and 1, with positive D-values indicating that the item discriminates in favor of the upper group and negative D-values indicating that the item discriminates in favor of the lower group. Items that require revision have D-values lower than .20; items with D-values greater than .30 require little or no revision; items with D-values over .40 are functioning satisfactorily (Crocker & Algina, 1986)

Total group of examinees. For the entire group of examinees, difficulty (p) values ranged from .05 to 1.00. Six items had difficulty levels below .30. Five of these items were simulation items that may need to be reviewed for possible revision. Most of these tasks (e.g., create a basic formula in a spreadsheet) were not intended to be difficult. One simulation item had no variability, because all participants performed the item correctly.

Item discrimination values ranged from .00 to .61. Sixty-four items were functioning very well with D-values equal to and greater than .40 (Crocker & Algina, 1986). Ten items had D-values below .20 and were recommended for review and/or revision (Crocker & Algina, 1986). Eight of these items were multiple-choice items, which should be carefully reviewed. Overall, these item statistics indicate that the ST²L can be used to assess the technology literacy of students, based on the indicators developed, at varying ability levels of computer literacy.

Beginner group of examinees. To examine how the ST²L tool functions with the lowest performing students, item statistics were calculated for the group of students who were rated as beginners by their responses on the

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

Table 3. Average Item Difficulty and Discrimination by Subsection for Level of Computer Experience

Section	All Students	Beginner	Intermediate	Advanced
		Level		
		Average Difficulty (p)		
Software Use and File Manipulation	74.43	68.67	75.01	79.66
Ethics, Safety, and Acceptable Use	83.93	80.44	84.91	85.65
Graphics, Presentation, and Video Editing	59.23	52.32	59.88	65.55
Spreadsheets	72.70	65.27	74.09	77.94
Browser Use and E-mail	84.63	79.77	85.90	87.24
Word Processing and Flowcharts	67.12	58.67	68.49	73.60
		Average Discrimination (D)		
Software Use and File Manipulation	0.31	0.32	0.27	0.31
Ethics, Safety, and Acceptable Use	0.36	0.38	0.32	0.37
Graphics, Presentation, and Video Editing	0.38	0.40	0.35	0.38
Spreadsheets	0.47	0.51	0.43	0.43
Browser Use and E-mail	0.43	0.48	0.37	0.43
Word Processing and Flowcharts	0.49	0.50	0.46	0.49

presurvey. For the beginner group, the item difficulties ranged from .03 to 1.00. Twelve items had *p*-values lower than .30, whereas 46 items had *p*-values between .80 and .99.

Item discrimination values for the beginner group ranged between -.02 and .65. Nine items had *D*-values below .20, indicating the need for review and/or revision (Crocker & Algina, 1986). Sixty-six items were functioning extremely well for the beginners' group with *D*-values greater than .40 (Crocker & Algina, 1986). These item statistics indicate that, with minimal revisions, the ST^2L will provide a well-functioning tool that can be used for assessing beginning students' technology literacy skills.

Average item difficulty and discrimination. We also examined the average item difficulty and discrimination for each section. Table 3 shows average item difficulty and item discrimination by subsection. Please note that, unlike the section scores provided to students, these values measure the average difficulty of the items within a section, as opposed to the average performance across the examinees. Trends in these measures are examined to identify potential problems. It is expected that the average difficulty of items for beginners will be higher than the average difficulty of the same items for intermediate and advanced level student. For the entire group and the beginners' group, the most difficult section was Graphics, Presentation, and Video Editing, and the easiest section for both groups was Browser Use and E-mail. When investigating the ST^2L by section, the tool functions well in discriminating between the best and worst performers. Average discrimination values that are over .30 for each section for each level of student indicate that the sections of items are functioning well.

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191
(U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

Table 4. Internal Consistency Reliability of ST²L Sections

Section	*n*	Nonstandardized KR-20
Software Use and File Manipulation	1513	.67
Ethics, Safety, and Acceptable Use	1442	.81
Graphics, Presentation, and Video Editing	1437	.72
Spreadsheets	1438	.82
Browser Use and E-mail	1400	.85
Word Processing and Flowcharts	1416	.86
Total for ST²L	1335	.95

Incorrect answer and distracter analysis. Additional analyses were conducted to examine patterns in the students' incorrect responses. For multiple-choice items, the percentage of participants who selected each wrong-answer distracter was examined to identify the most frequently selected distracters and the least frequently selected distracters. Four items had over 60% of the student participants select one incorrect distracter, while less than 10% of the participants selected a different incorrect distracter. These four items should be carefully reviewed for content.

None of the multiple-choice items required student participants to respond. Student participants could click the Continue button to skip the question. In these cases, a pop-up message stating, "You have not yet responded to all the items on the screen. Any items you leave blank will be scored as incorrect," warned student participants. Students had to respond by clicking Cancel or Continue in the message dialog box to continue. The percent of missing responses to items ranged from 0.0% to 1.85%. Students also had the option to skip steps in the performance-based simulated tasks. More than 10% of the participants selected the Skip Step option for 10 items. The development team should review these items to determine if they need to be revised.

ST²L reliability. We estimated reliability or internal consistency of scores across the student participants using the Kruder-Richardson 20 (KR-20) measure of internal consistency, which is used for dichotomously scored items (e.g., right or wrong, yes or no). We included all students who completed all items in at least one section in the analysis. To determine the reliability of the scores of the entire tool, we included all students who completed all items in the tool in the analysis. As shown in Table 4, the KR-20 reliability estimate for scores on the entire tool is .95. At the subsection level, the KR-20 reliability estimates range from .67 to.86 (see Table 4).

The degree to which the subsection scores correlate to the total score is an indication of the cohesiveness of the construct (see Table 5, p. 376). Please note that only participants who completed all sections received a total score, and thus, only those participants were used in the correlations. The ST²L inter-section correlations show positive relationships between subsection scores and the total score. Correlations of individual subsections with the

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

Table 5. Correlations between Scores on ST²L Subsections

Section of ST²L	1	2	3	4	5	6	7
Software Use and File Manipulation (1)	1						
Ethics, Safety, and Acceptable Use (2)	.52	1					
Graphics, Presentation, and Video Editing (3)	.57	.48	1				
Spreadsheets (4)	.55	.57	.61	1			
Browser Use and E-mail (5)	.55	.53	.50	.60	1		
Word Processing and Flowcharts (6)	.58	.54	.68	.68	.60	1	
Total on ST²L (7)	.77	.73	.81	.84	.78	.87	1

Note:

1 = Software Use and File Manipulation
2 = Ethics, Safety, and Acceptable Use
3 = Graphics, Presentation, and Video Editing
4 = Spreadsheets
5 = Browser Use and E-mail
6 = Word Processing and Flowcharts

$n = 1335$

$p < .0001$ for all

Table 6. Subsection and Total Scores by Computer Experience

Computer Experience**	1	2	3	4	5	6	Total	*n*
Beginner	71.78	77.45	49.34	61.35	75.55	51.30	64.84	404
Intermediate	78.81	82.27	58.02	71.04	82.46	63.23	72.77	773
Advanced	83.51	83.04	64.04	74.99	84.00	70.69	76.67	336

Note:

1 = Software Use and File Manipulation
2 = Ethics, Safety, and Acceptable Use
3 = Graphics, Presentation, and Video Editing
4 = Spreadsheets
5 = Browser Use and E-mail
6 = Word Processing and Flowcharts

** $p < .0001$

total ranged from .73 to .87. Correlations between scores on the subsections with the other subsections are strong and positive ($r > .45$), which affirms the cohesiveness of the construct. Please note that all correlations are statistically significant at $p < .0001$ level.

The weakest relationship with the total score and with the other subsections (Ethics, Safety, and Acceptable Use section) can be attributed to the subsection measuring a different aspect of technology literacy than the other subsections. Having strong technical skills (e.g., ability to use a spreadsheet) does not necessarily indicate that student participants can discriminate between ethical and nonethical uses of technology.

ST²L validity. To estimate construct validity of scores from the ST²L, we examined the relationships between scores on the ST²L and the rating from the presurvey (PISA) for (a) differences between computer experience levels and (b) correlations among the ST²L and the various scores of the PISA.

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191
(U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

Table 7. Correlations among ST²L Subsections and Presurvey Factors

Subsection	Frequency of Computer Use	Attitudes toward Computers	Comfort with Computer Tasks
Software Use and File Manipulation	.18	.23	.33
Ethics, Safety, and Acceptable Use	.10	.18	.25
Graphics, Presentation, and Video Editing	.17	.18	.34
Spreadsheets	.18	.17	.32
Browser Use and E-mail	.14	.17	.27
Word Processing and Flowcharts	.20	.19	.34
Average Total Score	.21	.23	.39

Note: $n = 1335$; $p < .0001$ for all

The first step was to analyze the differences across the computer experience levels. Table 6 shows the computer experience levels, as measured by the composite score of the survey across the subsection scores and the total scores. As predicted, there was a significant difference identified on overall performance based on computer experience ($F(2, 1332) = 52.65$, $p < .0001$), which demonstrates the validity of the ST²L in discriminating groups of students based on their expertise. A Tukey follow-up procedure with $\alpha <= .001$ confirmed that advanced users performed significantly better than both intermediate and basic users, and intermediate users performed significantly better than beginners.

Correlations of ST²L subsection scores with the presurvey factor scores were positive and all significant at $p < .0001$ (see Table 7). In general, the Comfort with Computer Tasks composite measure has the strongest correlations across all of the subsection scores as well as the total score. Comfort with Computer Tasks contained items that asked student participants to rate their comfort level with performing various technology-related tasks. The weakest relationship among the subsections was with Ethics, Safety, and Acceptable Use, because, as noted, this subsection measures a different aspect of technology literacy than the other subsections. Ethics, Safety, and Acceptable Use measures digital citizenship and students' ability to make responsible decisions, whereas the other subsections focus on students' technical skills. We used the simulations in the subsections of the ST²L, which were very similar to the computer tasks listed in the presurvey, to calculate the factor scores of the Comfort with Computer Tasks.

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

Table 8. Subscores and Mean Score by Demographics

Demographics	1	2	3	4	5	6	Total
All participants	77.98	81.18	57.10	69.37	80.97	61.70	71.62
Region**							
1 = Panhandle	72.78	77.54	51.65	60.88	78.12	52.36	65.98
2 = Crown	78.80	82.66	57.61	71.91	81.73	61.85	72.14
3 = East Central	80.28	82.43	58.60	70.59	81.90	64.08	72.85
4 = West Central	76.94	80.66	56.56	69.12	80.00	61.21	71.21
5 = South	84.85	85.04	64.91	76.40	85.19	74.31	78.81
Gender**							
Female	79.95	83.06	58.30	72.66	84.09	65.49	74.14
Male	75.96	79.21	55.86	65.96	77.70	57.71	68.95
Ethnicity**							
White	80.72	82.75	59.97	70.82	82.79	65.64	74.04
Black	70.35	77.33	48.28	64.52	76.50	50.78	64.99
Hispanic	78.67	80.48	58.72	71.07	79.89	62.10	72.22
Asian	81.49	83.76	62.90	77.29	87.78	68.45	77.48
Other	75.31	80.51	54.11	64.45	79.00	58.69	68.09
District Size							
Large	78.08	81.94	57.60	70.78	81.49	61.85	72.24
Medium	77.99	80.74	56.43	69.90	80.33	62.41	71.55
Small	77.89	80.88	57.17	67.83	81.00	61.08	71.16

Note:

1 = Software Use and File Manipulation
2 = Ethics, Safety, and Acceptable Use
3 = Graphics, Presentation, and Video Editing
4 = Spreadsheets
5 = Browser Use and E-mail
6 = Word Processing and Flowcharts

n = 1335

**$p < .0001$

Statistical Differences of ST²L

Table 8 presents subsection scores and mean scores for all sections (percent correct) by demographic variables. Overall, across sections and demographic conditions, the average score was 71.62%. We detected significant differences on overall performance based on region ($F[4, 1330] = 22.54$, $p < .0001$), gender ($F[1, 1333] = 37.02$, $p < .0001$), and ethnicity ($F[4, 1330] = 20.62$, $p < .0001$). We detected no differences across district size ($F[2, 1332] = 0.57$, $p = .57$). Instead of examining each demographic category separately, future research needs to examine models that include all of the demographic variables expected to influence student outcomes. Then the isolated differences among specific demographic categories, such as gender, can be examined while statistically controlling the confounding influences from the other demographic variables in the model.

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

Table 9. Questions, Responses, and Response Frequencies to Postsurvey

Question/Response	*n*	Percent
Did you find using this tool fun?		
None of the activities were fun (1)	167	12.82
Some of the activities were fun (2)	597	45.82
Most of the activities were fun (3)	539	41.37
How easy or difficult was using the tool for you?		
Difficult (1)	81	6.22
Just right (2)	665	51.04
Easy (3)	557	42.75
How well did this tool measure your computer skills and knowledge?		
Did not show my skills and knowledge (1)	88	6.75
Allowed me to show some of my skills and knowledge (2)	536	41.14
Was a good way to show most of my skills and knowledge (3)	679	52.11
Have you done activities at school that are similar to the ones you just completed within the tool?		
We have done none of these activities at school (1)	196	15.04
We have done very few of these activities at school (2)	285	21.87
We have done some of these activities at school (2)	477	36.61
We have done most of these activities at school (4)	345	26.48

Analysis of the Postsurvey

The postsurvey provided student participants with the opportunity to express their opinions about the ST²L (see Table 9). More than 80% of the sample who responded to the postsurvey (n = 1,303) indicated that some or most of the activities in the ST²L were fun. Few (6%) of the student participants found using the ST²L tool difficult.

Although the subsection scores were visible on the display, the students were also asked to rate how well they thought that the tool measured their technology knowledge and skills. More than half of the student participants indicated that the tool was a good way for them to demonstrate most of their skills and knowledge. Almost two thirds of the student participants indicated they had done some or most of the activities that the ST²L assessed at school.

Analysis of Proctor Survey

The responses of students participating in the ST²L were linked to one of 72 teachers/ proctors, who had access codes to participate in the post-ST²L

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

Table 10. Types of Problems Reported by Proctors

Type of Problem	Proctors	Percent
Login problems	13	44.83
Internet connection problems	1	3.45
Technical problems	15	51.72
Students did not understand directions	6	20.69
Students had problems with specific questions	13	44.83
Interruptions during the administration	10	34.48
Students engaged in off-task activities	3	10.34

administration survey. Twenty-eight proctors (39% of proctors with access codes) responded to at least part of the proctor survey. The respondents proctored the administration for 627 students (41% of all student participants). Results from these proctors indicated that more than half of the ST^{2}L sessions were administered in computer labs (57%), 18% were conducted in classrooms, 18% in media centers, and 7% in other areas. The responding proctors reported that, on average, 22 computers were used concurrently during the administration. Forty-six percent of the responding proctors reported that they administered the ST^{2}L alone; however, 50% of responding proctors reported that an additional facilitator was available to assist students in completing the ST^{2}L; one proctor (4%) reported three additional facilitators were available. Forty-three percent of the responding proctors reported that students completed the ST^{2}L in one session. Only one responding proctor stated that the administration took longer than two sessions. All of the responding proctors reported that students used Internet Explorer to complete the tool.

Proctors were provided an opportunity to report the types of problems that they encountered (see Table 10). The most common problem reported was related to the log on procedure for student participants. An analysis of the explanations showed that, in every case, the problem was a result of a spelling error in the passwords that were provided to the proctors. Only one proctor reported an Internet connection problem during the administration. Half of the responding proctors reported at least one technical problem, such as the installation of the Flash 9.0 Player plug-in, laptop batteries running out of power, or problems advancing from one section to the next. In each case, proctors reported resolving the problems so that student participants were able to continue the assessment.

Approximately 20% of the responding proctors reported that some students had problems understanding directions in the Word Processing and Flowcharts section. Approximately one third of the proctors' responses indicated that there were interruptions during the administration period (e.g., fire drill, assembly, end of class session, end of day). More than 90% of responding proctors rated the overall level of student effort and engagement observed during the administration of the ST^{2}L as high or very high.

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191
(U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

The success of the administration of the ST²L was confirmed when 100% of the responding proctors either agreed or strongly agreed with the statement "The administration of the ST²L ran smoothly."

Discussion

Prior to its development, the purpose of the ST²L was determined to be for low-stakes assessments for monitoring technology literacy of eighth grade students within the state of Florida. The results of the ST²L were intended to be used for NCLB reporting of aggregated school-level results, school district curricular purposes, and individual students identifying relative strengths and weaknesses in their technology literacy. Based on these stated low-stakes purposes, appropriate development processes were followed throughout the project. For the initial stage of construct definition, the indicators were carefully developed and based on state and national technology standards (NETS·S). The development team also followed sound development procedures for item writing and review and conducted usability analyses to ensure that the user interface and the simulated performance-based tasks were as clear and intuitive as possible. Finally, the pilot test team validated functioning according to a sound research design and sampling plan.

The purpose of the pilot test was to demonstrate the overall assessment quality by considering item analyses, reliability, and validity. The item *p*-values (item difficulty) for the entire group ranged from .05 to 1.0, with 66% of the items having values between .30 and .89. In contrast, item discrimination values ranged from .00 to .61, with 60% of the items functioning very well with *D*-values equal to or above .40. Although the ST²L is designed as a criterion-referenced or a mastery test, these results demonstrate substantial variability across the items and student participants. This is not an error in the assessment; rather, it is an indication of the diversity of technology literacy skills across the sample of student participants.

The item differential indices are lower (i.e., more difficult) for those examinees who were classified as being at a beginner level of computer experience. For the beginners' group, the item difficulties ranged from .03 to 1.00, whereas item discrimination values for the beginner group ranged between -.02 and .65. The entire group performed better than the beginners' group across the subsections and the total score. When compared to the entire group, the items were more discriminating across the subsections with the beginners' group. In addition, the average time to complete the ST²L increased when comparing the beginners group with the entire group.

Reliability was estimated for the tool for each subsection and for the total score using KR-20. All reliability estimates were found to be positive and above the social science acceptable threshold (KR-20 > 0.7) (Nunnally, 1978), with the exception of Software Use and File Manipulation, which had a KR-20 = .67. All of the subsection scores are significantly and positively correlated ($r > .45$). Further, the total score on the ST²L significantly and positively corre-

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

lates with each subsection score ($r > .75$). The high inter-section score correlations and total score correlations demonstrate a highly cohesive and internally consistent structure of the ST²L for the sample participants.

To estimate the construct validity of scores, we compared the results of the ST²L to the presurvey scores (PISA). Specifically, we used two approaches to demonstrate the relationships. First, we calculated the subsection scores and total scores across computer experience levels based on the results of the presurvey. Then we compared the results of the students in each of these levels on the total ST²L assessment. The advanced group performed significantly better than the intermediate group, and the intermediate group performed significantly better than the beginners' group. Further, subsection scores for the ST²L in each of the computer experience classifications differentiated in the same way.

Second, we calculated the correlations between the subsections of the ST²L and the presurvey measures of Frequency of Computer Use, Attitudes toward Computers, and Comfort with Computer Tasks. As hypothesized, all the correlations were significantly and positively correlated. Most of the measures were modest correlations. The Comfort with Computer Tasks measure, which was designed to match the indicators in the ST²L, had the highest relationships with each of the subsection scores and the total score. These analyses show the construct validity of the ST²L in relation to the stated indicators and the broad construct of technology literacy.

Limitations

The results of this pilot test must be interpreted with an understanding of the limitations. This pilot test was limited to eighth grade students (N = 1,561) from 13 school districts in Florida during the spring of 2008. Although the ST²L is intended to be software independent, students may not find the interface similar enough to the specific software suites that they are accustomed to using in their schools and homes. Thus, the ST²L may not adequately measure the knowledge and skills of these students. The validation process included comparing the ability of the ST²L to separate students by their perceived technology ability levels. For this study, we used the results from a self-assessment, which may not adequately represent the students' true abilities, to classify students to computer experience levels. A final limitation of the current pilot test is that variables were not included in the analysis for socioeconomic status, primary language other than English spoken at home (ESL), or special education status of the student participants, because a programming error resulted in the loss of this relevant information.

Conclusions

The process of developing and validating the ST²L can be used as a model for others who would like to develop innovative performance assessments

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191
(U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

of technology literacy skills. The successful development of ST²L required many different teams of people to carefully plan, monitor, and adjust many activities. The development of the ST²L followed an extensive, thorough process (based on CTT and DBR frameworks) for defining the indicators, which were developed within the framework of the NETS·S (1998) and state standards. The items were mapped to these indicators and provide measurements of them in innovative, relevant, performance-based ways. Finally, test-quality criteria show acceptable item analysis, reliability, and validity results for the tool. With few modifications, the ST²L was ready for low-stakes implementation and should become a useful tool of school districts for reporting aggregated technology literacy scores of schools for NCLB purposes and for helping students and districts target technology-related curricular needs.

In the 2008–09 school year, the ST²L was made available for districts in Florida to use with their students. In addition to using the tool for NCLB aggregated school-level reporting purposes, teachers can adapt the delivery to meet their students' instructional needs. Some teachers might have their students take all sections of the ST²L and analyze the results to determine which skills the class needs to develop. Then the teacher can deliberately integrate those technologies into the daily instructional activities to guide the students' technology literacy development. Others might administer a section of the ST²L, such as the spreadsheet section, as a pretest before beginning an authentic unit, such as collecting data about an ecosystem in the school's local environment. Students would use spreadsheets as they observe and record data, and later analyze the information to find patterns and create charts and tables to support their recommendations to their school district and community planners. At the conclusion of the unit, teachers could have the students take the assessment again as the posttest and help them identify the growth in their technology skills by comparing the changes in their scores.

Students can also use the ST²L to monitor and track the progress of their technology literacy skills. This information could guide their choices of courses or projects within courses. Students can share the results and feedback from the ST²L with their parents, and thus, open communication channels with their families about the importance of technology literacy skills.

Schools might have their entire student population take the ST²L each year. Schools could then measure and track the longitudinal growth of their schools' level of technology literacy skills. Special programs could be created to support the development of technology literacy of special groups of students or to support the development of the whole school's specific technology skills. Districts might use the longitudinal data collected to plan, fund, and monitor special technology initiatives.

The ST²L was designed to be a flexible tool for the assessment of student technology literacy in order to support the integration of technology into the

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

curriculum and students' daily learning experiences. The next steps require that future research look at how teachers, schools, and districts utilize this tool to determine if the assessment reports that the ST²L provides support the curricular decisions for which it is intended, which is to support students' acquisition of technology literacy skills through modification of students' instructional experiences.

Professional standards for testing require certain actions on the part of test developers. For example:

> Standard 1.1: A rationale should be presented for each recommended interpretation and use of test scores, together with a comprehensive summary of the evidence and theory bearing on the intended use or interpretation (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999, p. 17).
>
> Standard 1.4: If a test is used in a way that has not been validated, it is incumbent on the user to justify the new use, collecting new evidence if necessary (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education., 1999, p. 18).

These statements of test development standards make clear both the necessity of evaluating a final test product and of using an assessment solely for the purpose for which it was developed.

The ST²L was specifically developed as a low-stakes assessment and was designed to provide data related to the technology literacy of eighth grade students for district aggregated reporting, curriculum design, and student self-assessment. The tool, in its present form, is not suitable for use in high-stakes applications, such as computing school grades or evaluating individual student performance for promotion/retention. Prior to using the ST²L for high-stakes purposes, several additional procedures would need to be implemented. These steps include: (a) ensuring standardized administration procedures (consistent, proctored environment, time limits, etc.); (b) developing multiple, parallel test forms (to ensure test security and equitable conditions); (c) creating an appropriate scoring model; and (d) conducting additional data analyses (Cizek, 2001).

Validation of measurement instruments is an ongoing process. This is especially true when dealing with the measurement of technology literacy while using technology, because technology is perpetually changing. The capabilities of the hardware and software continue to improve and new innovations are introduced. As a result, developing valid and reliable instruments and assessment tools to measure the construct is difficult and an ongoing process.

The development team is already in the process of mapping the items developed for the ST²L to the new standards of the NETS·S (ISTE, 2007).

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191
(U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

Then the team will reorganize and revise them and develop a new set of innovative items to create a reliable and valid objective performance measure of technology literacy skills.

Author Notes

Tina N. Hohlfeld is an assistant professor of instructional technology at the University of South Florida. Her research focuses on examining the relationship between technology integration and positive outcomes for K–12 students. Her research has been published in Computers & Education, The Quarterly Review of Distance Education, International Journal of Education Policy and Leadership, and Journal of Interactive On-line Learning. Correspondence regarding this article should be addressed to Tina Hohlfeld, College of Education, Department of Secondary Education, University of South Florida, 4202 E. Fowler Avenue, EDU 105, Tampa, FL 336200-5650. Email: thohlfeld@coedu.usf.edu

Albert D. Ritzhaupt is an assistant professor of educational technology in the School of Teaching and Learning at the University of Florida. His primary research areas focus on the development of technology-enhanced instruction and technology integration in education. His publications have appeared in the Journal of Computing in Higher Education, Journal of Research on Technology in Education, International Journal of Education Policy and Leadership, Journal of Information Systems Education, Computers & Education, Behavior Research Methods, *and* Computers in Human Behavior. *E-mail: aritzhaupt@gmail.com*

Ann E. Barron is a professor of instructional technology at the University of South Florida. Her research interests include Web-based distance learning, multimedia instruction, and audio technology. She serves on the editorial boards of the Journal of Educational Computing Research, Computers in Human Behavior, Journal of Computing in Higher Education, *and* Journal of Educational Multimedia and Hypermedia. *E-mail: barron@usf.edu*

References

American Educational Research Association, American Psychological Association (APA), & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, D.C.

Apple Computer, Inc. (1995). *Changing the conversation about teaching, learning, & technology: A report on 10 years of ACOT research.* Cupertino, CA: Apple Computer, Inc. Retrieved September 30, 2009, from http://imet.csus.edu/imet1/baeza/PDF%20Files/Upload/10yr.pdf

Axelson, M. (2005). Online assessments of technology literacy: The perfect Petri dish. *Insight, 5*, 1–24.

Certiport. (2009). The Internet and Computing Core Certification (IC³®). Retrieved September 30, 2009, from http://www.certiport.com

Cizek, G. J. (Ed.). (2001). *Setting performance standards: Concepts, methods, and perspectives.* Mahwah, NJ: Lawrence Erlbaum Associates.

Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory.* Belmont, CA: Wadsworth Group/Thomson Learning.

Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5–8.

Florida Department of Education. (2009). *Final LEA EETT Part I: Entitlement project reporting guidance.* Retrieved September 30, 2009, from http://www.fldoe.org/BII/Instruct_Tech/EETT/Part1/word/2008-09-EETT-PART-I-FINAL-EVAL-REPORT.doc

Gamire, E., & Pearson, G. (Eds.). (2006). *Tech tally approaches to assessing technological literacy.* Washington, D.C.: The National Academies Press.

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

Georgia Department of Education (2008). *One stop shop for educators: Georgia technology literacy assessment toolkit*. Retrieved September 30, 2009, from http://gadoe.georgiastandards.org/toolkit.aspx

Harmes, J., Barron, A. & Kemker, K. (2007). Development and implementation of an online tool to assess teacher technology skills. In R. Carlsen et al. (Eds.), *Proceedings of Society for Information Technology and Teacher Education International Conference 2007* (pp. 802–805). Chesapeake, VA: AACE.

International Computer Driving License (ICDL). (2009). *The international computer driving license program*. Retrieved September 30, 2009, from http://65.99.150.19/icdlus-lms-webclient/homepage/index.jsp

International Society of Technology in Education (ISTE). (1998). *NETS for students.* Retrieved September 30, 2009, from http://www.iste.org/Content/NavigationMenu/NETS/ForStudents/1998Standards/NETS_for_Students_1998.htm

International Society of Technology in Education (ISTE). (2007). *NETS for students, refreshed.* Retrieved September 30, 2009, from http://www.iste.org/Content/NavigationMenu/NETS/ForStudents/2007Standards/NETS_for_Students_2007.htm

International Technology Education Association (ITEA). (2007). *Standards for technological literacy* (3rd ed.). Reston, VA: International Technology Education Association (ITEA). Retrieved September 30, 2009, from http://www.iteaconnect.org/TAA/PDFs/xstnd.pdf

K to Eighth Power. (2008).*Computer literacy and technology education.* Retrieved September 30, 2009, from http://www.kto8.com/index.php

Kay, K., & Honey, M. (2005). Beyond technology competency: A vision of information communication technology literacy to prepare students for the 21st century. *Insight, 5*, 1–18.

Kemker, K. (2007). Leveraging laptops. *T.H.E. Journal, 34*(7), 32. Retrieved on September 30, 2009, from http://thejournal.com/articles/2007/07/01/curriculumbased-reform--florida.aspx

Kerlinger, F. (1986). *Foundations of behavioral research* (3rd ed.). New York: Holt, Rinehart, and Winston.

Learning Point Associates. (2005). *NETS for students: Achievement rubric grades PK–12.* Retrieved September 30, 2009, from http://www.ncrel.org/tech/nets/p-12rubric.pdf

Lennon, M., Kirsch, I., Von Davier, M., Wagner, M., & Yamamoto, K. (2003). *Feasibility study for the PISA ICT literacy assessment*. Retrieved September 30, 2009, from http://www.oecd.org/dataoecd/35/13/33699866.pdf

Metiri Group. (2009). *National trends report: Enhancing Education through Technology (EETT) Round 6, fiscal year 2007.* Washington, D.C.: The State Educational Technology Directors Association (SETDA). Retrieved September 30, 2009, from http://www.setda.org

Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA). (2007). *National Assessment Program: ICT literacy years 6 and 10 report 2005.* Carlton South, VIC, Australia: Ministerial Council on Education, Employment, Training and Youth Affairs. Retrieved September 30, 2009, from http://www.mceetya.edu.au/verve/_resources/NAP_ICTL_2005_Years_6_and_10_Report.pdf

National Assessment Governing Board (NAGB). (2009). *Technological literacy framework for the 2012 National Assessment of Educational Progress, discussion draft–7/24/09*. National Assessment Governing Board. Retrieved September 30, 2009, from http://www.edweek.org/media/nagb_assessment_devel_comm_aug_7-09.pdf

Nunnally, J. (1978). *Psychometric theory.* New York: McGraw-Hill.

Organisation for Economic Co-Operation and Development (OECD). (2005). *Are students ready for a technology-rich world? What PISA studies tell us.* Paris: OECD Publications.

Parshall, C. G., Harmes, J. C., Rendina-Gobioff, G., & Jones, P. K. (2004). *Final evaluation report: Teacher technology literacy inventory: Development and field test.* Florida Department of Education. Retrieved September 30, 2009, from http://www.flstar.org/docs/TeachTechLitFinalEval.pdf

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191
(U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

Parshall, C. G., Spray, J. A., Kalohn, J. C., & Davey, T. (2002). *Practical considerations in computer-based testing.* New York: Springer.
Qualifications and Curriculum Authority. (2008). *KS3 ICT assessment tasks*. Retrieved September 30, 2009, from http://testsandexams.qcda.gov.uk/18908.aspx
Quellmalz, E. S., & Kozma, R. (2003). Designing assessments of learning with technology. *Assessment in Education, 10*(3), 389–407. Retrieved September 30, 2009, from http://robertkozma.com/images/kozma_ed_assessment_quellmalz.pdf
Russell M., & Higgins, J. (2003). *Assessing effects of technology on learning: Limitations of today's standardized tests.* Boston College: Technology and Assessment Study Collaborative. Retrieved September 30, 2009, from http://escholarship.bc.edu/intasc/3/
Stansbury, M. (2008). States struggle with assessing technology literacy. *eSchool News Online.* Retrieved September 30, 2009, from http://www.eschoolnews.com/news/top-news/?i=53017
Taylor, L. (2005). *ICT literacy: Equipping students to succeed in an information-rich, technology-based society.* Educational Testing Service. Retrieved September 30, 2009, from http://www.ets.org/Media/Tests/ICT_Literacy/pdf/ICT_Equipping_Students_to_Succeed.pdf
T.E.S.T., Inc. (2007). *TAGLIT: Taking a good look at technology.* Retrieved September 30, 2009, from http://www.testkids.com/taglit/?CFID=99530&CFTOKEN=26228741
U.S. Department of Education. (2001). *No Child Left Behind: Enhancing Education through Technology Act of 2001.* Retrieved September 30, 2009, from http://www.ed.gov/policy/elsec/leg/esea02/pg34.html
U.S. Department of Education. (2009). *Evaluation of the Enhancing Education through Technology Program: Final report.* Washington, D.C.: Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service. Retrieved September 30, 2009, from http://www.ed.gov/about/offices/list/opepd/ppss/reports.html
Wenglinsky, H. (2005). *Using technology wisely: The keys to success in schools.* New York: Teachers College Press.

Appendix

ST^2L Indicators

I. Essential Operational Skills

The student can:

1. Use help functions within an application for assistance.
2. Respond appropriately to information presented in a dialog box. (e.g., replace a file dialog).
3. Select correct printer.
4. Use print preview.
5. Change page orientation between landscape and portrait.
6. Print a specific page range.
7. Demonstrate practical keyboarding skills.
8. Identify and locate the standard menu bar.
9. Toggle between two open software applications.
10. Create a new file.
11. Locate and open a specific file.
12. Rename a file.
13. Move a file to a different location.
14. Search for specific files.
15. Use "Save As…" to change the name of the working file.
16. Use "Save As…" to save a file to a different location.

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

II. Constructing and Demonstrating Knowledge

The student can:

1. Select the best device to complete a given task, such as digital cameras, scanners, and external storage devices.
2. Select appropriate uses for word processing software.
3. Use the ordered and unordered list features of a word processor.
4. Use the table creation feature of a word processor.
5. Insert a hyperlink into a document.
6. Insert an image into a document.
7. Set page margins within a word processing document.
8. Adjust line spacing within a word processing document.
9. Insert an object using the drawing tools feature of a word processor.
10. Edit images within software using cropping.
11. Edit images within software using resizing.
12. Edit images within software using rotating.
13. Edit images within software using brightness/contrast.
14. Edit images within software using duplicating.
15. Select appropriate uses for Web browser software.
16. Identify a Web browser.
17. Identify and use the address bar in a Web browser.
18. Identify and use the back function in a Web browser.
19. Identify and use the Refresh function in a Web browser.
20. Identify and use the bookmarks/favorites elements in a Web browser.

III. Communication and Collaboration

The student can:

1. Use e-mail to send a message.
2. Use e-mail to receive/open a message.
3. Use e-mail to forward a message.
4. Use e-mail to reply to a message.
5. Use e-mail to add attachments to a message.
6. Select appropriate uses for presentation software.
7. Create new slides within presentation software.
8. Enter content within presentation software.
9. Play a slide show within presentation software.
10. Perform basic digital video editing by removing a section of video.
11. Perform basic digital video editing by adding narration.
12. Perform basic digital video editing by adding music.
13. Insert an edited video clip into presentation software.

IV. Independent Learning

The student can:

1. Perform Web searches that produce relevant results.
2. Use the advanced search features of search engines. (e.g., Boolean, date limits, language, etc.).

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191
(U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

3. Access information through online resources including encyclopedias, libraries, education and government websites, and electronic catalogs (a.k.a. card catalogs).
4. Evaluate Internet sites for accuracy.
5. Select appropriate uses for graphic organizer software.
6. Create flowcharts as a learning strategy.
7. Create concept maps as a learning strategy.
8. Select appropriate uses for spreadsheet software.
9. Enter data into a spreadsheet.
10. Format data in a spreadsheet.
11. Delete data in a spreadsheet.
12. Use spreadsheets to compute basic formulas.
13. Use spreadsheets to create a graph.
14. Import and export data (e.g., copying and pasting from spreadsheet to presentation software).

V. Ethical, Legal, and Safety Issues

The student can:

1. Differentiate between appropriate and inappropriate use of school computers (acceptable use policy).
2. Use and appropriately cite electronic references.
3. Understand and follow copyright laws pertaining to software and/or Internet resources, including duplicating and/or plagiarizing text and media files.
4. Identify an appropriate procedure to follow when a peer is using the computer inappropriately.
5. Identify an appropriate procedure to follow when inappropriate content is encountered on a computer.
6. Display an awareness of potentially inappropriate language while using technology.
7. Display an awareness of potentially inappropriate media use in regards to technology.
8. Display an awareness that technology is in a state of continual change/advancement.
9. Identify security risks that are involved with giving out personal information (e.g., fake eBay sign-in to steal password).
10. Understands there is no guarantee of privacy on a network.
11. Recognize and report potential online predators (e.g., strangers asking inappropriate questions).
12. Recognize the risks of downloading files and documents.
13. Recognize the permanency of electronic data.
14. Maintain password security.
15. Understand the need for virus scans, pop-up blockers, spyware blockers, firewalls, and filters.

Copyright © 2010, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.