Final Evaluation Report

Student Tool for Technology Literacy (ST[2]L) Development and Pilot Test

prepared for the

Office of Technology Learning and Innovation
Florida Department of Education

Dr. Ann E. Barron
Dr. Tina N. Hohlfeld
University of South Florida

Dr. Albert D. Ritzhaupt
University of North Carolina Wilmington

August 1, 2008

Table of Contents

List of Figures

List of Tables

Executive Summary

The purpose of the Student Tool for Technology Literacy (ST^{2}L) is to help districts evaluate students' technology literacy for No Child Left Behind (NCLB) reporting requirements and to provide an objective measure for targeting technology-related instruction and activities for students. This pilot test project correlated the results of the ST^{2}L with the results from a self-assessment technology competency and attitudes survey to establish the validity of the ST^{2}L for eighth grade students. In addition, the functioning of the items on the ST^{2}L was examined. Recommendations for improvement of the ST^{2}L assessment tool are included.

Development

The domain of content and behaviors of student technology literacy for the ST^{2}L was delineated in the following areas: essential operational skills; constructing and demonstrating knowledge; communication and collaboration; independent learning; and ethical, legal and safety issues (all of which were based on the National Educational Technology Standards from 2000). Based on the low stakes purpose of the assessment, the ST^{2}L was designed and items were written following appropriate test development processes. Technology supervisors and classroom teachers identified performance indicators with attention to state and national technology standards. Assessment items were then developed to measure the final set of indicators. Item types included multiple selection and simulated performance tasks. During the development of the tool, extensive usability analyses were conducted to ensure that the user interface and content were as clear and intuitive as possible. The final version of the tool was pilot tested according to a sound research design and sampling plan.

For the pilot test, the researchers sought a target sample that would be representative of the state. First, school districts that had participated in Florida's Enhancing Education through Technology (EETT) Leveraging Laptops (2006-07) program were invited to participate. The EETT Leveraging Laptops participants were selected because the initial group had been stratified to be representative of districts and schools across the state in size, rural and urban areas, and access to technology. A second wave of school districts was then invited in order to reach the sample stratification requirements. All participating districts were categorized as small, medium, or large, and districts were selected from five different regions of the state. Final eighth grade participants came from 13 diverse school districts, yielding a sample size of 1,561 students from 40 different schools. Examination of demographic variables (district size and region) found that the final sample satisfactorily represented the distribution of students across the state. Prior to completing the ST^2L, participants were asked to complete a self-assessment related to their uses of and attitudes toward technology. Participants also completed a post-assessment survey related to their perceptions of the ST^2L.

Results

The ST^2L was found to be a sound assessment tool for the intended purpose. Test quality criteria showed acceptable item analysis, reliability, and validity results. The mean total score for participants was 71.62%, with highest scores in the *Ethic, Safety, and Acceptable Use* section (81.18%) and lowest scores in the *Word Processing and Flowcharts* section (61.70%). The item *p*-values (item difficulty) for the entire group showed the difficulty of the items ranged from 0.05 to 1.00, and item discrimination values ranged from 0.00 to 0.61. While the ST^2L is designed as a criterion-referenced or a mastery test, these results show substantial variability across the items

and student participants. This is an indication of the diversity of technology literacy skills across the sample of student participants. In support of this interpretation, the item difficulty indices were lower (i.e., harder) for those students who were classified as being at a beginner level of computer experience. Additionally, the average time to complete the ST^2L increased when comparing the entire group with the beginners group.

Reliability was estimated for the inventory scores for each subsection and for the total score, using internal consistency reliability, or K-R 20 reliability. All reliability estimates were found to be positive and above the social science acceptable threshold (K-R 20 > 0.7), with the exception of *Software Use and File Manipulation* which had a K-R 20 = 0.67 -- which is still moderately strong (Nunnaly, 1978).

Two validity methods were designed to address construct validity. First, the subsection scores and total scores for the ST^2L were calculated across computer experience levels (the beginner, intermediate, and advanced levels were determined based on the results of a pre-survey that assessed students' computer use and attitudes). The results showed that the advanced group performed significantly better (overall and in each of the subsections) than the intermediate group on the ST^2L assessment, and the intermediate group performed significantly better than the beginners group. Second, the correlations between the sub sections of the ST^2L and the pre-survey measures of *Frequency of Computer Use*, *Attitudes towards Computers*, and *Computer Task Comfort* were calculated. As hypothesized, all of the correlations were significantly and positively correlated. These analyses confirm the construct validity of the ST^2L in relation to the stated indicators and the broad construct of technology literacy.

Recommendations

With few modifications, the ST^{2}L should be ready for a low stakes implementation. It is likely to become a useful tool for school districts for NCLB purposes and for helping students and districts target technology-related curricular needs. Recommendations include reviewing 24 unique items that were tagged in one or more quality dimensions (item difficulty values below 0.30; item discrimination below 0.20; more than 60% or less than 10% of the participants selecting an incorrect distracter; more than 10% of the student participants choosing to omit a step; and whether or not a user response was recorded). These items need to be carefully reviewed by the development team to determine what revisions, if any, should be made.

Limitations

This study was limited to 8th grade students (N=1,561) from 13 school districts in Florida during the spring of 2008. The findings may not generalize to different populations or different time periods in the academic year. It is also possible that the results are biased, based upon a difference in the schools that were asked to participate in the pilot study and those that were not. Also, although the ST^{2}L is intended to be software-independent, students may not find the interface similar enough to the software suites that they are accustomed to using in their schools and homes (due to constant changes and updates in software design).

The validation process included comparing the ability of the ST^{2}L to separate students by their ability levels related to technology. In order to make this comparison, another measure was needed to separate the students into ability levels. For this study, the results from a self-assessment were used to classify students. This self-assessment by students may not adequately represent the students' true ability. A final limitation to the current pilot test is that the evaluation

did not include variables for the socio-economic status (SES), primary language other than English spoken at home (ESOL), or special education status of the student participants.

As specified, this tool was developed for low stakes purposes. In order for it to be used for high-stakes purposes, several additional steps would need to be taken. Among these are: (a) the utilization of standardized, secure administration procedures, (b) the development of multiple, parallel test forms, (c) the development of an appropriate scoring model based on a formal standard setting process, and (d) the implementation of additional data analyses including confirmatory factor analysis and differential item functioning.

Conclusions

The Student Tool for Technology Literacy (ST^{2}L) was developed according to low stakes test purposes. For low stakes applications, the ST^{2}L is clearly satisfactory. Beginning with the framework of the NETS-S and the technology standards of Florida, an extensive, thorough process was followed for defining the indicators. The assessment items are mapped to these indicators and provide measurement of the indicators in innovative, relevant, performance-based ways. Finally, test quality criteria demonstrate reasonable item analysis, reliability, and validity results for a relatively short, criterion referenced test (CRT) or mastery assessment. With few modifications, the ST^{2}L can be ready for a low stakes implementation. The tool may be beneficial as districts report aggregated data for NCLB purposes and teachers target technology related curricular needs.

Background

The Enhancing Education Through Technology (EETT) section of the No Child Left

Behind Act (NCLB) mandates that districts demonstrate evidence of student technology literacy.

Consequently, the Office of Technology Learning and Innovation (OTLI) at the Florida

Department of Education (FLDOE) funded a competitive grant to develop an objective

assessment tool for measuring student technology literacy. The Student Tool for Technology

Literacy (ST^{2}L) is intended to be used for the low stakes purposes of evaluating technology

literacy for NCLB reporting requirements.

One of the goals of the NCLB Act is

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" (U.S. Department of Education, SEC. 2402).

Beginning in the 2006-2007 academic year, schools were required to document the

technology literacy of their students. In order to receive funds related to the NCLB initiative,

schools must also show that they "are effectively integrating technology and proven teaching

practices into instruction" (U.S. Department of Education, SEC. 2403).

Achieving technology literacy for students and teachers is a challenging task requiring

extensive planning, funding, and training. As outlined by the International Society for

Technology in Education (ISTE), basic technology skills for students and teachers include the

ability to responsibly use appropriate technology to communicate; solve problems; and access,

manage, integrate, evaluate, and create information (ISTE, 2002). Because the NCLB Act

requires documentation that students and teachers have achieved technology literacy, school

districts and states are searching for valid and reliable assessment tools. A variety of strategies

are being implemented to address the need for measuring and monitoring technology literacy,

10

including achievement rubrics, self-assessment surveys, and skill/performance inventories (Georgia Department of Education, 2008; Stansbury, 2008).

For example, some states and school districts are using the *National Educational Technology Standards* (*NETS) for Students: Achievement Rubric* and the *NETS for Students: Achievement Rubric* that were developed by the North Central Regional Educational Laboratory (Learning Point Associates, 2005). These rubrics outline the proficiency levels related to the ISTE standards. Another option for self-assessment includes the *Taking a Good Look at Instructional Technology* (TAGLIT) suite of assessment tools that are designed to help school leaders gather, analyze, and report information about how technology is used for teaching and learning. Many agree, however, that the best way to measure technology skills is through performance assessments. To fill the void of options in this area, several commercial tools have been developed that provide assessment, training, and certification of technology skills. For example, *Computer Skills for Life* by International Computer Driving License, *iSkills™ Assessment* by Electronic Testing Services, *Internet and Computer Core Certification* by Certiport, and *K to Eighth Power* all provide technology skills training and assessment. However, these commercial tools can be expensive for a district or state to implement.

There are a few state-sponsored initiatives underway to provide performance-based assessments for schools, without the price tag of a commercial product. North Carolina was the first state to develop a required set of online assessments for students.

> Beginning at grade eight, all students must meet the computer skills graduation requirement based on the competencies of the K–8 computer skills curriculum as defined in the North Carolina *Standard Course of Study* (Public Schools of North Carolina, 2008, p. 1).

Their online assessments consist of multiple choice items, as well as skill-based items. Immediate feedback is provided to the user, and performance is tracked for evaluation.

In a similar initiative, the Florida Department of Education (FLDOE) funded the

development of an online assessment tool. The Florida project was designed to help districts and

schools assess the basic technology skills of their students and to target technology-related

curricular initiatives. A specific goal of the project was to incorporate direct performance of

technology related tasks, when possible, while staying within the boundaries of an objective,

automatically-scored assessment. This report focuses on the validation of this Florida-based tool,

the Student Tool for Technology Literacy (ST^2L).

Development of the Student Tool for Technology Literacy

The Florida Department of Education (FLDOE) awarded a developmental grant to the

Pinellas County School District, who collaborated with the Florida Center for Interactive Media

(FCIM) and measurement consultants, to identify indicators of technology literacy appropriate

for the 8th-grade level and to design a tool to objectively measure student competency of each

indicator. The emphasis on a practical tool for monitoring underpins the framework for a low

stakes assessment. To emphasize the low stakes purposes, the word *tool* was used in the

assessment title in order to minimize test anxiety among teachers and students. The strategy for

developing the technology tool was as follows: 1) technology standards were identified; 2)

grade-level expectations/benchmarks for these standards were developed; 3) indicators for the

benchmarks were outlined, and 4) specific knowledge assessment items were written and

specific performance or skill assessment items were designed and programmed.

Alignment with National and State Technology Standards

In Florida, technology literacy is defined as:

The ability to responsibly use appropriate technology to communicate; solve
problems; and access, manage, integrate, evaluate, and create information to
improve learning in all subject areas and to acquire lifelong knowledge and skills
in the 21st century (Florida Department of Education, 2003).

The development of the 8th-grade student technology standards for Florida was based on the National Education Technology Standards for Students (NETS-S) established by the International Society for Technology in Education (ISTE, 2007), the standards of other states, and standards used by individual school districts within Florida. The student technology categories and standards were initially developed at the EETT statewide meeting held on August 12 and 13, 2005. Subsequently, a draft list of benchmarks was created at a follow-up meeting on August 15, 2005. Next, the FLDOE incorporated the technology standards into the Sunshine State Standards (SSS), for each curricular area by grade level.

Identifying Technology-Related Indicators

After the standards and benchmarks were identified, an Advisory Group and an Expert Review Panel worked jointly to create the indicators for the ST^{2}L. Members of the Advisory Group were recruited at the September 19, 2005, meeting of the Florida Council of Instructional Technology Leaders (FCITL). First, the draft of the student technology standards and benchmarks was reviewed. Next, the Advisory Group drafted indicators for the standards and benchmarks. Thus, this meeting resulted in an expanded matrix of standards, benchmarks, and their associated indicators. The following week, all members of the FCITL group were contacted via email and asked to join the Advisory Group for the remainder of the project. To ensure representation from a cross-section of school districts within the state, additional instructional technology experts were contacted from districts that were not represented at the FCITL meeting.

An Expert Review Panel was also formed, consisting of selected technology experts with experience in one or more of the following areas: instructional technology, student assessment, information science, and measurement. On October 26-27, 2005, the Expert Review Panel met with members of the FLDOE and FCIM. Feedback on the standards and benchmarks from

13

districts across the state was utilized to create draft indicators. In November, 2005, staff from

FLDOE met with the measurement team at FCIM to complete the indicator review. Decisions

were based on the measurability of the indicators and the types of items that could be developed

in relation to each indicator. The full and final set of indicators on which the ST^2L was based is

included in Appendix A.

Finalizing Technology-Related Indicators

The next step was to develop a statewide survey and send it out to the Advisory Group

members, each of whom was asked to forward the survey link to at least 50 professional middle

school educators or library media specialists. It was requested that this review group of

professionals be different from the original group who had established the first set of standards

and benchmarks. The purpose of the survey was to solicit feedback about the appropriateness of

the indicators before assessment items were developed. Each indicator had two evaluation

questions devoted to it: 1) the grade level appropriateness of the indicator, and 2) the clarity of

the wording. The participants had radio button choices with a text box for explanatory

comments.

Data from the survey were compiled in late December 2005 and January 2006. Results

indicated that the majority of the indicators were appropriate expectations of technology-related

knowledge for an 8th-grader. Based on the survey responses and the comments regarding clarity

and appropriateness, the indicators were revised.

The indicators were finalized at the second Expert Review Panel meeting held on

February 2, 2006 in Pinellas County. The panel decided to remove all indicators pertaining to

hardware-specific use, because hardware-specific test items that would be relevant to a broad

array of schools and settings would be too difficult to implement. The remaining indicators were

grouped into software types (e.g., word processing, spreadsheet, etc.), and simulated item examples were created in order to measure each indicator. The final set of indicators is included in Appendix A.

Item Writing

Teachers and media specialists from the Leon County School District drafted items. All participants were familiar with instructional technology and the capabilities of middle school students. The Expert Review Panel decided that the indicators would not be revised or removed from the assessment unless (during the item-writing phase) the measurement team determined that some indicators could not be assessed. Thus, the item-writing team was instructed to only write items, and if necessary, to send any indicators that could not be assessed to the measurement team for review. Item writing began in February, and by the end of the month, all sections had been completed except the section related to *Ethical, Legal, and Safety Issues*. These items were written in collaboration with library media specialists in June.

A primary goal for the ST^{2}L was the measurement of technology skills through direct examinee action, whenever possible. It was regarded as critical that actual skill use, and not simply knowledge, be assessed. Thus, a variety of item types were developed, including selected-response items as well as performance-based tasks. Altogether, the final tool consists of 107 items: 40 are selected-response and 67 are steps within performance-based tasks.

The selected-response item types include text-based multiple-choice and true/false items and text-based multiple-choice items with graphics and image map selections (where examinees respond to a prompt by clicking directly on some part of an image). The performance-based tasks 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* (ITTS), which was the technology assessment instrument previously developed for teachers by the FLDOE (Harmes, Barron & Kemker, 2007). An example of the screen layout for a simulated task can be found in Appendix H. 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 multi-step item indicator measurement format was used when the wording of an indicator implied a series of steps. For example, an indicator may require an examinee to *cut and paste* a sample text, two actions which are frequently paired in computer use. Such an indicator might be assessed by two separate steps within one individual item, wherein the steps would be given equal proportions of the single item score that formed the measurement of a given indicator.

As in the ITTS, the correct completion of any given ST^2L item is 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 in a number of ways. Once an examinee performs an action that would constitute a response to a performance-based item, a *Submit Step?* dialogue box appears. At that point, the examinee must choose to either confirm or cancel the action just taken. Additionally, a *Skip Step* button is available to allow the examinee to bypass any given 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 then displays the screen image needed to correctly complete the next item.

The tool was programmed in *Flash* and designed to be administered over the Internet. Brief instructions and a few practice items are provided to give the examinee basic information

on how to take the test. Examinees are required to complete the practice items before beginning

the actual tool (see Proctor's Guide in Appendix C).

Usability Testing

After the programming was complete for approximately two-thirds of the indicators, initial usability tests were conducted in Pinellas County.

Usability Test Design

Groups of approximately seven students were selected by the participating schools' technology instructors (based on their level of technological literacy) in order to get a representative sample of students from basic to advanced levels. Each group completed the assessment and participated in a follow-up discussion group session. A total of 104 students participated in this phase of the usability test.

Students began using the ST^2L tool simultaneously at individual computers and took an average of 30 minutes to complete the three modules. Students were encouraged to ask questions and express feedback during the process. An on-screen feedback button was also available for students to record their comments and questions.

Members of the measurement team observed the students, and the project director led the discussion group after all of the students had completed their items. This discussion group followed a basic interview protocol, which consisted of open-ended questions to stimulate conversation and to collect feedback related to content and usability issues. Discussion group sessions with the students lasted from 40 to 65 minutes. An audio recording of the discussion sessions was transcribed as a reference for future revisions of the tool.

Interviews were also conducted with at least one technology teacher/coordinator at each school and the principal at one school. These interviews, which ranged from five 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 Group Discussions

During the group discussions, some students were more vocal than others; however, the responses to general questions such as, *Did you like the tool?,* were consistently positive. Every student who spoke had a favorable view of the tool as well as the experience of being part of the usability testing. In addition, virtually every student found the tool's interface easy to use.

Students were also provided the opportunity to submit feedback about individual items during the ST^{2}L assessment (via a feedback button). Data from students' feedback were collected, coded, and analyzed in order to derive a set of themes that related to students' expressed concerns about the tool. Overall, the students expressed high levels of satisfaction with the ST^{2}L. There were, however, two commonly cited concerns. There were 57 comments regarding students' confusion over the assigned tasks, and more than half of these were directed to items in the Browser section. Similarly, 38 students were concerned with the clarity of the wording of some instructions. Fifty-seven comments from students expressed a desire to increase the instructive use of the tool. However, the central purpose of the tool is assessment, and interlacing the instrument with didactic sections would interfere with the natural flow of the simulations. Forty students expressed concerns about the user interface, and twenty-three students expressed concerns about the authenticity of the tasks in the simulations.

Descriptive Data

The median score for the usability assessment was 27.9 out of 36 (with a standard deviation of 4). Overall, 75% of the test takers received a 70% or greater on the assessment, thus indicating a relatively easy set of items. The percentile ranks are as follows:

- The mean score at the 25th percentile was 25.25

- At the 50th percentile, the mean score was 28.00

- At the 75th percentile, the mean score was 31.00

Item Analysis

For the entire group of examinees, difficulty values ranged from .17 to .97. Ten items had *p*-values ranging between .90 and .99, ten items had *p*-values from .80 to .89, nine items had *p*-values from .70 to .79, and seven items had *p*-values below .70. This indicates that for this group of examinees, the assessment was relatively easy.

An analysis of the discrimination values of items answered incorrectly by more than half the examinees during the usability testing demonstrated these items to be truly difficult measures of advanced skills, generally answered correctly by high-level users. Discrimination values (*D*-values) ranged from -0.02 to 0.46. A number of the items had low *D*-values, including one that had a value just below zero. However, most of the small *D*-values were accompanied by very high *p*-values. An exception to this case was question 6 of the Software Use section, which had an item difficulty value of .44 and a *D* value of .07. In general, the high *p*-values and low *D*-values, which were seen in this usability test of the tool, are relatively typical of mastery testing, as opposed to norm-referenced testing (NRT) where both a broader range of *p*-values, including lower values and high *D*-values would be expected.

Reliability

The internal consistency reliability estimate for scores on the entire tool (during the usability test) was .74. This is considered (by measurement criteria) as acceptable (Nunnaly, 1978). As discussed earlier, this assessment performs like a CRT or mastery test, in which examinees are generally expected to achieve at a level around 80%. Within this framework, there

will be 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 an NRT.

Pilot Study Design

A pilot study to validate the ST^{2}L was conducted in the spring of 2008 with 1,561 students from 40 different schools in 13 diverse school districts in the state of Florida. The districts that volunteered were mapped by region (as defined by the FLDOE Office of School Improvement); these school districts represented all five regions of the state. Figure 1 illustrates locations of the participating districts, each of which is marked with a star (Broward, Duval, Escambia, Jackson, Lake, Lee, Madison, Manatee, Miami-Dade, Palm Beach, Pinellas, Polk, and St. Johns).
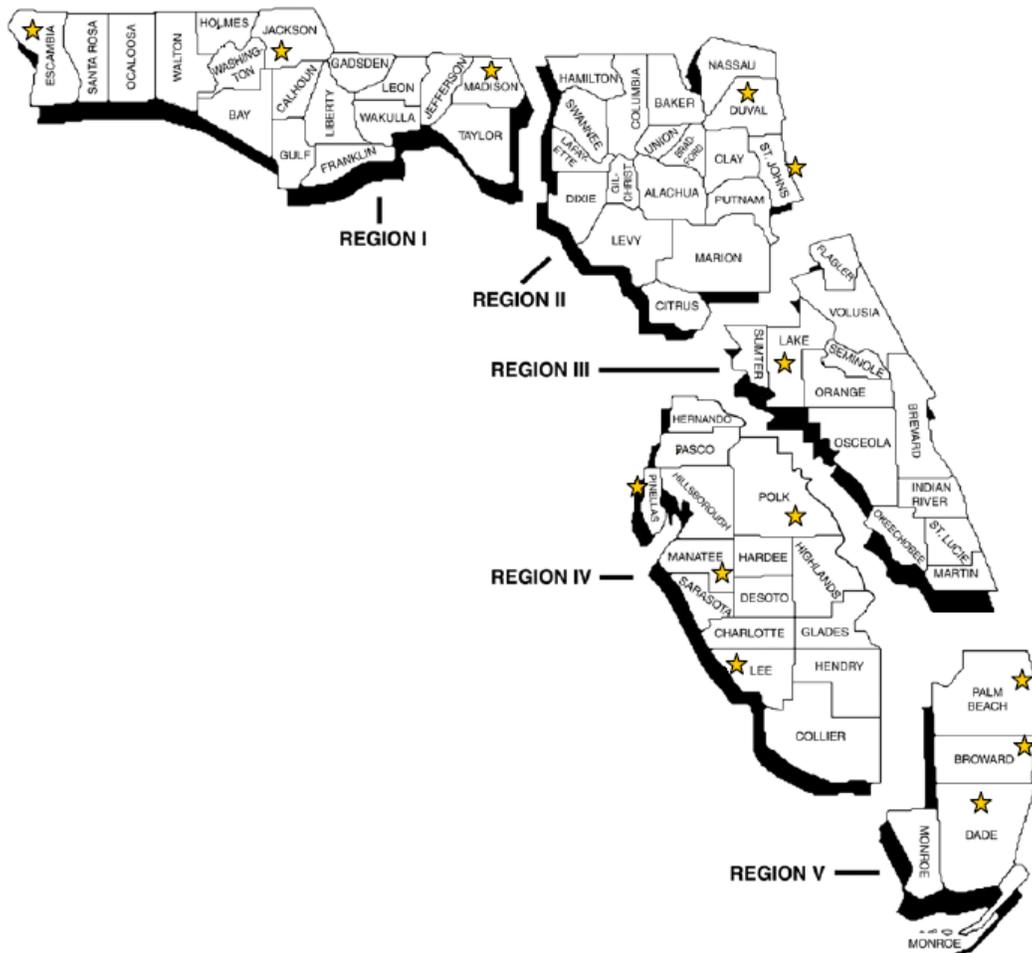
Figure 1. State map designating participating districts by region.

The sample was further categorized to ensure representation from small, medium, and large districts. Large school districts were defined as those that had more than 100,000 students; medium school districts had greater than 40,000 and less than 100,000 students; and small districts had less than 40,000 students. Five large school districts, four medium school districts, and four small school districts participated in the pilot study. All student participants were 8th graders and were recruited from middle/junior/high schools.

Initially, school districts that participated in the Enhancing Education through Technology (EETT) Leveraging Laptops (2006-07) program in the state of Florida were invited to participate (Escambia, Lake, Madison, Manatee, Miami-Dade, Pinellas, and Polk), and then a second wave of school districts were invited based on the sample stratification requirements (Broward, Duval, Jackson, Lee, Palm Beach, Polk, and St. Johns). The EETT Leveraging Laptops participants were selected because the initial group had been stratified across the state by size, rural and urban areas, and access to technology. A letter (see Appendix B) was sent to each school district outlining the program's intent, an overview of the requirements for administration, and an information sheet requesting details about potential participants.

Schools districts that volunteered were asked to submit a variety of information about each of the participating schools (i.e., the percentage of students on free or reduced lunch, an approximate number of student participants, the degree of technology exposure, an estimated number of sessions for completion, headphone availability, and whether the administration of the tool would occur in a classroom or computer lab). The researchers used the collected information about potential student participants from the school districts to inform the sampling process.

To facilitate the administration of the instrument, a proctor guide (see Appendix C) was created, and a web conference was scheduled for proctors and technology leaders in each of the

districts. The proctor guide outlined administrative and technical requirements. The information was presented in the web conference, and school districts were provided with the opportunity to ask specific questions of a designated representative of the pilot study team.

The following sequence was used for the pilot test: (1) completion of a pre-survey related to computer use and attitudes, (2) completion of the ST^2L in one of four randomly assigned section sequences, and (3) completion of a post-survey. The pre-survey 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 socio-economic status (see Appendix D). The remaining items on the pre-survey were modeled after the items from the context questionnaires about the students' use of and access to technology from the *Programme for International Student Assessment (PISA)*. The PISA items were modified slightly to match the indicators on the ST^2L (PISA, 2003). The PISA questionnaire has been rigorously analyzed to demonstrate both reliability and validity across diverse international populations (PISA, 2003).

The ST^2L was organized purposefully into four different sequences in an effort to (1) control for an ordering effect and (2) obtain valuable data from potential non-completers (e.g., those who did not complete all sections of the ST^2L). The post survey included five items designed to measure the participants' satisfaction, their perception of how well the ST^2L measured their ability, their perceived level of difficulty, and whether the tool reflected skills that they had learned at school (see Appendix E).

After completion of the ST^2L administration, proctors were asked to complete a survey in order 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 available to student participants (e.g., headsets), and an overall evaluation

of the proctors' perception of administrative issues. This information also allowed the

identification of any systematic errors. The proctor survey is shown in Appendix F.

Results

Participants

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

Table 1 illustrates the size category (small, medium, and large) of the school districts involved in the pilot program. Approximately 53% of the students in Florida are in large school districts, 26.24% within medium-sized school districts, and 20.57% in small-sized school districts. The final sample represents a satisfactory representation of medium-sized school districts (27.5%); however, the final sample fell short in representing large school districts and has over representation from small school districts at 32.72% and 39.79%, respectively.

Table 1. *Comparison of the population and obtained and final sample by district sizes*.

District size	Students in Florida		Obtained Sample		Final Sample	
	Students	**Percent**	**Students**	**Percent**	**Students**	**Percent**
Large	1,416,837	53.19	530	33.95	495	32.72
Medium	699,080	26.24	426	27.29	416	27.50
Small	548,056	20.57	605	38.76	602	39.79
Total	2,663,973	100.00	1,561	100.00	1,513	100.00

Another factor used to determine the degree of success in obtaining a sample representative of the state was the regions in which the students reside. Table 2 illustrates the regional classification (Panhandle, Crown, East Central, West Central and South) of the school districts involved in the pilot program. The regions of the Florida have the following total student

26

enrollment: Panhandle region (9.37%), Crown (26%), 19% East Central (19%), West Central (25%), and South (33%). The final sample represents a satisfactory representation of students within the Crown and East Central regions with 15.33% and 21.15%, respectively. The final sample from the South region was under represented at 14.54%. Finally, the sample was over represented with participants from the Panhandle and West Central regions at 16.79% and 32.19%, respectively.

Table 2. *Comparison of the population and obtained and final sample by regions.*

	Students in Florida		Obtained Sample		Final Sample	
Region	**Students**	**Percent**	**Students**	**Percent**	**Students**	**Percent**
Panhandle	249,645	9.37	258	16.53	254	16.79
Crown	348,760	13.09	233	14.93	232	15.33
East Central	514,402	19.31	321	20.56	320	21.15
West Central	664,911	24.96	494	31.65	487	32.19
South	886,255	33.27	255	16.34	220	14.54
Total	2,663,973	100.00	1,561	100.00	1,513	100.00

In order to validate the function of the ST^{2}L tool for populations of students with diverse abilities, students were grouped into levels of technology literacy: beginner, intermediate, advanced experience levels. In order to assess these abilities, an instrument was developed based on the computer instrument used by PISA in 2003. This survey was administered as part of the pre-survey before the students began the ST^{2}L assessment. To differentiate the natural abilities of students completing the ST^{2}L, a composite score was created from the computer task comfort level and the frequency of use sections of the pre-survey for each of the participants in the obtained sample. The composite measure was internally consistent with Cronbach's alpha calculated at α=.87 for these data. The scores were split into quartiles in which the fourth quartile was designated as advanced (top 25%), the first quartile was designated as beginner (bottom 25%), and the second and third quartiles were used to represent intermediate computing skills

(middle 50%). Table 3 provides the distribution of students across these computer experience classifications.

Table 3. *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

Table 4 shows the distribution of students across gender and ethnicity groups for the eighth grade students in Florida, the obtained, and the final sample. As summarized in Table 4, the distribution of males to females was similar. Both the obtained sample and the final sample were similar in percentages of each ethnic group; however, when compared to the 8th grade population in Florida schools, the samples overrepresented White students and underrepresented Hispanic students.

Table 4. *Comparison of population with obtained and final sample by gender and ethnicity*.

	8th Grade Students in Florida		Obtained Sample		Final Sample	
	Students	**Percent**	**Students**	**Percent**	**Students**	**Percent**
Gender						
Female	9,7846	49.05	783	50.16	766	50.63
Male	101,498	50.88	778	49.84	747	49.37
Ethnicity						
White	95,058	47.65	878	56.25	854	56.44
Black	45,837	22.98	323	20.69	318	21.02
Hispanic	47,877	24.00	233	14.93	218	14.41
Asian	4,456	2.23	45	2.88	43	2.84
Other	6,263	3.14	82	5.25	80	5.29

These analyses suggest that the obtained sample is similar to the demographics of students in Florida on some demographic characteristics. Though the stratified sample deviates from the population norms on some demographic characteristics (region and district size, White and Hispanic students), 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.

Descriptive Statistics of Pre-Survey

To provide an overview of the student participants of the pilot study, the response frequencies, averages, standard deviations and reliability measures of the pre-survey questions are provided. The pre-survey included items that measure a participant's level of comfort with various computer tasks, frequency of computer use, and overall attitudes towards computer technology. Exploratory factor analyses were conducted in each of the domains to explore the internal structure of the factors. The factor analyses were executed with a simple random sample of 50% of the total participants or *n*=780, which is approximately a 20:1 participant-to-item ratio.

Frequency of computer use. A total of 14 different items constitute this section. Exploratory factor analyses were conducted to investigate the structure of the item loadings using an orthogonal rotation (Promax) as the factors are anticipated to be correlated. The optimal model resulted in a two factor model using the proportion and Kaiser criteria. Additional factor models specifying two, three, and four factors were explored. The items did not meaningfully load or have simple structures, and thus, it was decided to combine the individual items logically. The combined internal consistency reliability measured using Cronbach's alpha for the total dataset is 0.79. As noted by Nunnaly (1978), this level of internal consistency is more than

29

acceptable for the social sciences. These items were combined to form a composite measure to represent the *Frequency of Computer Use*.

Table 5 shows the number and percentage of student participants related to the duration of their computer use. The majority of student participants (68.74%) report using a computer for more than five years, followed by 18.97% using computers from three to five years, 10.05% using computers for one to three years, and 2.25% using computers for less than one year. As depicted in Table 5, the vast majority of student participants have been using computers for more than three years.

Table 5. *Distribution of students using computers over different time-periods*.

Duration of Use	Students	Percent
Less than one year	32	2.12
One to three years	151	9.98
Three to five years	288	19.04
More than five years	1042	68.87
Total	1,513	100.00

Table 6 illustrates the sample population's use of computer technology by location. As illustrated, 65.70% of students use computers at their home almost every day of the week, and 41.51% of students report everyday use at school. Approximately 70% of the student participants indicated using a computer at least a few times each week at school. Less than 2% of the sample reported not using a computer at school at all.

Table 6. *Frequency, mean and standard deviation of computer use by location*.

Location	M	SD	1	2	3	4	5
At Home	4.38	1.07	4.89	3.24	6.48	19.70	65.70
At School	3.98	1.07	1.78	9.19	19.89	27.63	41.51
At Other Places	2.82	1.21	15.40	27.23	27.56	19.56	10.24

Table 6. *Frequency, mean and standard deviation of computer use by location.*

Location	M	SD	1	2	3	4	5

1=Never
2=Less than once a month
3=Between once a week and once a month
4=A few times each week
5=Almost every day

Table 7 illustrates the student participants' frequency of computer use for specific purposes. Sixty-five percent or more of the participants report using computers for electronic communication, to download files over the Internet, and to look up information on the Internet at least a few times a week. In contrast, sixty-five percent or more of the student participants report using a computer for spreadsheets or video editing never or less than once a month.

Table 7. *Frequency of computer use for specific purposes.*

Computer Use Task	M	SD	1	2	3	4	5
The internet to look up information about people, things, or ideas?	3.89	1.08	2.71	9.58	19.63	31.92	36.15
Games on a computer?	3.58	1.32	9.32	14.14	18.11	25.71	32.72
Word processing (e.g., Microsoft Word, Works)?	3.41	1.21	7.87	15.20	26.70	28.09	22.14
Drawing, painting or graphics programs on a computer?	2.87	1.32	18.57	24.45	23.46	18.70	14.81
Presentation software (e.g., Microsoft PowerPoint, Apple Keynote)?	2.81	1.26	18.31	24.32	27.36	18.18	11.83
Spreadsheets (e.g., Microsoft Excel, Works)?	2.14	1.20	40.65	25.31	18.31	10.71	5.02
Video Editing (e.g., Moviemaker, iMovie)?	2.14	1.31	45.80	20.42	15.66	10.31	7.80
Electronic communication (e.g., e- mail, instant messaging)?	3.98	1.38	11.37	6.28	9.52	18.37	54.46
The Internet to download software, music, or files?	3.95	1.29	7.87	8.53	12.23	23.13	48.25
The computer to help you learn school material?	3.14	1.32	14.94	17.38	24.98	24.32	18.37

1=Never
2=Less than once a month
3=Between once a week and once a month
4=A few times each week
5=Almost every day

Computer task comfort. The computer task comfort section of the survey is composed of 19 different items. Following the same procedure, to investigate the factor structure, exploratory factor analyses were conducted using an orthogonal rotation (Promax). The optimal model resulted in a three factor model using the proportion and Kaiser criteria. The structure was indiscernible, and thus, factor models were explored with two, four, and five factors specified. The items did not result in a meaningful factor loading, and thus, the items were logically assigned to a composite titled *Computer Task Comfort*. The internal consistency reliability (Cronbach's alpha) was very high for these data at .87.

Table 8 shows the item average, standard deviation, and response frequencies for participants' comfort across a variety of computer tasks. The tasks with which 90% or more students reported being most comfortable and able to do very well by themselves include: open a file, scroll up/down a screen, save a computer document or file, print a document or file, delete a file or document, search the Internet for a topic, and play computer games. Conversely less than 40% of students were as comfortable with the tasks: use software to find and get rid of computer viruses, insert a formula into a spreadsheet, use a spreadsheet to make a chart, and edit a video.

Table 8. *Computer task comfort*.

Computer Task	M	SD	1	2	3	4
Use software to find and get rid of computer viruses	2.83	0.92	7.67	29.28	35.23	27.83
Open a file	3.91	0.38	0.59	0.99	5.62	92.80
Create/edit a document	3.81	0.54	1.45	2.71	9.12	86.72
Scroll up and down a document on the screen	3.92	0.39	1.06	1.06	2.38	95.51
Save a computer document or file	3.94	0.31	0.40	0.99	2.58	96.03
Print a computer document or file	3.91	0.41	0.86	1.59	3.17	94.38
Delete a computer document or file	3.90	0.39	0.46	1.72	5.02	92.80
Move files from one place	3.63	0.67	1.32	6.54	20.36	71.78

Table 8. *Computer task comfort.*

Computer Task	M	SD	1	2	3	4
to another on a computer						
Search for a topic on the Internet	3.94	0.33	0.59	0.73	2.84	95.84
Copy or download files or music from the Internet	3.70	0.63	1.06	5.88	15.33	77.73
Attach a file to an e- mail message	3.44	0.84	3.97	11.17	21.48	63.38
Use Insert a formula into a spreadsheet	2.76	1.09	19.50	16.13	33.11	31.26
Use a spreadsheet to make a chart	2.88	1.09	16.52	16.59	29.54	37.34
Create a presentation	3.64	0.73	3.30	5.49	14.81	76.40
Play computer games	3.88	0.45	1.06	1.98	4.49	92.47
Create a multimedia presentation	3.38	0.80	3.30	10.18	32.12	54.40
Change the size of a picture	3.78	0.56	1.32	3.11	11.96	83.61
Write and send email	3.83	0.56	1.65	3.64	4.63	90.09
Edit a video	2.97	0.92	6.54	23.86	35.69	33.91

1=I don't know what this means
2=I know what this means but I cannot do it
3=I can do this with help from someone
4=I can do this very well by myself

Attitudes towards computers. The attitudes towards computers factor is composed of five different items. Exploratory factor analyses were conducted using an orthogonal rotation (Promax) to examine the structure and item loadings. The optimal model resulted in a one factor model using the proportion and Kaiser criteria. The one factor was labeled *Attitudes towards Computers* and has a moderate degree of internal consistency reliability as measured by Cronbach's alpha at $\alpha = .62$. Table 9 shows the item average, standard deviation, and response frequencies for student participants' attitudes towards computers.

Approximately 90% or more of the student participants agree that using computers helps them with their schoolwork (92.13%), using a computer is really fun (95.64%), and that it is

important for them to work with a computer (89.49%). Seventy percent or more student

participants also agreed they lose track of their time when using a computer (72.9%) and that

people ask them for help with their computer (73.23%). The only item with a mean lower than

the mid score of the scale (< 3) was *People ask me for help with their computer*.

Table 9. *Attitudes towards computers*.

Attitude toward computers	M	SD	1	2	3	4
Using a computer helps me with my schoolwork.	3.42	0.69	1.92	5.82	40.85	51.42
It is important to me to work with a computer.	3.40	0.71	1.32	9.05	37.74	51.88
I think playing or working with a computer is really fun.	3.62	0.60	1.12	2.97	28.42	67.48
I lose track of time when I am using a computer.	3.10	0.97	7.14	20.89	26.77	45.21
People ask me for help with their computer.	2.99	0.86	5.68	20.75	42.63	30.93

1=Strongly disagree
2=Disagree
3=Agree
4=Strongly agree

Descriptive Statistics and Statistical Differences of ST^{2}L

Subscores and mean scores for all sections (percent correct) are presented by

demographic variables in Table 10. Note that only *n*=1,335 student participants completed

enough of the items in each section to earn a composite score. Thus, only those student

participants are included in the descriptive analysis.

Overall, across sections and demographic conditions, the average score was 71.62%.

Students performed best on the *Ethics, Safety and Acceptable Use* (81.18%) section, followed by

Browser Use and Email (80.97%) section. The lowest scores were in the *Graphics, Presentation,

and Video Editing* (57.10%) and *Word Processing and Flowcharts* (61.70%) sections of the

ST^{2}L. These results demonstrate that the students' strongest competencies were the acceptable

and ethical uses of computer within school environments.

Significant differences were detected on overall performance based on region ($F(4, 1330) = 22.54$, $p < .01$), gender ($F(1, 1333) = 37.02$, $p < .01$), and ethnicity ($F(4, 1330) = 20.62$, $p < .01$). No differences were detected across district size ($F(2, 1286) = 1.71$, $p = .18$). Student participants from the South region performed significantly better than all other regions, while student participants from the Panhandle region performed significantly lower than all other regions as confirmed by a Tukey follow-up procedure. Females, overall, scored approximately 5 points higher than males on the ST^2L. Finally, Asians and Whites performed significantly better than students classified as Black or Other. Those students classified as Hispanic performed significantly better than those classified as Black. Overall, these findings provide strong evidence that the overall performance scores varied a great deal across the demographic variables.

Table 10. *Sub scores and mean score by 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

Table 10. *Sub scores and mean score by demographics.*

	1	2	3	4	5	6	Total*
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

1=Software Use and File Manipulation
2=Ethics, Safety and Acceptable Use
3=Graphics, Presentation, and Video Editing
4=Spreadsheets
5=Browser Use and Email
6=Word Processing and Flowcharts
*N=1,335 for Total

ST^2L Test Quality and Characteristics

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. The analysis of individual items usually begins with calculating its difficulty (*p*). The difficulty of an item is measured by the percent of examinees able to respond correctly to the item (Crocker & Algina, 1986). Values for item difficulty range between 0 and 1; the items with the lowest *p*-value are the most difficult. After the difficulty of an item has been determined, then the discrimination (*D*) of the item can be calculated. 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. For tests with a large number of items (107) that are scored either right or wrong, the discrimination of the item is measured by its point biserial correlation (Crocker & Algina, 1986). Item discrimination values range between 0 and 1, with higher scores indicating better discrimination.

In order to calculate item discrimination for every item in the pilot study, participants were filtered to include only those who completed all sections of the assessment (*n*=1,335). For calculating item difficulty and item discrimination, the total score for the assessment was

computed by summing all of the correct items. This is a different measure than was used to score

each section of the test in order to provide feedback. Because the purpose of the ST^{2}L tool is to

provide guidance and structure for identifying the technology literacy skills that students need to

learn, the functioning of the items was examined by ability group as well as for the whole group

of participants in the pilot test.

Students were separated into four quartiles or ability groups based on their total score on

the ST^{2}L for the item analysis. First, item statistics are reported for the entire group of pilot test

participants (see Table 11). The percent correct are reported by quartile to see how the tool

functions at different computer literacy levels. Students with the lowest computer literacy ability

are in quartile 1, while students with the highest computer literacy ability are in quartile 4. The

second column, Item Type, indicates whether each item was a selected response multiple choice

item (MC), a dichotomous Yes/ No or A/B item, or a step in a performance-based simulation

task (SIM).

Table 11. *Item Difficulty, Discrimination, and Percent Correct by Quartile for Entire Group (N=1,335).*

Item	Type	Item Difficulty p	Item Discrimination D	Percent Correct by Quartile			
				Q1	Q2	Q3	Q4
Software Use and File Manipulation							
SoftwareUse1	MC	0.73	0.39	46.9	67.9	82.5	93.9
SoftwareUse2	MC	0.80	0.36	59.5	75.7	87.9	94.7
SoftwareUse3	MC	0.89	0.31	76.7	87.5	94.1	97.4
SoftwareUse4	MC	0.63	0.32	40.5	57.8	68.1	84.5
SoftwareUse5	MC	0.91	0.32	79.0	88.5	97.2	97.7
SoftwareUse6	Y/N	0.86	0.19	78.4	80.7	89.0	95.3
SoftwareUse7	Y/N	0.44	0.14	37.9	30.4	43.8	62.9
SoftwareUse8	Y/N	0.62	0.34	38.2	56.1	67.8	84.5
SoftwareUse9	Y/N	0.31	0.19	21.6	22.0	30.5	49.7
SoftwareUse10	MC	0.95	0.34	87.2	95.6	98.3	99.7
SoftwareUse11	MC	0.88	0.32	72.6	90.5	92.1	96.2
SoftwareUse12	MC	0.95	0.27	86.9	96.0	96.6	98.8
SoftwareUse13	MC	0.43	0.03	45.2	37.8	41.0	47.1
SoftwareUse14	MC	0.78	0.32	57.7	77.4	81.6	95.6
SoftwareUse15	A/B	0.89	0.37	74.1	91.2	92.9	97.1

Table 11. *Item Difficulty, Discrimination, and Percent Correct by Quartile for Entire Group (N=1,335).*

Item	Type	Item Difficulty *p*	Item Discrimination D	Q1	Q2	Q3	Q4
FileManipulation1	SIM	0.74	0.41	47.8	69.3	85.6	93.9
FileManipulation2	SIM	0.78	0.43	51.9	73.3	90.4	94.4
FileManipulation3	SIM	0.88	0.43	69.4	89.5	95.5	98.3
FileManipulation4	SIM	0.68	0.41	39.9	65.9	75.7	89.8
Ethics, Safety and Acceptable Use							
EthicsSafeUse1	A/B	0.89	0.08	84.8	91.2	90.4	91.5
EthicsSafeUse2	MC	0.81	0.40	58.6	79.4	89.3	98.3
EthicsSafeUse3	Y/N (A/B)	0.64	0.16	50.7	60.8	68.6	73.4
EthicsSafeUse4	MC	0.90	0.52	71.7	92.2	97.7	100.0
EthicsSafeUse5	MC	0.91	0.50	72.6	93.2	98.9	98.5
EthicsSafeUse6	MC	0.92	0.50	74.9	95.3	97.7	100.0
EthicsSafeUse7	A/B	0.86	0.14	79.9	87.8	88.1	88.0
EthicsSafeUse8	A/B	0.93	0.25	84.3	93.2	96.6	96.2
EthicsSafeUse9	A/B	0.95	0.30	87.2	94.9	98.9	99.1
EthicsSafeUse10	A/B	0.91	0.42	75.8	93.6	96.6	98.8
EthicsSafeUse11	A/B	0.78	0.43	52.8	75.7	87.6	94.7
EthicsSafeUse12	A/B	0.96	0.39	85.7	99.0	99.2	99.7
EthicsSafeUse13	A/B	0.82	0.44	59.2	80.7	89.8	97.4
EthicsSafeUse14	A/B	0.94	0.31	86.0	95.3	97.5	98.5
EthicsSafeUse15	A/B	0.94	0.46	81.6	97.0	98.9	99.7
EthicsSafeUse16	MC	0.18	0.03	16.9	15.5	17.8	22.2
EthicsSafeUse17	A/B	0.95	0.35	83.1	99.0	97.7	100.0
EthicsSafeUse18	A/B	0.89	0.32	75.2	86.8	94.6	99.1
EthicsSafeUse19	A/B	0.94	0.48	77.6	98.7	99.2	99.7
EthicsSafeUse20	A/B	0.70	0.26	53.6	64.9	75.4	85.7
EthicsSafeUse21	A/B	0.67	0.29	46.9	62.2	71.2	86.0
EthicsSafeUse22	MC	0.89	0.47	69.1	91.6	95.8	98.3
EthicsSafeUse23	MC	0.87	0.43	68.5	87.8	93.8	98.0
EthicsSafeUse24	MC	0.83	0.47	57.1	84.8	93.2	95.9
EthicsSafeUse25	MC	0.91	0.50	73.2	93.9	97.7	99.1
Graphics, Presentation, and Video Editing							
Graphics1	SIM	0.75	0.43	47.2	68.9	87.3	96.2
Graphics2	SIM	0.13	0.24	3.2	6.4	15.8	26.6
Graphics3	SIM	0.44	0.33	18.7	39.2	50.6	65.2
Graphics4	SIM	0.75	0.39	52.5	78.7	80.2	90.4
Video1	SIM	0.84	0.29	70.0	85.5	88.4	92.7
Video2	SIM	0.30	0.28	15.2	18.6	33.1	51.2
Video3	SIM	0.24	0.38	3.5	10.1	24.0	55.6
Video4	SIM	0.50	0.48	16.0	39.5	57.1	85.7
Video5	SIM	0.33	0.39	8.5	22.3	37.6	61.7
Presentation1	SIM	0.95	0.49	82.8	99.0	99.4	99.7
Presentation2	SIM	0.81	0.37	60.9	82.1	89.6	91.2
Presentation3	SIM	0.69	0.48	37.3	62.2	80.8	95.6
Presentation4	SIM	0.97	0.45	87.5	99.3	99.7	100.0

Table 11. *Item Difficulty, Discrimination, and Percent Correct by Quartile for Entire Group (N=1,335).*

Item	Type	Item Difficulty *p*	Item Discrimination D	Percent Correct by Quartile			
				Q1	Q2	Q3	Q4
Spreadsheets							
SpreadSheet11	SIM	0.92	0.47	76.1	96.3	96.3	98.3
SpreadSheet12	SIM	0.94	0.48	83.1	98.0	97.7	99.4
SpreadSheet13	SIM	0.49	0.38	22.5	44.3	56.5	73.1
SpreadSheet14	SIM	0.85	0.56	57.7	89.2	95.2	98.8
SpreadSheet15	SIM	0.75	0.40	52.5	79.7	81.1	87.7
SpreadSheet21	SIM	0.86	0.51	62.7	89.5	94.1	98.0
SpreadSheet22	SIM	0.05	0.17	0.9	2.4	4.0	12.9
SpreadSheet23	SIM	0.70	0.53	32.1	72.6	82.5	92.1
SpreadSheet24	SIM	0.78	0.61	41.4	82.1	92.4	98.0
SpreadSheet25	SIM	0.92	0.52	75.8	97.0	98.0	99.1
SpreadSheet26	SIM	0.91	0.54	71.1	94.3	97.7	100.0
SpreadSheet27	SIM	0.81	0.50	54.8	82.4	91.2	93.9
SpreadSheet28	SIM	0.86	0.53	62.1	91.2	93.5	99.1
SpreadSheet29	SIM	0.33	0.43	5.8	18.9	35.9	68.4
Browser Use and Email							
InternetSearching1	SIM	0.83	0.33	68.2	85.8	88.4	90.4
InternetSearching2	SIM	0.94	0.49	79.0	97.3	99.4	100.0
InternetSearching3	SIM	0.88	0.35	73.2	91.9	94.6	92.7
InternetSearching4	SIM	0.96	0.57	84.6	99.7	100.0	100.0
InternetSearching5	SIM	0.92	0.49	75.2	96.0	96.9	98.8
InternetSearching6	SIM	0.13	0.27	2.3	4.7	13.0	29.8
InternetSearching7	SIM	0.78	0.46	49.9	81.4	88.4	94.2
InternetSearching8	SIM	0.91	0.57	69.1	96.6	99.2	99.7
EmailUse11	SIM	0.76	0.41	51.9	72.0	83.9	95.0
EmailUse12	SIM	0.91	0.40	78.7	96.0	93.5	96.8
EmailUse13	SIM	0.90	0.48	72.3	95.6	96.3	97.4
EmailUse14	SIM	0.95	0.48	82.5	97.0	99.4	99.4
EmailUse15	SIM	0.87	0.55	62.4	90.9	96.9	99.7
EmailUse16	SIM	0.96	0.46	87.2	98.0	100.0	99.7
EmailUse21	SIM	0.82	0.41	60.4	84.5	89.8	95.0
EmailUse22	SIM	0.83	0.43	63.9	84.8	90.7	93.3
EmailUse23	SIM	0.88	0.51	65.9	91.2	95.2	98.5
EmailUse24	SIM	1.00	0.00	100.0	100.0	100.0	100.0
Word Processing and Flowcharts							
WordProcessing11	SIM	0.25	0.39	2.9	14.9	24.9	56.1
WordProcessing12	SIM	0.81	0.46	56.6	82.4	89.0	94.7
WordProcessing13	SIM	0.38	0.46	7.6	23.3	45.2	73.7
WordProcessing14	SIM	0.89	0.56	63.3	94.6	97.5	99.4
WordProcessing15	SIM	0.82	0.61	48.4	84.8	96.1	98.8
WordProcessing16	SIM	0.79	0.44	54.5	80.7	88.7	90.1
WordProcessing17	SIM	0.85	0.55	58.0	91.2	94.6	97.4
WordProcessing18	SIM	0.87	0.57	61.5	92.9	97.2	98.3
WordProcessing19	SIM	0.70	0.54	34.7	70.6	83.6	92.7

Table 11. *Item Difficulty, Discrimination, and Percent Correct by Quartile for Entire Group (N=1,335).*

Item	Type	Item Difficulty *p*	Item Discrimination D	Percent Correct by Quartile Q1	Q2	Q3	Q4
WordProcessing110	SIM	0.69	0.59	27.1	65.2	86.7	94.4
WordProcessing21	SIM	0.44	0.22	28.3	43.2	47.5	56.1
WordProcessing22	SIM	0.53	0.53	16.3	37.8	66.1	90.9
WordProcessing23	SIM	0.38	0.44	8.5	23.0	46.9	71.4
FlowCharting1	SIM	0.82	0.48	55.4	82.1	93.5	97.4
FlowCharting2	SIM	0.79	0.43	56.6	79.1	89.3	92.1
FlowCharting3	SIM	0.62	0.45	30.6	54.4	72.6	89.8
FlowCharting4	SIM	0.81	0.55	50.4	81.8	93.5	98.3
FlowCharting5	SIM	0.64	0.49	32.4	52.4	81.4	87.7

For the entire group of examinees, difficulty (*p*) values ranged from 0.05 to 1.00. Six items had difficulty levels below .30 (see Table 12). Five of these items were simulation items that may need to be reviewed to determine if they need revision (SpreadSheet22, InternetSearching6, Graphics2, Video3, and WordProcessing11). One item was a multiple choice item, EthicsSafeUse16. Most of these tasks (e.g., use a spreadsheet to create a basic formula) are not intended to be difficult. One simulation item, EmailUse24, had no variability because all participants performed the item correctly. Seventy-one items had difficulty values between 0.30 and 0.89; twenty-one items had *p*-values between 0.90 and 0.99.

Table 12. *Items and Indicators with the Difficulty Levels Below 0.30.*

Item and Indicator	Type	Item Difficulty *p*	Item Discrimination D	Percent Correct by Quartile Q1	Q2	Q3	Q4
EthicsSafeUse16 (Recognizes the permanency of electronic data.)	MC	0.18	0.03	16.9	15.5	17.8	22.2
Graphics2 (Edit images within software using cropping.)	SIM	0.13	0.24	3.2	6.4	15.8	26.6
Video3 (Perform basic digital video editing by adding narration.)	SIM	0.24	0.38	3.5	10.1	24.0	55.6
SpreadSheet22 (Use spreadsheets to compute basic formulas.)	SIM	0.05	0.17	0.9	2.4	4.0	12.9
InternetSearching6 (Use the advanced search features of search engines. (e.g. Boolean, date limits, language, etc))	SIM	0.13	0.27	2.3	4.7	13.0	29.8

WordProcessing11 (Set page margins within a word processing document.)	SIM	0.25	0.39	2.9	14.9	24.9	56.1

Item discrimination values ranged from 0.00 to 0.61. Sixty-four items are functioning very well with *D*-values equal to and greater than 0.40 (Crocker & Algina, 1986). Ten items (see Table 13) had *D*-values below 0.20, and are recommended for review and/or revision (Crocker & Algina, 1986) (EmailUse24, EthicsSafeUse16, SoftwareUse13, EthicsSafeUse1, SoftwareUse7, EthicsSafeUse7, EthicsSafeUse3, SpreadSheet22, SoftwareUse9, and SoftwareUse6). EmailUse24 and SpreadSheet22 were simulation items. Six items were dichotomous scores, multiple selection items (EthicsSafeUse1, SoftwareUse7, EthicsSafeUse7, EthicsSafeUse3, SoftwareUse9, SoftwareUse6), and two were multiple choice (EthicsSafeUse16 and SoftwareUse13) that should be carefully reviewed.

Table 13. *Items and Indicators with the Discrimination Levels Below 0.20*.

		Item Difficulty	Item Discrimination	Percent Correct by Quartile			
Item and Indicator	**Type**	***p***	***D***	**Q1**	**Q2**	**Q3**	**Q4**
SoftwareUse6 (Access information through online resources including encyclopedias, libraries, education and government websites, and electronic catalogs (a.k.a. card catalogs).)	Y/N	0.86	0.19	78.4	80.7	89.0	95.3
SoftwareUse7 (Access information through online resources including encyclopedias, libraries, education and government websites, and electronic catalogs (a.k.a. card catalogs).)	Y/N	0.44	0.14	37.9	30.4	43.8	62.9
SoftwareUse9 (Access information through online resources including encyclopedias, libraries, education and government websites, and electronic catalogs (a.k.a. card catalogs).)	Y/N	0.31	0.19	21.6	22.0	30.5	49.7
SoftwareUse13 (Select the best device to complete a given task such as digital cameras, scanners, and external storage devices.)	MC	0.43	0.03	45.2	37.8	41.0	47.1
EthicsSafeUse1 (Understand and follow copyright laws pertaining to	A/B	0.89	0.08	84.8	91.2	90.4	91.5

Table 13. *Items and Indicators with the Discrimination Levels Below 0.20.*

		Item Difficulty	Item Discrimination	Percent Correct by Quartile			
Item and Indicator	Type	p	D	Q1	Q2	Q3	Q4
software and/or Internet resources, including duplicating and/or plagiarizing text and media files.)							
EthicsSafeUse3 (Display an awareness of potentially inappropriate media use in regards to technology.)	Y/N (A/B)	0.64	0.16	50.7	60.8	68.6	73.4
EthicsSafeUse7 (Identify security risks that are involved with giving out personal information (e.g. fake eBay sign in to steal password).)	A/B	0.86	0.14	79.9	87.8	88.1	88.0
EthicsSafeUse16 (Recognizes the permanency of electronic data.)	MC	0.18	0.03	16.9	15.5	17.8	22.2
SpreadSheet22 (Use spreadsheets to compute basic formulas.)	SIM	0.05	0.17	0.9	2.4	4.0	12.9
EmailUse24 (Use email to send a message)	SIM	1.00	0.00	100.0	100.0	100.0	100.0

Overall, these item statistics indicate that the ST^2L can be used to assess the technology literacy of students, based on the indicators developed, at varying ability levels of computer literacy. To examine how the ST^2L 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 pre-survey (see Table 14).

Table 14. *Item Difficulty, Discrimination, and Percent Correct by Quartile for Beginner Group (N=343).*

		Item Difficulty	Item Discrimination	Percent Correct by Quartile			
Item	Type	p	D	Q1	Q2	Q3	Q4
Software Use and File Manipulation							
SoftwareUse1	MC	0.63	0.34	42.5	59.3	61.8	89.5
SoftwareUse2	MC	0.73	0.35	50.6	71.6	77.5	90.7
SoftwareUse3	MC	0.84	0.31	71.3	82.7	86.5	96.5
SoftwareUse4	MC	0.53	0.28	33.3	44.4	62.9	70.9
SoftwareUse5	MC	0.85	0.38	66.7	86.4	88.8	97.7
SoftwareUse6	Y/N	0.81	0.26	70.1	70.4	86.5	95.4
SoftwareUse7	Y/N	0.36	0.02	40.2	28.4	30.3	44.2
SoftwareUse8	Y/N	0.56	0.34	34.5	42.0	66.3	79.1
SoftwareUse9	Y/N	0.24	0.07	19.5	23.5	21.4	33.7
SoftwareUse10	MC	0.91	0.45	74.7	95.1	97.8	97.7

Table 14. *Item Difficulty, Discrimination, and Percent Correct by Quartile for Beginner Group (N=343).*

Item	Type	Item Difficulty p	Item Discrimination D	Percent Correct by Quartile			
				Q1	Q2	Q3	Q4
SoftwareUse11	MC	0.85	0.31	67.8	84.0	91.0	96.5
SoftwareUse12	MC	0.92	0.31	79.3	97.5	95.5	96.5
SoftwareUse13	MC	0.42	-0.02	44.8	48.2	36.0	40.7
SoftwareUse14	MC	0.74	0.24	60.9	70.4	71.9	94.2
SoftwareUse15	A/B	0.88	0.47	70.1	88.9	100.0	94.2
FileManipulation1	SIM	0.66	0.47	35.6	58.0	83.2	88.4
FileManipulation2	SIM	0.67	0.49	32.2	65.4	79.8	91.9
FileManipulation3	SIM	0.83	0.53	56.3	84.0	92.1	97.7
FileManipulation4	SIM	0.60	0.48	25.3	54.3	74.2	87.2
Ethics, Safety and Acceptable Use							
EthicsSafeUse1	A/B	0.90	0.10	85.1	92.6	93.3	90.7
EthicsSafeUse2	MC	0.76	0.32	58.6	70.4	82.0	93.0
EthicsSafeUse3	Y/N (A/B)	0.66	0.15	55.2	64.2	69.7	74.4
EthicsSafeUse4	MC	0.85	0.63	51.7	92.6	94.4	100.0
EthicsSafeUse5	MC	0.85	0.57	56.3	87.7	96.6	97.7
EthicsSafeUse6	MC	0.87	0.55	62.1	96.3	93.3	98.8
EthicsSafeUse7	A/B	0.85	0.22	77.0	81.5	84.3	97.7
EthicsSafeUse8	A/B	0.87	0.26	73.6	85.2	92.1	96.5
EthicsSafeUse9	A/B	0.93	0.34	80.5	95.1	97.8	100.0
EthicsSafeUse10	A/B	0.87	0.46	66.7	91.4	92.1	98.8
EthicsSafeUse11	A/B	0.75	0.43	44.8	82.7	83.2	90.7
EthicsSafeUse12	A/B	0.94	0.42	77.0	100.0	100.0	100.0
EthicsSafeUse13	A/B	0.79	0.46	49.4	86.4	84.3	95.4
EthicsSafeUse14	A/B	0.93	0.28	82.8	95.1	95.5	98.8
EthicsSafeUse15	A/B	0.91	0.54	66.7	98.8	97.8	100.0
EthicsSafeUse16	MC	0.17	-0.02	18.4	12.4	18.0	17.4
EthicsSafeUse17	A/B	0.92	0.31	75.9	97.5	96.6	100.0
EthicsSafeUse18	A/B	0.84	0.26	71.3	82.7	88.8	94.2
EthicsSafeUse19	A/B	0.89	0.61	60.9	97.5	98.9	100.0
EthicsSafeUse20	A/B	0.67	0.32	42.5	64.2	82.0	79.1
EthicsSafeUse21	A/B	0.57	0.23	43.7	46.9	71.9	66.3
EthicsSafeUse22	MC	0.84	0.52	56.3	88.9	94.4	97.7
EthicsSafeUse23	MC	0.81	0.47	52.9	87.7	91.0	94.2
EthicsSafeUse24	MC	0.77	0.49	44.8	77.8	94.4	90.7
EthicsSafeUse25	MC	0.88	0.57	62.1	95.1	97.8	98.8
Graphics, Presentation, and Video Editing							
Graphics1	SIM	0.69	0.45	33.3	69.1	80.9	90.7
Graphics2	SIM	0.08	0.20	1.2	3.7	9.0	16.3
Graphics3	SIM	0.37	0.33	13.8	30.9	49.4	53.5
Graphics4	SIM	0.71	0.49	35.6	80.3	78.7	89.5
Video1	SIM	0.82	0.33	59.8	85.2	91.0	91.9
Video2	SIM	0.24	0.24	12.6	18.5	22.5	43.0
Video3	SIM	0.15	0.31	1.2	7.4	16.9	36.1
Video4	SIM	0.36	0.44	5.8	25.9	42.7	70.9

Table 14. *Item Difficulty, Discrimination, and Percent Correct by Quartile for Beginner Group (N=343).*

Item	Type	Item Difficulty p	Item Discrimination D	Percent Correct by Quartile			
				Q1	Q2	Q3	Q4
Video5	SIM	0.22	0.33	4.6	13.6	21.4	48.8
Presentation1	SIM	0.90	0.56	65.5	96.3	97.8	100.0
Presentation2	SIM	0.74	0.46	43.7	77.8	84.3	90.7
Presentation3	SIM	0.58	0.50	19.5	53.1	71.9	88.4
Presentation4	SIM	0.94	0.54	78.2	97.5	100.0	100.0
Spreadsheets							
SpreadSheet11	SIM	0.88	0.55	63.2	96.3	96.6	95.4
SpreadSheet12	SIM	0.90	0.57	70.1	97.5	95.5	98.8
SpreadSheet13	SIM	0.43	0.42	16.1	27.2	55.1	73.3
SpreadSheet14	SIM	0.74	0.63	29.9	79.0	92.1	95.4
SpreadSheet15	SIM	0.67	0.48	33.3	67.9	86.5	81.4
SpreadSheet21	SIM	0.80	0.61	41.4	82.7	96.6	97.7
SpreadSheet22	SIM	0.01	0.09	0.0	0.0	2.3	3.5
SpreadSheet23	SIM	0.60	0.51	24.1	51.9	75.3	89.5
SpreadSheet24	SIM	0.67	0.63	18.4	70.4	86.5	94.2
SpreadSheet25	SIM	0.87	0.56	62.1	90.1	98.9	98.8
SpreadSheet26	SIM	0.83	0.57	49.4	84.0	97.8	100.0
SpreadSheet27	SIM	0.76	0.62	36.8	84.0	86.5	96.5
SpreadSheet28	SIM	0.77	0.53	41.4	84.0	92.1	91.9
SpreadSheet29	SIM	0.19	0.35	3.5	4.9	22.5	44.2
Browser Use and Email							
InternetSearching1	SIM	0.78	0.35	60.9	85.2	83.2	83.7
InternetSearching2	SIM	0.91	0.55	70.1	95.1	98.9	100.0
InternetSearching3	SIM	0.87	0.42	66.7	88.9	96.6	94.2
InternetSearching4	SIM	0.92	0.65	71.3	98.8	100.0	100.0
InternetSearching5	SIM	0.87	0.57	62.1	92.6	98.9	96.5
InternetSearching6	SIM	0.06	0.16	1.2	4.9	6.7	11.6
InternetSearching7	SIM	0.73	0.54	36.8	79.0	87.6	90.7
InternetSearching8	SIM	0.85	0.61	54.0	90.1	98.9	98.8
EmailUse11	SIM	0.65	0.39	39.1	56.8	77.5	84.9
EmailUse12	SIM	0.89	0.56	66.7	91.4	98.9	98.8
EmailUse13	SIM	0.86	0.59	55.2	93.8	96.6	98.8
EmailUse14	SIM	0.92	0.56	71.3	97.5	97.8	100.0
EmailUse15	SIM	0.80	0.59	47.1	82.7	93.3	98.8
EmailUse16	SIM	0.94	0.57	75.9	100.0	98.9	100.0
EmailUse21	SIM	0.70	0.44	46.0	59.3	79.8	93.0
EmailUse22	SIM	0.78	0.52	48.3	82.7	91.0	91.9
EmailUse23	SIM	0.82	0.55	50.6	88.9	92.1	96.5
EmailUse24	SIM	1.00	0.00	100.0	100.0	100.0	100.0
Word Processing and Flowcharts							
WordProcessing11	SIM	0.15	0.32	0.0	2.5	22.5	36.1
WordProcessing12	SIM	0.76	0.51	46.0	72.8	92.1	94.2
WordProcessing13	SIM	0.26	0.36	5.8	7.4	36.0	53.5
WordProcessing14	SIM	0.83	0.60	47.1	86.4	97.8	98.8

Table 14. *Item Difficulty, Discrimination, and Percent Correct by Quartile for Beginner Group (N=343).*

Item	Type	Item Difficulty p	Item Discrimination D	Percent Correct by Quartile			
				Q1	Q2	Q3	Q4
WordProcessing15	SIM	0.74	0.65	26.4	80.3	91.0	98.8
WordProcessing16	SIM	0.73	0.56	35.6	75.3	85.4	96.5
WordProcessing17	SIM	0.79	0.61	40.2	82.7	94.4	98.8
WordProcessing18	SIM	0.84	0.62	51.7	87.7	96.6	98.8
WordProcessing19	SIM	0.60	0.52	18.4	64.2	73.0	84.9
WordProcessing110	SIM	0.58	0.56	12.6	53.1	74.2	90.7
WordProcessing21	SIM	0.40	0.23	21.8	44.4	42.7	52.3
WordProcessing22	SIM	0.41	0.47	10.3	27.2	42.7	81.4
WordProcessing23	SIM	0.27	0.38	5.8	13.6	28.1	60.5
FlowCharting1	SIM	0.76	0.54	44.8	74.1	85.4	97.7
FlowCharting2	SIM	0.73	0.46	35.6	80.3	85.4	89.5
FlowCharting3	SIM	0.55	0.50	17.2	49.4	64.0	87.2
FlowCharting4	SIM	0.69	0.57	28.7	65.4	85.4	96.5
FlowCharting5	SIM	0.48	0.46	14.9	42.0	57.3	79.1

For the beginner group, the item difficulties ranged from 0.03 to 1.00. Twelve items (see Table 15) had *p*-values less than 0.30 (SpreadSheet22, InternetSearching6, Graphics2, Video3, WordProcessing11, EthicsSafeUse16, SpreadSheet29, Video5, Video2, SoftwareUse9, WordProcessing13, and WordProcessing23), while 46 items had *p*-values between 0.80 and 0.99. Every beginner participant selected the correct response on one item (EmailUse24). When the items were examined for the lowest quartile of participants rated as beginners, only five participants were able to respond correctly to the SpreadSheet22 item.

Table 15. *Items and Indicators with Difficulty Less than 0.30 for Beginner Group (N=343).*

Item	Type	Item Difficulty p	Item Discrimination D	Percent Correct by Quartile			
				Q1	Q2	Q3	Q4
SoftwareUse9 (Access information through online resources including encyclopedias, libraries, education and government websites, and electronic catalogs (a.k.a. card catalogs).)	Y/N	0.24	0.07	19.5	23.5	21.4	33.7

Table 15. *Items and Indicators with Difficulty Less than 0.30 for Beginner Group (N=343).*

Item	Type	Item Difficulty p	Item Discrimination D	Percent Correct by Quartile Q1	Q2	Q3	Q4
EthicsSafeUse16 (Recognizes the permanency of electronic data.)	MC	0.17	-0.02	18.4	12.4	18.0	17.4
Graphics2 (Edit images within software using cropping.)	SIM	0.08	0.20	1.2	3.7	9.0	16.3
Video2 (Perform basic digital video editing by removing a section of video.)	SIM	0.24	0.24	12.6	18.5	22.5	43.0
Video3 (Perform basic digital video editing by adding narration.)	SIM	0.15	0.31	1.2	7.4	16.9	36.1
Video5 (Use "Save As…" to change the name of the working file.)	SIM	0.22	0.33	4.6	13.6	21.4	48.8
SpreadSheet22 (Use spreadsheets to compute basic formulas.)	SIM	0.01	0.09	0.0	0.0	2.3	3.5
SpreadSheet29 (Change page orientation between landscape and portrait.)	SIM	0.19	0.35	3.5	4.9	22.5	44.2
InternetSearching6 (Use the advanced search features of search engines. (e.g. Boolean, date limits, language, etc))	SIM	0.06	0.16	1.2	4.9	6.7	11.6
WordProcessing11 (Set page margins within a word processing document.)	SIM	0.15	0.32	0.0	2.5	22.5	36.1
WordProcessing13 (Adjust line spacing within a word processing document.)	SIM	0.26	0.36	5.8	7.4	36.0	53.5
WordProcessing23 (Print a specific page range.)	SIM	0.27	0.38	5.8	13.6	28.1	60.5

Item discrimination values for the beginner group ranged between -0.02 and 0.65. Nine items (see Table 16) had *D*-values under 0.20, indicating the need for review and/or revision (Crocker & Algina, 1986) (EthicsSafeUse16, SoftwareUse13, EmailUse24, SoftwareUse7, SoftwareUse9, SpreadSheet22, EthicsSafeUse1, EthicsSafeUse3, and InternetSearching6). Sixty-

six items were functioning extremely well for the beginners group with *D*-values greater than 0.40 (Crocker & Algina, 1986).

Table 16. *Items and Indicators with Discrimination Values under 0.20 for the Beginner Group (N=343).*

Item and Indicator	Type	Item Difficulty *p*	Item Discrimination *D*	Percent Correct by Quartile			
				Q1	Q2	Q3	Q4
SoftwareUse7 (Access information through online resources including encyclopedias, libraries, education and government websites, and electronic catalogs (a.k.a. card catalogs).)	Y/N	0.36	0.02	40.2	28.4	30.3	44.2
SoftwareUse9 (Access information through online resources including encyclopedias, libraries, education and government websites, and electronic catalogs (a.k.a. card catalogs).)	Y/N	0.24	0.07	19.5	23.5	21.4	33.7
SoftwareUse13 (Select the best device to complete a given task such as digital cameras, scanners, and external storage devices.)	MC	0.42	-0.02	44.8	48.2	36.0	40.7
EthicsSafeUse1 (Understand and follow copyright laws pertaining to software and/or Internet resources, including duplicating and/or plagiarizing text and media files.)	A/B	0.90	0.10	85.1	92.6	93.3	90.7
EthicsSafeUse3 (Display an awareness of potentially inappropriate media use in regards to technology.)	Y/N (A/B)	0.66	0.15	55.2	64.2	69.7	74.4
EthicsSafeUse16 (Recognizes the permanency of electronic data.)	MC	0.17	-0.02	18.4	12.4	18.0	17.4
SpreadSheet22 (Use spreadsheets to compute basic formulas.)	SIM	0.01	0.09	0.0	0.0	2.3	3.5
InternetSearching6 (Use the advanced search	SIM	0.06	0.16	1.2	4.9	6.7	11.6

Table 16. *Items and Indicators with Discrimination Values under 0.20 for the Beginner Group (N=343).*

Item and Indicator	Type	Item Difficulty *p*	Item Discrimination *D*	Percent Correct by Quartile			
				Q1	Q2	Q3	Q4
features of search engines. (e.g. Boolean, date limits, language, etc))							
EmailUse24 (Use email to send a message)	SIM	1.00	0.00	100.0	100.0	100.0	100.0

These item statistics indicate that with minimal revisions the ST^{2}L will provide a well-functioning tool that can be used for assessing beginning students' technology literacy skills as well as students in other ability levels (intermediate, advanced). The average item difficulty and discrimination for each section were examined. Average item difficulty and item discrimination by subsection are shown in Table 17. 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 students.

Table 17. *Average Item Difficulty and Discrimination by Subsection for Entire Group and Beginners' Group.*

Subsection	Average Item Difficulty *p*		Average Item Discrimination *D*	
	Entire Group	Beginners' Group	Entire Group	Beginners' Group
Software Use and File Manipulation	0.74	0.69	0.31	0.32
Ethics, Safety and Acceptable Use	0.84	0.80	0.36	0.38
Graphics, Presentation, and Video Editing	0.59	0.52	0.38	0.40
Spreadsheets	0.73	0.65	0.47	0.51
Browser Use and Email	0.85	0.80	0.43	0.48
Word Processing and Flowcharts	0.67	0.59	0.49	0.50

For the entire group and the beginners' group, the most difficult section was *Graphics,*
Presentation, and Video Editing, with 59% of the entire group getting these items correct and
52% of the beginners' group getting these items correct. The easiest section for both groups was
Browser Use and Email with 85% of the participants in the entire group and 80% of the
participants in the beginners' group getting the items correct. Participants in both groups did
almost as well in the *Ethics, Safety and Acceptable Use* section (84% and 80%). Next in order of
success for both groups were *Software Use and File Manipulation* (74% and 69%), *Spreadsheets*
(73% and 65%), and *Word Processing and Flowcharts* (67% and 59%). When investigating the
ST^{2}L by section, the tool functions well in discriminating between the best and worst performers
in every section.

   Time analysis. Time is an important consideration when assessing student skills. The
average time to complete each section of the ST^{2}L was calculated. Only those individuals who
completed at least one section were included in the analysis, and total time was recorded for each
student, which could consist of several sessions over the three-week testing period. Further,
students could have left in the middle of a section, and returned at a latter time, in which the time
lapse would have been recorded. The median time for completing all sections of the tool was just
over 37 minutes (see Table 18). However, the amount of time needed to complete all sections of
the ST^{2}L ranged from approximately 13.72 minutes to 183.35 minutes -- just a little over three
hours. The longest amount of total time spent (151.47 minutes) in a single section was for the
Ethics, Safety and Acceptable Use, which was the section with the highest overall subscores.
Visual inspection of observations that took the least amount of time to complete the tool revealed
that participants were from different schools in different districts.

Table 18. *Average Time in Minutes to Complete Each Subsection by the Entire Group*.

Subsection	N	Median	Mean	Std Dev	Minimum	Maximum
Software Use and File Manipulation	1512	6.37	6.81	3.08	0.53	66.43
Ethics, Safety and Acceptable Use	1440	8.01	8.80	5.85	0.85	151.47
Graphics, Presentation, and Video Editing	1413	4.00	4.37	2.41	0.62	44.13
Spreadsheets	1411	4.58	5.06	2.98	0.27	46.82
Browser Use and Email	1379	6.15	6.58	2.95	0.62	46.12
Word Processing and Flowcharts	1371	6.95	7.50	3.06	0.87	48.03
Total Time for All Sections	1268	37.16	38.90	12.12	13.72	183.35

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 ST^{2}L tool increases by less than two minutes. The median time for beginners to complete all six sections of the tool was 29.24 minutes while the average time was less than 41.54 minutes. These are reasonable time requirements for 8th 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.

Table 19. *Average Time in Minutes to Complete Each Subsection by the Beginner Group*.

Subsection	N	Median	Mean	Std Dev	Minimum	Maximum
Software Use and File Manipulation	403	6.47	7.37	4.31	0.53	66.43
Ethics, Safety and Acceptable Use	376	8.44	9.07	4.53	1.08	43.85
Graphics, Presentation, and Video Editing	364	4.28	4.70	3.29	0.62	44.13
Spreadsheets	362	4.98	5.40	3.03	0.27	39.47
Browser Use and Email	363	6.80	7.00	2.50	0.68	22.83
Word Processing and Flowcharts	357	7.35	8.13	4.08	1.00	48.03
Total Time for All Sections	310	39.24	41.54	11.55	18.92	88.88

Incorrect answer and distracter analysis. Additional analyses were conducted to examine patterns in the students' wrong answer 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 (see Table 20). Four items (see Table 21) had over 60% of the student participants' select one incorrect distracter while less than 10% of the participants selected a different incorrect distracter (SoftwareUse3, SoftwareUse11, SoftwareUse13, and EthicsSafeUse16). In addition, some participants were able to advance through the tool skipping multiple choice or A/B Yes/No items.

Table 20. *Percentage of Students Selecting Each Multiple Choice Wrong Answer Distracter*.

MC Item	*n*	%	
SoftwareUse1			
a	190	46.80	
b	60	14.78	
c	156	38.42	
SoftwareUse2			
a	146	48.67	
b	68	22.67	
d	86	28.67	
SoftwareUse3			
a	9	5.70	**
b	116	73.42	*
c	33	20.89	
SoftwareUse4			
a	211	38.57	
c	257	46.98	
d	79	14.44	
SoftwareUse5			
a	59	45.74	
b	51	39.53	
d	19	14.73	
SoftwareUse10			
a	16	23.19	
c	40	57.97	
d	13	18.84	
SoftwareUse11			
a	136	78.16	*
c	15	8.62	**

51

Table 20. *Percentage of Students Selecting Each Multiple Choice Wrong Answer Distracter.*

MC Item	*n*	%	
d	23	13.22	
SoftwareUse12			
a	25	31.65	
b	14	17.72	
c	40	50.63	
SoftwareUse13			
a	809	94.29	*
b	23	2.68	**
d	26	3.03	**
SoftwareUse14			
b	146	45.63	
c	151	47.19	
d	23	7.19	**
EthicsSafeUse2			
a	188	72.31	*
c	72	27.69	
EthicsSafeUse4			
a	44	31.88	
b	60	43.48	
d	34	24.64	
EthicsSafeUse5			
a	40	30.77	
b	29	22.31	
c	61	46.92	
EthicsSafeUse6			
a	35	31.25	
b	42	37.50	
c	35	31.25	
EthicsSafeUse16			
a	764	64.91	*
b	51	4.33	**
d	362	30.76	
EthicsSafeUse22			
a	42	25.77	
b	51	31.29	
c	70	42.94	
EthicsSafeUse23			
a	75	42.61	
c	28	15.91	
d	73	41.48	
EthicsSafeUse24			
a	116	48.13	
b	54	22.41	
c	71	29.46	
EthicsSafeUse25			

Table 20. *Percentage of Students Selecting Each Multiple Choice Wrong Answer Distracter*.

MC Item	*n*	%
a	38	30.65
b	32	25.81
c	54	43.55

* > 60% selected distracter
** < 10% selected distracter

Table 21. *Multiple Choice Items and Indicators with Distracters Selected Most and Least Often*.

Item/Indicator/Item stem/Item Distracters by Selection Frequency		
SoftwareUse3 (Select appropriate uses for web browser software.) *Your teacher has asked you to research hurricane storm systems. Which of the following is the best software for this assignment?*		
a. E-mail	9	5.70 **
b. Presentation	116	73.42 *
c. Spreadsheet	33	20.89
d. Web browser (correct)	1188	88.99
SoftwareUse11 (Select the best device to complete a given task such as digital cameras, scanners, and external storage devices.) *Which of the following devices is the best choice to film a clip of the school play?*		
a. Projector	136	78.16 *
b. Video camera (correct)	1,329	87.84
c. MP3 Player	15	8.62 **
d. Digital Microscope	23	13.22
SoftwareUse13 (Select the best device to complete a given task such as digital cameras, scanners, and external storage devices.) *Which of the following devices is the best choice to make a paper copy of your science assignment?*		
a. Scanner	809	94.29 *
b. Projector	23	2.68 **
c. Printer (correct)	645	42.63
d. Hard Drive	26	3.03 **
EthicsSafeUse16 (Recognizes the permanency of electronic data.) *After you have deleted an email off a computer, how long could it exist?*		
a. It is gone immediately	764	64.91 *
b. The remainder of the school year	51	4.33 **
c. It could be permanent (correct)	261	18.1
d. Until the computer is reformatted	362	30.76

* > 60% selected distracter
** < 10% selected distracter

These items might reflect misunderstandings in intent, and consequently, require

revisions or at minimum careful review and consideration in light of these results. For example,

SoftwareUse13 is worded as *Which of the following devices is the best choice to make a paper*

copy of your science assignment?, and more than half of the student participants selected scanner as the correct answer. Students may have interpreted this question as already having a paper copy, and needing to make a duplicate. In such a case, a scanner would be the best answer as opposed to a printer, which was the correct answer for this item.

None of the multiple choice items required student participants to respond. Student participants could have clicked the *Continue* button to skip the question. In these cases, student participants were warned with a pop-up message in a dialog box that stated, *You have not yet responded to all the items on the screen. Any items you leave blank will be scored as incorrect.* Students had to respond by clicking *Cancel* or *Continue* in the message dialog in order to continue. The number of participants for each item who did not have a user response recorded is depicted in Table 22. The percent of missing responses to items ranged from 0.0% to a high of 1.85% for the SoftwareUse6 item.

Table 22. *Participants' Responses that Were Not Recorded for All Multiple Choice and A/B Yes/No Items*.

Item	n	%
SoftwareUse1	6	0.40
SoftwareUse2	24	1.59
SoftwareUse3	20	1.32
SoftwareUse4	20	1.32
SoftwareUse5	18	1.19
SoftwareUse6	28	1.85
SoftwareUse7	26	1.72
SoftwareUse8	27	1.78
SoftwareUse9	26	1.72
SoftwareUse10	8	0.53
SoftwareUse11	10	0.66
SoftwareUse12	5	0.33
SoftwareUse13	10	0.66
SoftwareUse14	6	0.40
SoftwareUse15	0	0.00
EthicsSafeUse1	6	0.42
EthicsSafeUse2	5	0.35
EthicsSafeUse3	7	0.49

Table 22. *Participants' Responses that Were Not Recorded for All Multiple Choice and A/B Yes/No Items.*

Item	n	%
EthicsSafeUse4	4	0.28
EthicsSafeUse5	7	0.49
EthicsSafeUse6	4	0.28
EthicsSafeUse7	10	0.69
EthicsSafeUse8	12	0.83
EthicsSafeUse9	9	0.62
EthicsSafeUse10	12	0.83
EthicsSafeUse11	8	0.55
EthicsSafeUse12	10	0.69
EthicsSafeUse13	7	0.49
EthicsSafeUse14	9	0.62
EthicsSafeUse15	8	0.55
EthicsSafeUse16	4	0.28
EthicsSafeUse17	6	0.42
EthicsSafeUse18	4	0.28
EthicsSafeUse19	7	0.49
EthicsSafeUse20	4	0.28
EthicsSafeUse21	6	0.42
EthicsSafeUse22	5	0.35
EthicsSafeUse23	9	0.62
EthicsSafeUse24	9	0.62
EthicsSafeUse25	12	0.83

Students also had the option to skip steps in the performance-based simulated tasks. Thus, the percentage of students who chose to skip steps for each simulation item was also examined (see Table 23). Ten items (see Table 24) had greater than 10% of the participants select the skip step option (FileManipulation4, Video2, SpreadSheet22, SpreadSheet24, SpreadSheet29, WordProcessing11, WordProcessing13, WordProcessing19, WordProcessing110, FlowCharting3). These items should be reviewed. Some of the tasks (e.g., Use the table creation feature of a word processor) may have been skipped because the students did not know the correct response, while other items may reflect a situation in which the simulated environment did not provide students' known path for execution of the required task (e.g., removing a section of video).

Table 23. *Number and Percent of Participants who Omitted the Step, Had No User Response Recorded, or Were Missing the Entire Section of Items.*

Item	Omit Step n	Omit Step %	No User Response n	No User Response %	Missing Section n
FileManipulation1	34	2.25	1	0.07	0
FileManipulation2	43	2.84	0	0.00	0
FileManipulation3	54	3.57	0	0.00	0
FileManipulation4	155	10.24	0	0.00	0
Graphics1	29	2.02	7	0.49	76
Graphics2	75	5.22	116	8.07	76
Graphics3	32	2.23	5	0.35	76
Graphics4	37	2.57	0	0.00	76
Video1	33	2.30	0	0.00	76
Video2	215	14.96	0	0.00	76
Video3	95	6.61	0	0.00	76
Video4	86	5.98	0	0.00	76
Video5	68	4.73	0	0.00	76
Presentation1	44	3.06	1	0.07	76
Presentation2	45	3.13	1	0.07	76
Presentation3	73	5.08	8	0.56	76
Presentation4	42	2.92	0	0.00	76
SpreadSheet11	50	3.48	0	0.00	75
SpreadSheet12	49	3.41	0	0.00	75
SpreadSheet13	0	0.00	2	0.14	75
SpreadSheet14	86	5.98	1	0.07	75
SpreadSheet15	66	4.59	0	0.00	75
SpreadSheet21	95	6.61	0	0.00	75
SpreadSheet22	378	26.29	2	0.14	75
SpreadSheet23	107	7.44	1	0.07	75
SpreadSheet24	172	11.96	1	0.07	75
SpreadSheet25	65	4.52	0	0.00	75
SpreadSheet26	71	4.94	0	0.00	75
SpreadSheet27	143	9.94	0	0.00	75
SpreadSheet28	83	5.77	0	0.00	75
SpreadSheet29	339	23.57	0	0.00	75
InternetSearching1	41	2.93	1	0.07	113
InternetSearching2	47	3.36	0	0.00	113
InternetSearching3	34	2.43	134	9.57	113
InternetSearching4	47	3.36	0	0.00	113
InternetSearching5	44	3.14	71	5.07	113
InternetSearching6	116	8.29	4	0.29	113
InternetSearching7	69	4.93	1	0.07	113
InternetSearching8	71	5.07	0	0.00	113
EmailUse11	28	2.00	0	0.00	113
EmailUse12	35	2.50	0	0.00	113
EmailUse13	45	3.21	0	0.00	113
EmailUse14	46	3.29	0	0.00	113

Table 23. *Number and Percent of Participants who Omitted the Step, Had No User Response Recorded, or Were Missing the Entire Section of Items.*

	Omit Step		No User Response		Missing Section
Item	n	%	n	%	n
EmailUse15	58	4.14	0	0.00	113
EmailUse16	38	2.71	0	0.00	113
EmailUse21	43	3.07	0	0.00	113
EmailUse22	66	4.71	1	0.07	113
EmailUse23	96	6.86	0	0.00	113
EmailUse24	0	0.00	0	0.00	113
WordProcessing11	339	23.94	197	13.91	97
WordProcessing12	85	6.00	8	0.56	0
WordProcessing13	361	25.49	119	8.40	97
WordProcessing14	88	6.21	5	0.35	97
WordProcessing15	126	8.90	55	3.88	97
WordProcessing16	69	4.87	12	0.85	97
WordProcessing17	95	6.71	62	4.38	97
WordProcessing18	80	5.65	2	0.14	97
WordProcessing19	267	18.86	4	0.28	97
WordProcessing110	245	17.30	0	0.00	97
WordProcessing21	122	8.62	0	0.00	97
WordProcessing22	131	9.25	0	0.00	97
WordProcessing23	121	8.55	0	0.00	97
FlowCharting1	95	6.71	0	0.00	97
FlowCharting2	54	3.81	1	0.07	97
FlowCharting3	156	11.02	0	0.00	97
FlowCharting4	68	4.80	0	0.00	97
FlowCharting5	79	5.58	0	0.00	97

Table 24. *Simulation Items and Indicators with the Most Participants Who Choose to Select the Omit Item Option.*

SIM Item and Indicator	n	%	
FileManipulation4 (Search for specific files.)	155	10.24	*
Video2 (Perform basic digital video editing by removing a section of video.)	215	14.96	*
SpreadSheet22 (Use spreadsheets to compute basic formulas.)	378	26.29	*
SpreadSheet24 (Use spreadsheets to create a graph.)	172	11.95	*
SpreadSheet29 (Change page orientation between landscape and portrait.)	339	23.57	*
WordProcessing11 (Set page margins within a word processing document.)	349	28.01	*
WordProcessing13 (Adjust line spacing within a word processing document.)	369	27.98	*
WordProcessing19 (Insert a hyperlink into a document.)	270	18.87	*
WordProcessing110 (Use the table creation feature of a word processor.)	249	17.38	*

Table 24. *Simulation Items and Indicators with the Most Participants Who Choose to Select the Omit Item Option.*

SIM Item and Indicator	n	%	
FlowCharting3 (Create flowcharts as a learning strategy.)	157	11.08	*

* > 10% of participants chose to omit step

In addition, during the data cleaning phase, it was discovered that some students did not have their response recorded for simulation items. This would occur when a student responded in a way that the tool had not anticipated. If the participant had a section score recorded, it was assumed that the student had completed the section and missing answers were incorrect answers and should have been recorded as wrong; thus, the response was imputed as a wrong response. Analyses were conducted to examine the number of participants who had imputed wrong answers supplied for each item but had no user response recorded. Results are depicted in Table 23, along with omitted items and sections that were not attempted. The programming staff should carefully investigate these items to ascertain user responses in future efforts.

As can be gleaned from Table 23, the following simulation items appear to have greater numbers of participants not have their responses recorded than the rest of the items: Graphics2, InternetSearching3, InternetSearching5, WordProcessing11, WordProcessing13, and WordProcessing17. This would suggest that the programming code for capturing the user response from these items may need to be reviewed, or user interface testing may need to be conducted to determine how students are responding to these items.

Reliability. Another measure of test quality is reliability or internal consistency of scores across the student participants. In this case, reliability was estimated using a measure of internal consistency known as Kruder-Richardson 20 (K-R 20), which is used for dichotomously scored items (e.g., right or wrong; yes or no). For an assessment in which subsection scores are reported, it is necessary to estimate reliability of the scores for both the subsections and the entire

test. As shown in Table 25, the reliability estimate for scores on the entire tool is K-R 20 = .95. This is a very high level of internal consistency reliability. At the subsection level, K-R 20 reliability estimates range from = .67 for *Software Use and File Manipulation* to.86 for *Word Processing and Flowcharts*. All of the individual subsections have more than acceptable internal consistency reliability (K-R 20 greater than .70), with the exception of the *Software Use and File Manipulation* subsection. Though the ST^{2}L is really a criterion-referenced test or mastery test, there is a modest degree of variability in scores across items and across student participants. This modest level of variability leads to higher measures of association, and thus, higher estimates of reliability for these data.

Table 25. *Internal consistency reliability of ST^{2}L sections*.

Section on ST^{2}L	n	Non-Standardized K-R 20
Software Use and File Manipulation	1513	0.67
Ethics, Safety and Acceptable Use	1442	0.81
Graphics, Presentation, and Video Editing	1437	0.72
Spreadsheets	1438	0.82
Browser Use and Email	1400	0.85
Word Processing and Flowcharts	1416	0.86
Total for ST^{2}L	1335	0.95

Inter-section correlations are presented in Table 26. The degree to which the subsection scores correlate to the total score is an indication of the cohesiveness of the construct. Please note that only participants who completed all sections received a total score, and thus, only those participants were used in correlations involving a total score (*n*=1,335). The ST^{2}L inter-section correlations show positive relationships between subsection scores and the total score. Correlations of individual subsections with the total inventory range from .73 (*Ethics, Safety and Acceptable Use*) to .87 (*Word Processing and Flowcharts*). All of the subsections correlate strongly and positively with the total score (greater than .70). The *Ethics, Safety, and Acceptable*

Use section has the weakest relationship with the total score, which can be attributed to the subsection measuring a somewhat different construct 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 non-ethical uses of information and communication technology, and vice versa.

Correlations between scores on the individual subsections with the other subsections are strong and positive (greater than $r=.50$), which affirms the cohesiveness of the construct. As expected, *Ethics, Safety and Acceptable Use* has the lowest relationship with the other subsections, for the reason previously described. The strongest subsection correlations are between the *Word Processing and Flowcharts* and *Spreadsheets* ($r=.68$) sections, and *Word Processing and Flowcharts* and *Graphics, Presentation, and Video Editing* ($r=.68$) sections. Please note that all correlations are statistically significant at $\alpha = .05$.

Table 26. *Correlations (all significant at $\alpha=.05$) between scores on ST^2L subsections*.

	1	2	3	4	5	6	7
Software Use and File Manipulation (1)	1						
Ethics, Safety and Acceptable Use (2)	0.52	1					
Graphics, Presentation, and Video Editing (3)	0.57	0.48	1				
Spreadsheets (4)	0.55	0.57	0.61	1			
Browser Use and Email (5)	0.55	0.53	0.50	0.60	1		
Word Processing and Flowcharts (6)	0.58	0.54	0.68	0.68	0.60	1	
Total on ST^2L (7)	0.77	0.73	0.81	0.84	0.78	0.87	1

Validity. To estimate construct validity of scores from the ST^2L, the relationship between scores on the ST^2L and the rating from the pre-survey (designed to measure frequency of computer use, attitudes towards computers, and comfort with various computer tasks) was examined. These relationships were analyzed in two ways: (1) an examination of the differences between computer experience levels and (2) the correlations among the ST^2L and the various sub scores of the assessment.

The first step was to analyze the differences across the computer experience levels. Participants' computer experience was measured with a composite variable that was created by combining the responses to the items that asked how often students engaged in different computer tasks and their self-rating about how well they perform different computer tasks. These computer scores were rank ordered and then participants were separated into three levels or groups – beginner, intermediate, and advanced. Table 27 shows the computer experience levels, as measured by the composite scores on the *Computer Task Comfort* and *Frequency of Use* section 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 ($F(2, 1332)= 52.65, p< .01$), which demonstrates the validity of the ST^2L in discriminating groups of students based on their experience levels. A Tukey follow-up procedure that maintained $\alpha = .05$ confirmed that advanced users performed significantly better than both intermediate and basic users; and intermediate users performed significantly better than beginners.

Table 27. *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

1=Software Use and File Manipulation

Table 27. *Subsection and total scores by computer experience.*

Computer Experience	1	2	3	4	5	6	Total*	n
2=Ethics, Safety and Acceptable Use								
3=Graphics, Presentation, and Video Editing								
4=Spreadsheets								
5=Browser Use and Email								
6=Word Processing and Flowcharts								
*N=1335 for Total								

All correlations of ST^{2}L subsection scores with factor scores on the various sections of the survey were positive and significant at $\alpha < .05$ (see Table 28). The strongest relationship was between the *Word Processing and Flowcharts* subsection of the ST^{2}L and the *Comfort of Computer Tasks* composite measure of the survey (.35). In general, the *Comfort of Computer Tasks* composite measure has the strongest correlations across all of the subsection scores as well as the total score. The *Comfort of Computer Tasks* contained items that asked student participants to rate their comfort level with performing various technology-related tasks. These tasks were very similar to those measured through the simulations in the subsections of the ST^{2}L. The weakest relationship was between the *Ethics, Safety, and Acceptable Use* subsection and the *Frequency of Computer Use* composite (.10). As the *Ethics, Safety, and Acceptable Use* subsection measures a different construct of technology literacy than the other subsections, the weaker relationship with the survey sections is expected and should not be interpreted as a weakness of that subsection or the assessment as a whole. The composite score for *Use and Comfort* had positive relationships with all of the subsections (from .20 to .30) and with the average total score of the ST^{2}L (.33).

Table 28. *Correlations (all significant at $\alpha = .05$) among ST^{2}L Subsections and pre-survey factors.*

Subsection	Frequency of Computer Use	Attitudes towards Computers	Comfort of Computer Tasks	Composite Use and Comfort

Table 28. *Correlations (all significant at α = .05) among ST²L Subsections and pre-survey factors.*

Subsection	Frequency of Computer Use	Attitudes towards Computers	Comfort of Computer Tasks	Composite Use and Comfort
Software Use and File Manipulation	0.18	0.23	0.33	0.29
Ethics, Safety and Acceptable Use	0.10	0.18	0.25	0.20
Graphics, Presentation, and Video Editing	0.17	0.18	0.34	0.29
Spreadsheets	0.18	0.17	0.32	0.28
Browser Use and Email	0.14	0.17	0.27	0.23
Word Processing and Flowcharts	0.20	0.19	0.34	0.30
Average Total Score	0.21	0.23	0.39	0.33

Descriptive Statistics of Post-Survey

The post-survey provided student participants with the opportunity to express their opinions about the ST²L. Out of the 1,513 students in the final sample, *n*=1,303 students completed the post-survey. The question, responses, and response frequencies for the post-survey items are shown in Table 29. As can be gathered, only 12.82% of the participants indicated that none of the activities in the ST²L were fun. Approximately 81% of the sample indicated that some or most of the activities in the ST²L were fun.

Another important aspect was the level of difficulty that student participants experienced. Only 6.22% of the student participants indicated that the ST²L was difficult. Approximately 43% indicated that the activities were easy, and half of the participants perceived that the activities were just right. Equally important, more than 60% of the student participants indicated that none of the activities were confusing to complete.

63

Student participants were also asked to rate how well they thought that the tool measured their knowledge and skills. To facilitate this process, the subscores were visible to the student participants when responding to this question as shown in Appendix E. Less than 7% of the student participants indicated the tool was a poor measure. More than half (52.11%) of the student participants indicated the tool was a good way for them to demonstrate most of their skills and knowledge.

A final consideration was to evaluate the degree to which student participants perceived that they had gained their skills and knowledge from experiences at school. Approximately 63% of the student participants indicated they had done some or most of the activities at school. The remaining student participants indicated that very few or none of the activities found within the ST^{2}L were engaged in while at school.

Table 29. *Questions, responses and response frequencies to post-survey*.

Question/Response	Students	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
Were any of the activities confusing for you (e.g., did you get stuck on any of the activities)?		
Yes	802	61.55
No	501	38.45

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 Proctor Survey

Prior to the analysis of the ST^2L data, the proctor survey was used to determine whether any of the data were compromised by external circumstances (e.g., fire alarm during administration). Twenty-nine proctors responded to at least part of the post-ST^2L administration survey. After carefully reviewing the proctor responses, it was determined that no extenuating circumstances required the removal of student participants from the dataset.

Table 30 shows the locations in which the ST^2L was administered. The proctors reported that the majority of the student participants completed the ST^2L in a computer lab (58.62%). The proctors also reported that on average 29 computers were used concurrently during the administration, ranging from nine to 40.

Table 30. *Locations of ST^2L administration*.

Location	Proctors	Percent
Media center	5	17.24
Classroom	5	17.24
Computer lab	17	58.62
Other	2	6.90
Total	29	100.00

In terms of the administration of the ST^2L, 14 (48.28%) of the proctors said the tool was administered by one proctor, 11 (48.28%) indicated that 2 proctors participated, and the remaining proctor indicated four facilitators. Fifteen (51.72%) of the proctors indicated that the administration was completed in one session. Of those who indicated more than one session only one (7.14%) indicated 1.5 sessions, 9 (64.29%) indicated it was completed in two sessions, and

the remaining proctor (7.14%) indicated seven sessions. All of the 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 and clarify with explanations. Table 31 summarizes this information. Thirteen (44.83%) of the proctors reported student participants experienced problems logging on to the ST^{2}L. An analysis of the explanations showed that, in every case, the problem was a result of a spelling error in the passwords (the password tangerine was misspelled) that were provided to the proctors. Only one proctor reported an Internet connection problem during the administration, which is an encouraging result of the administration.

Table 31. *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

When investigating the technical problems encountered, 15 (51.72%) of the proctors reported at least one technical problem. A review of explanations revealed a variety of technical problems, including the installation of the Flash 9.0 Player, laptop batteries running out of power, and problems advancing from one section to the next. In each case, proctors reported resolving the problems (e.g., installation of Flash plug-in, recharging batteries, refreshing page) so that student participants were able to continue the assessment.

Six (20.69%) proctors reported that some students had problems understanding the instructions. A review of the explanations showed general misunderstandings (e.g., not being able to use hot keys), and two proctors reported that the items in the Word Processing and

Flowcharts section were difficult for students to understand. Ten proctors (34.48%) reported that there were interruptions during the administration period (e.g., fire drill, assembly, end of class session, end of day). Although three proctors (10.34%) reported that some students engaged in off- task activities (e.g., browsing the Internet or socializing with peers) during the administration period, 27 proctors (93.1%) rated the overall level of student effort and engagement observed during the administration of the ST^2L as high or very high. The success of the administration of the ST^2L was confirmed when 100% of the proctors either agreed or strongly agreed with the statement, *The administration of the ST^2L ran smoothly*.

Summary of Evaluation

An assessment tool or instrument is developed for a specific purpose. In this case, the purpose of the assessment was to create a tool for monitoring technology literacy of 8th grade students within the state of Florida for NCLB reporting and school district curricular purposes. In advance, the purpose of the tool was determined to be for low stakes assessments. Appropriate low stakes applications include identifying weaknesses of individual students in the appropriate uses of technology, identifying the technology curricular needs of students by a school district, and reporting for monitoring purposes in which no high-stakes consequences are attached. 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*). Sound development procedures were followed for item writing and review, and usability analyses were conducted to ensure that the user interface and the simulated performance-based tasks were as clear and intuitive as possible. Finally, the pilot test was conducted according to a sound research design and sampling plan.

The purpose of this study was to demonstrate the overall validity and reliability of the ST^{2}L for the stated purposes. Assessment quality was evaluated by considering item analyses, reliability, and validity. The item *p*-values (item difficulty) for the entire group showed the difficulty of the items ranged from 0.05 to 1.0. Sixty-six percent of the items had difficulty values between 0.30 and 0.89; 20% of the items had *p*-values between 0.90 and 0.99; and the remaining had *p*-values below 0.30, indicating these items were the most difficult for the student participants. In contrast, item discrimination values ranged from 0.00 to 0.61. Sixty percent items are functioning very well with *D*-values equal to or above 0.40; 31% are functioning moderately

well with *D*-values between 0.20 and 0.40. Nine percent of the items had *D*-values below 0.20; the majority of examinees performed very similarly on these items, and thus, the items do not discriminate among student participants. While the ST^2L 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 beginner group, the item difficulties ranged from 0.03 to 1.00, while item discrimination values for the beginner group ranged between -0.02 and 0.65. In comparison to the entire group, twelve as opposed to six items had difficulty levels below 0.30, which means that the student participants in the beginners group found the items more difficult to answer correctly. Additionally, the average time to complete the ST^2L increased when comparing the entire group with the beginners group. Further, the entire group performed better than the beginners group across the subsections and the total score, as evidenced by the average item difficultly scores. When compared to the entire group, the items also were more discriminating across the all subsections with the beginners group.

Reliability was estimated for the inventory for each subsection and for the total score, using internal consistency reliability, or K-R 20 reliability. All reliability estimates were found to be positive and above the social science acceptable threshold (K-R 20 > 0.7), with the exception of *Software Use and File Manipulation* which had a K-R 20 = 0.67 -- which is still moderately strong (Nunnaly, 1978). All of the subsection scores are significantly and positively correlated (*r* > 0.45). Further, the total score on the ST^2L significantly, highly, and positively correlates with each subsection score (*r* > 0.7). In particular, the *Word Processing and Flowcharts* subsection

correlates most strongly with the other subsections. These high inter-section score correlations and total score correlations demonstrate a highly cohesive and internally consistent structure of the ST^2L for the sample participants.

To estimate the construct validity of scores, the ST^2L was related to the pre-survey scores. Put simply, one assumed valid measure of technology literacy was compared to the results of the ST^2L as a way to demonstrate the measures are meaningfully related. Specifically, two approaches were used to demonstrate the relationships. First, the subsection scores and total scores were calculated across computer experience levels (the beginner, intermediate, and advanced levels were determined based on the results of the pre-survey). The results showed that the advanced group performed significantly better than the intermediate group on the total ST^2L assessment, and the intermediate group performed significantly better than the beginners group on the total score. Further, subsection scores for the ST^2L in each of the computer experience classifications are differentiated in the same way, in favor of the advanced over intermediate and the intermediate over the beginner classification.

Second, the correlations between the sub sections of the ST^2L and the pre-survey measures of *Frequency of Computer Use*, *Attitudes towards Computers*, and *Computer Task Comfort* were calculated. As hypothesized, all the correlations were significantly and positively correlated. Most of the measures were modest correlations. In particular, the *Computer Task Comfort* measure had the highest relationships with each of the subsection scores and the total score. The measures in the *Computer Task Comfort* section were specifically designed to match with the indicators found in the ST^2L; thus, the stronger relationships were anticipated. These analyses show the construct validity of the ST^2L in relation to the stated indicators and the broad construct of technology literacy.

Recommendations

Results from the pilot test support several recommendations about items that should be reviewed and considered for possible revision. Several quality dimensions (item difficulty values below 0.30; item discrimination below 0.20; more than 60% or less than 10% of the participants selecting an incorrect distracter; more than 10% of the student participants choosing to omit a step; and whether or not a user response was recorded) were used to identify items needing review. A summary of these quality dimensions in relation to potentially problematic items is depicted in Table 32. A total of 24 unique items were tagged as potentially problematic according to at least one of these quality dimensions.

Overall, six of the items (EthicsSafeUse16, SpreadSheet22, SoftwareUse13, WordProcessing11, SoftwareUse3, and SoftwareUse11) were tagged for not meeting three or more quality dimensions. Eleven of the items were tagged with falling short on at least two quality dimensions, and the remaining seven items fell short on one. These items need to be carefully reviewed by the development team to determine what revisions, if any, should be made -- especially the six items that were tagged with three or more quality dimension problems.

Though an item may be tagged by one of the quality dimensions, this does not necessarily require revision. For instance, Video3 required student participants to edit a video file by adding narration. It may be that the student participants did not know how to accomplish this task, and thus, the results may inform curricular modifications. Further, those items that did not record a user response may only need review by the programming staff to ascertain how the students were able to respond without securing their actions. However, in some cases, modification of the item may be the best alternative. For instance, EthicsSafeUse16 asks, *After you have deleted an email off a computer, how long could it exist?* with the answer options *it is gone immediately*, *the*

remainder of the school year, *it could be permanent*, and *until the computer is reformatted*. The item was tagged by five quality dimensions. The appropriate answer here is that *it could be permanent* as removal of information from an operating system index does not necessarily remove it from the physical hard drive. However, 764 (approximately 50%) student participants selected *it is gone immediately* and 362 (approximately 24%) selected *until the computer is reformatted*. To an average computer user, upon removal of an email from the email client software (meaning not accessible in deleted bin), the email is no longer accessible to them at an application layer. Further, reformatting a hard drive could make the raw data inaccessible to most users. It is relatively clear that the student participants did not understand that the intent of the question was to ask if the content could be available to someone such as a law enforcement specialist, and/or there is validity in the distracters.

Table 32. *Summary of potentially problematic items by quality dimensions*.

Item	Type	*p*-value < 0.30	*D*-Values < 0.20	Distracter selected by > 60%	Distracter selected <10%	Omit step chosen by > 10%	Response Not Recorded	# Tagged
EthicsSafeUse16	MC	x	x	x	x		x	5
SpreadSheet22	SIM	x	x			x	x	4
SoftwareUse13	MC		x	x	x		x	4
WordProcessing11	SIM	x				x	x	3
SoftwareUse3	MC			x	x		x	3
SoftwareUse11	MC			x	x		x	3
Graphics2	SIM	x					x	2
InternetSearching6	SIM	x					x	2
SoftwareUse6	Y/N		x				x	2
SoftwareUse7	Y/N		x				x	2
SoftwareUse9	Y/N		x				x	2
EthicsSafeUse1	A/B		x				x	2
EthicsSafeUse3	Y/N (A/B)		x				x	2
EthicsSafeUse7	A/B		x				x	2
SpreadSheet24	SIM					x	x	2
WordProcessing13	SIM					x	x	2
WordProcessing19	SIM					x	x	2
Video3	SIM	x						1
EmailUse24	SIM		x					1
FileManipulation4	SIM					x		1

Table 32. *Summary of potentially problematic items by quality dimensions.*

Item	Type	*p*-value < 0.30	D-Values < 0.20	Distracter selected by > 60%	Distracter selected <10%	Omit step chosen by > 10%	Response Not Recorded	# Tagged
Video2	SIM					x		1
SpreadSheet29	SIM					x		1
WordProcessing110	SIM					x		1
FlowCharting3	SIM					x		1

The final recommendation pertains to the use of scores from this ST^2L. Use of a single aggregated score may be misrepresentative of actual examinee skills and knowledge. Though the scores are strongly correlated, as shown in the results, student participants performed better and worse in different sections of the ST^2L. Only reporting a composite score of technology literacy would obscure the different patterns of knowledge and skills. The most informative approach would be to report the results of this ST^2L as a profile comprised of subsection scores, rather than a single summary score. Any mastery decisions that are made using scores from this instrument should be arrived at after using an accepted standard setting method and that examines subsection scores separately.

Limitations

This study was limited to eighth grade students (N=1,561) from 13 school districts in Florida during the spring of 2008. The findings may not generalize to different populations or different time periods. It is also possible that the results are biased, based upon differences in the demographics of the schools that agreed to participate in the pilot study and those that did not.

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 (APA, 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 (APA, 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 ST^2L was specifically developed as a low stakes assessment and designed to provide data related to the technology literacy of eighth grade students for district reporting requirements. It is also an appropriate tool for curriculum design purposes for teachers. However, the tool, in its present form, is not suitable for use in high-stakes applications such as evaluating student performance, computing school grades, or student promotion/retention. Prior to using the ST^2L 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.

Information and communication technology is perpetually changing. As a result, developing valid and reliable instruments and assessment tools to measure the construct is an ongoing and difficult process. For instance, at the debut of the ST^2L project, the most current version of NETS-S (ISTE Standards) was published in 2000. These standards, along with the Florida technology standards, formed the basis for the indicators. Since then, new NETS-S standards have been released. Therefore, the ST^2L does not map as well with the latest standards as it did with the previous version. Further, the most commonly used productivity software (Microsoft Office 2003) has been revised and the newest released version of Microsoft's software package (Microsoft Office 2007) has a fundamentally changed human-computer

74

interface paradigm. While the ST^2L is intended to be software-independent, students may not find the interface similar enough to the software suites that they are accustomed to using in their schools and homes. Thus, the ST^2L may not adequately measure the knowledge and skills of these students.

The validation process included comparing the ability of the ST^2L to separate students by their ICT ability levels. In order to make this comparison, another measure was needed to separate the students into ability levels. For this study, the results from a self-assessment were used to classify students. This self-assessment by students may not adequately represent the students' true ability.

A final limitation to the current pilot test is that the evaluation did not include variables for socio-economic status (SES), primary language other than English spoken at home (ESOL), or special education status of the student participants. The NCLB requires the demonstration of student technology literacy across the student's race, ethnicity, gender, family income, geographic location, and disability. The current study only included the race, ethnicity, gender, and geographic location of the student participants. While efforts were made to collect SES information and ESOL from the students, a programming error resulted in the loss of this relevant information. Future research efforts should target special education, ESOL, and SES status as relevant dimensions.

Conclusions

The *Student Tool for Technology Literacy* was developed using established best practices for test development, according to low stakes testing purposes. Additional steps would be needed to use the ST^2L for high-stakes purposes. For low stakes applications, the ST^2L is clearly a satisfactory tool for educators to employ. An extensive, thorough process was followed for

defining the indicators, which were developed within the framework of the ISTE-NETS-S and state standards. The items are well mapped to these indicators, and provide measurement 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 ST^2L should be ready for a low stakes implementation, and should become a useful tool of school districts for reporting NCLB purposes and for helping students and districts to target technology-related curricular needs.

References

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

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.

Florida Department of Education, Office of Educational Technology. (2003). *Florida educator accomplished practice #12: Technology*. Retrieved June 20, 2008, from http://etc.usf.edu/L4L/E-EAP12.html

Georgia Department of Education (2008). *One stop shop for educators: Georgia technology literacy assessment toolkit*. Retrieved June 20, 2008 from http://www.georgiastandards.org/toolkit.aspx

Harmes, J. C., Barron, A. E., & Kemker, K. (2007, March). *Development and implementation of an online tool to assess teacher technology skills*. Paper presented at the annual international conference of the Society for Information Technology & Teacher Education, San Antonio, TX.

ISTE, International Society for Technology in Education. (2002). *National Educational Technology Standards for Teachers: Preparing teachers to use technology*. Eugene, OR: Author.

ISTE, International Society of Technology in Education (2007). *NETS for Students*. Retrieved June 20, 2008 from http://www.iste.org/Content/NavigationMenu/NETS/ForStudents/2007Standards/NETS_for_Students_2007.htm

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

Public Schools of North Carolina (2008). *Test of computer skills (graduation requirement)*. Retrieved June 18, 2008 from http://www.ncpublicschools.org/accountability/testing/computerskills/

Stansbury, M. (2008). States struggle with assessing technology literacy. *eSchool News Online*. Retrieved June 20, 2008 from http://www.eschoolnews.com/news/top-news/?i=53017

U.S. Department of Education. (2001). *No Child Left Behind: Enhancing Education Through Technology Act of 2001*. Retrieved June 20, 2008 from http://www.ed.gov/policy/elsec/leg/esea02/pg34.html

Appendix

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.
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. Identifies and locates 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.

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 email to send a message.
2. Use email to receive/open a message.
3. Use email to forward a message.
4. Use email to reply to a message.
5. Use email 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)
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. Recognizes the permanency of electronic data.
14. Maintain password security.
15. Understand the need for virus scans, pop up blockers, spy ware blockers, firewalls and filters.

Appendix B. Participation Letter.

February 18, 2008

Contact Name
School District Name
School Contact Email

Dear Contact Name:

The eleven districts who received EETT Leveraging Laptops grants in 2006-07 are stratified across the state and vary in size, rural and urban areas, and access to technology. These districts, including Escambia County, provide a diverse sample population for the Student Tool for Technology Literacy (ST^{2}L) field test this upcoming April. Our goal is to test the tool with approximately 500 students in the state and ask for your participation to help meet our goal.

We are looking for eighth grades teachers and students, who may or may not be in a school which participated in your Leveraging Laptops grant. Here are some items to consider and share with principals when making your decision to participate:

- The tool takes approximately 1.5 hours to complete, and may need to span across two class periods to administer, although it is preferable that it is completed in one continuous time block.
- Proctors will be asked to complete a short survey after administering the ST^{2}L to provide feedback on testing conditions, question difficulty, and technology performance.

If you are willing to volunteer schools in your district and coordinate with the validation team, please fill out the attached information sheet. We appreciate and commend your commitment to student achievement in all areas, including technology literacy.

Sincerely,
Kate Kemker
Chief, Bureau of Instruction and Innovation

ST^{2}L Field Test Information Sheet (*Example*)

School Name	% Free and Reduced Lunch at school	Approximate number of 8th graders to participate	High Technology Access (laptops or other exposure, please explain)?	Anticipate 1 or 2 class periods to complete	Are headphones available for students requiring accommodations?	Lab or classroom setting?
XXXX Middle School (example)	45%	22	Laptops from Leveraging Laptops in 1 class period, history	1 – block schedule 1.5 hour period	Yes	Lab

ST2L Field Test Proctor Guide

The Student Tool for Technology Literacy (ST^{2}L) is intended to evaluate 8th grade students' technology literacy. You will be administering the ST^{2}L as part of the field test before the statewide release of the tool. Please follow the outlined procedures below to ensure the evaluation process is as consistent and accurate as possible. We encourage you to take the tool yourself, using the "Teacher" login, prior to administering the tool to the students. Being familiar with the ST^{2}L will help to ensure you are able to answer questions and provide assistance to students. Head to http://beta.fcim.org/st2l/, pull down your district, school, and name. The User Name is "teacher" and the Password is "teacher." Clicking on "Start Regular Section" will take you to the ST^{2}L screen the students will see upon logging in.

> **Following the administration of the ST^{2}L, we ask for your participation in a short survey to provide additional information about proctoring the tool. You can find the survey by logging back into http://beta.fcim.org/st2l/, with User Name "teacher" and Password "teacher" and clicking on "Go to Proctor Survey."

Testing Environment

1. A testing location free of noise and distractions is ideal.

2. Each student will need a computer to use the tool.

3. IEP requirements such as headphones and other accommodations should be provided for students that require them.

4. Hardware/Software Requirements

a. All computers should have Flash Player version 9. (To determine current version of Flash on your computers, visit this website: http://www.adobe.com/go/tn_15507). We recommend you use Internet Explorer, Mozilla Firefox, or Apple Safari to administer the tool.

5. Use the following link to access the ST^{2}L: http://beta.fcim.org/st2l/

6. The students will login to the system using their teacher assigned ID and password. Use the spreadsheet emailed to you to assign student User Names and Passwords. You should save a copy of the list for your records if a student needs to complete the ST^{2}L at a later date and time.

7. The following is an image showing the login screen for the ST^{2}L. The student will use the drop down menus to select (1) District Name, (2) School Name, (3) Teacher, (4) User Name, and (5) Password to login to the system (passwords are case sensitive).
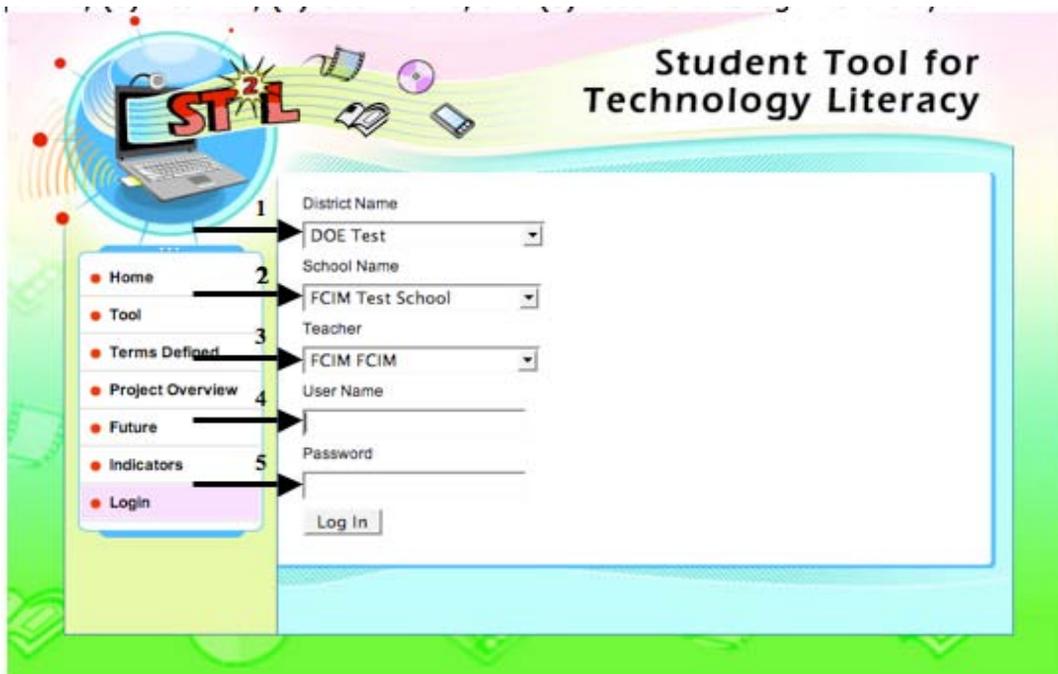
Administering the Tool

1. Explain the purpose of the ST^{2}L and this field test:

     The Florida Department of Education is creating a software tool that will help gauge 8th grade students' technology skills.

     To make sure the tool works properly, it is being tested with a select group of 8th graders in your district as well as other districts in the state.

     This is a very special group that gets to see and use the tool first before the rest of the state.

2. Make it clear that we are testing the tool not the students:

This tool is not at all like the FCAT. No one will know how you scored on the test; instead, we want to know how well the software performs.

This field test is similar to a beta test of a video game. Before a game is released, programmers and testers go through the game to ensure everything is working properly.

3. Students should do their own work and not worry about what their neighbor is doing, because they will be doing the activities in different orders.

4. Students should be encouraged to do their best even though their scores have no academic repercussions, because we need to know that this tool will provide accurate information about student technology skills.

5. Encourage students to take their time; there is no need to rush.

If a student needs help while using the tool, do not lead them to the answer in any way.

a. You can provide clarification.

b. You can help with other issues not related to completing the tasks (for example: a technical issue or a problem with the equipment).

i. You cannot actually tell them how to complete the item.

ii. You should encourage them to guess or use the **Skip This Step** button if they are unsure about how to complete a task and do not want to take a guess.

If you have technical questions during the administration of the tool, please contact FCIM Support at (850) 645-1759 or support@fcim.org for help.

Thank you for your time and effort!

Appendix D. Pre-Survey.

Background Survey

Background questions

*1. Gender:

▢　Male

▢　Female

*2. Age:

[   ]

*3. Race:

▢　White

▢　Black

▢　Hispanic

▢　Asian

▢　Other

* 4. Are you on a free or reduced lunch program?

▢　Yes

▢　No

* 5. Do you usually speak English with your family at home?

▢　Yes

▢　No

Frequency of computer use

* 6. How long have you been using computers?

☐ Less than one year

☐ One to three years

☐ Three to five years

☐ More than five years

* 7. How often do you use a computer at these locations?
Scale:

- 5: Almost every day
- 4: A few times each week
- 3: Between once a week and once a month
- 2: Less than once a month
- 1: Never

	5	4	3	2	1
At home	☐	☐	☐	☐	☐
At school	☐	☐	☐	☐	☐
At other places	☐	☐	☐	☐	☐

* 8. How often do you use:
Scale:

- 5: Almost every day
- 4: A few times each week
- 3: Between once a week and once a month
- 2: Less than once a month
- 1: Never

	5	4	3	2	1
The Internet to look up information about people, things, or ideas?	▢	▢	▢	▢	▢
Games on a computer?	▢	▢	▢	▢	▢
Word processing (e.g., Microsoft Word, Works)?	▢	▢	▢	▢	▢
Drawing, painting or graphics programs on a computer?	▢	▢	▢	▢	▢
Presentation software (e.g., Microsoft PowerPoint, Apple Keynote)?	▢	▢	▢	▢	▢
Spreadsheets (e.g., Microsoft Excel, Works)?	▢	▢	▢	▢	▢
Video Editing (e.g., Moviemaker, iMovie)?	▢	▢	▢	▢	▢
A computer for electronic communication (e.g., e-mail, instant messaging)?	▢	▢	▢	▢	▢
The Internet to download software, music, or files?	▢	▢	▢	▢	▢
The computer to help you learn school material?	▢	▢	▢	▢	▢

General comfort towards computers

*9. Rate how well you can do each of these tasks on a computer
Scale:

- 4: I can do this very well by myself
- 3: I can do this with help from someone
- 2: I know what this means but I cannot do it
- 1: I don't know what this means

	4	3	2	1
Use software to find and get rid of computer viruses	☐	☐	☐	☐
Open a file	☐	☐	☐	☐
Create/edit a document	☐	☐	☐	☐
Scroll up and down a document on the screen	☐	☐	☐	☐
Save a computer document or file	☐	☐	☐	☐
Print a computer document or file	☐	☐	☐	☐
Delete a computer document or file	☐	☐	☐	☐
Move files from one place to another on a computer	☐	☐	☐	☐
Search for a topic on the Internet	☐	☐	☐	☐
Copy or download files or music from the Internet	☐	☐	☐	☐
Attach a file to an e-mail message	☐	☐	☐	☐
Insert a formula into a spreadsheet	☐	☐	☐	☐
Use a spreadsheet to make a chart	☐	☐	☐	☐
Create a presentation (e.g., using Microsoft PowerPoint, Apple Keynote)	☐	☐	☐	☐
Play computer games	☐	☐	☐	☐
Create a multi-media presentation (with sound, pictures, and video)	☐	☐	☐	☐
Change the size of a picture (by cropping, re-sizing, etc.)	☐	☐	☐	☐
Write and send e-mails	☐	☐	☐	☐
Edit a video	☐	☐	☐	☐

89

*10. Have you taken a class in basic computer skills?

◯ Yes, I took a class at school or somewhere else

◯ No, I have not taken any computer classes

General attitudes towards computers

*11. To what extent do you agree with the following statements?
Scale:

- 4: Strongly agree
- 3: Agree
- 2: Disagree
- 1: Strongly disagree

	4	**3**	**2**	**1**
Using a computer helps me with my schoolwork.	◯	◯	◯	◯
It is important to me to work with a computer.	◯	◯	◯	◯
I think playing or working with a computer is really fun.	◯	◯	◯	◯
I lose track of time when I am using a computer.	◯	◯	◯	◯
People ask me for help with their computer.	◯	◯	◯	◯

[Submit]

Student post-ST2L Survey

*1. Did you find using this tool fun?

◻ Most of the activities were fun

◻ Some of the activities were fun

◻ None of the activities were fun

*2. How easy or difficult was using the tool for you?

◻ Easy

◻ Just right

◻ Difficult

*3. How well did this tool measure your computer skills and knowledge?

Here is your score:	
Technology Operations and Concepts	8%
Research and Information Fluency	0%
Communication and Collaboration	0%
Critical Thinking, Problem Solving, and Decision Making	0%
Creativity and Innovation	0%
Digital Citizenship	33%

◻ This tool was a good way to show most of my skills and knowledge

◻ This tool allowed me to show some of my skills and knowledge

◻ This tool did not show my skills and knowledge

*4. Have you done activities at school that are similar to the ones you just completed within the tool?

☐ We have done most of these activities at school

☐ We have done some of these activities at school

☐ We have done very few of these activities at school

☐ We have done none of these activities at school

*5. Were any of the activities confusing for you (e.g., did you get stuck on any of the activities)?

☐ Yes

☐ No

If yes, then please describe

[text box]

[Submit]

Appendix F. Proctor Survey.

Proctor Follow Up Survey

The responses on this form are confidential and will only be used for the Student Tool for Technology Literacy (ST2L) validation by the researchers. Please respond with your best response to help the validation.

* 1. Where did the students complete the ST2L you proctored?

☐ Media center

☐ Classroom

☐ Computer lab

☐ Other

if "Other", please specify: [   ]

* 2. How many computers were used at the same time in the room?

[   ]

* 3. Which software browsers did the students use to complete the tool? Check all that apply.
☐ Internet Explorer
☐ Mozilla Firefox
☐ Opera
☐ Netscape
☐ Safari
☐ Other
☐ Don't know

* 4. Were there any problems with students logging on to the ST2L?

☐ Yes

☐ No

if "Yes", please describe:

[text box]

* 5. Were there any problems maintaining an Internet connection during the administration?

☐ Yes

☐ No

if "Yes", please describe:

[text box]

* 6. Were there any other technical issues (hardware or software) during the administration?

☐ Yes

☐ No

if "Yes", please describe:

[text box]

* 7. How many proctors (facilitators) were available to assist students in completing the ST2L?

Insert number amount: [text box]

* 8. Were there any problems with students understanding the directions?

☐ Yes

☐ No

if "Yes", please describe:

[text box]

* 9. Were there any specific questions that caused problems for students?

⬤ Yes

⬤ No

if "Yes", please describe:

[text box]

*10. Were there any interruptions during the administration period (e.g., fire drill, assembly, end of class session, end of day)?

⬤ Yes

⬤ No

if "Yes", please describe:

[text box]

*11. Were students engaged in off-task activities (such as, browsing the Internet or socializing with peers) during the administration period?

⬤ Yes

⬤ No

if "Yes", please describe:

[text box]

* 12. What did students do if they finished early?

[text box]

* 13. Were headsets available for students who needed them?

⬤ Yes

⬤ No

if "Yes", how many students used head sets? [text box]

95

* 14. Did the students complete the ST2L in more than one sitting?

☐ Yes

☐ No

if "Yes", how many sittings: [   ]

*15. In your expert opinion, rate the overall student effort and levels of engagement that you observed during the administration of the tool.

☐ Very high

☐ High

☐ Medium

☐ Low

☐ Very low

* 16. Rate the extent to which you feel the following statement is true: The administration of the ST2L ran smoothly.

☐ Strongly Agree

☐ Agree

☐ Neither agree, nor disagree

☐ Disagree

☐ Strongly Disagree

Please clarify, if necessary:

[   ]

* 17. Please describe any other information you feel is important for the validation of this tool.

[   ]

Submit

Appendix G. Example Delivery Environment.

Screen after completion of pre-survey.

Confirmation between sections.
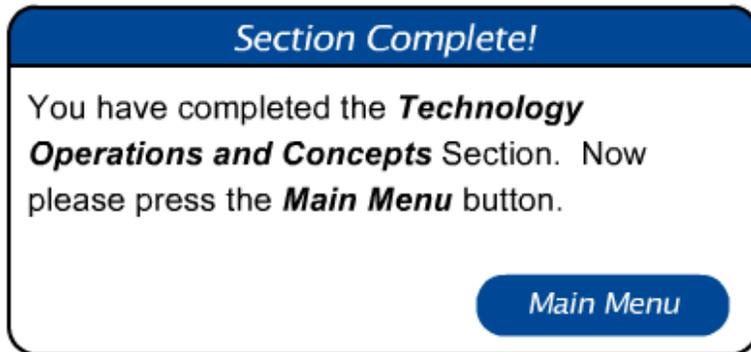
Section Complete!

You have completed the **Technology Operations and Concepts** Section. Now please press the **Main Menu** button.

Main Menu

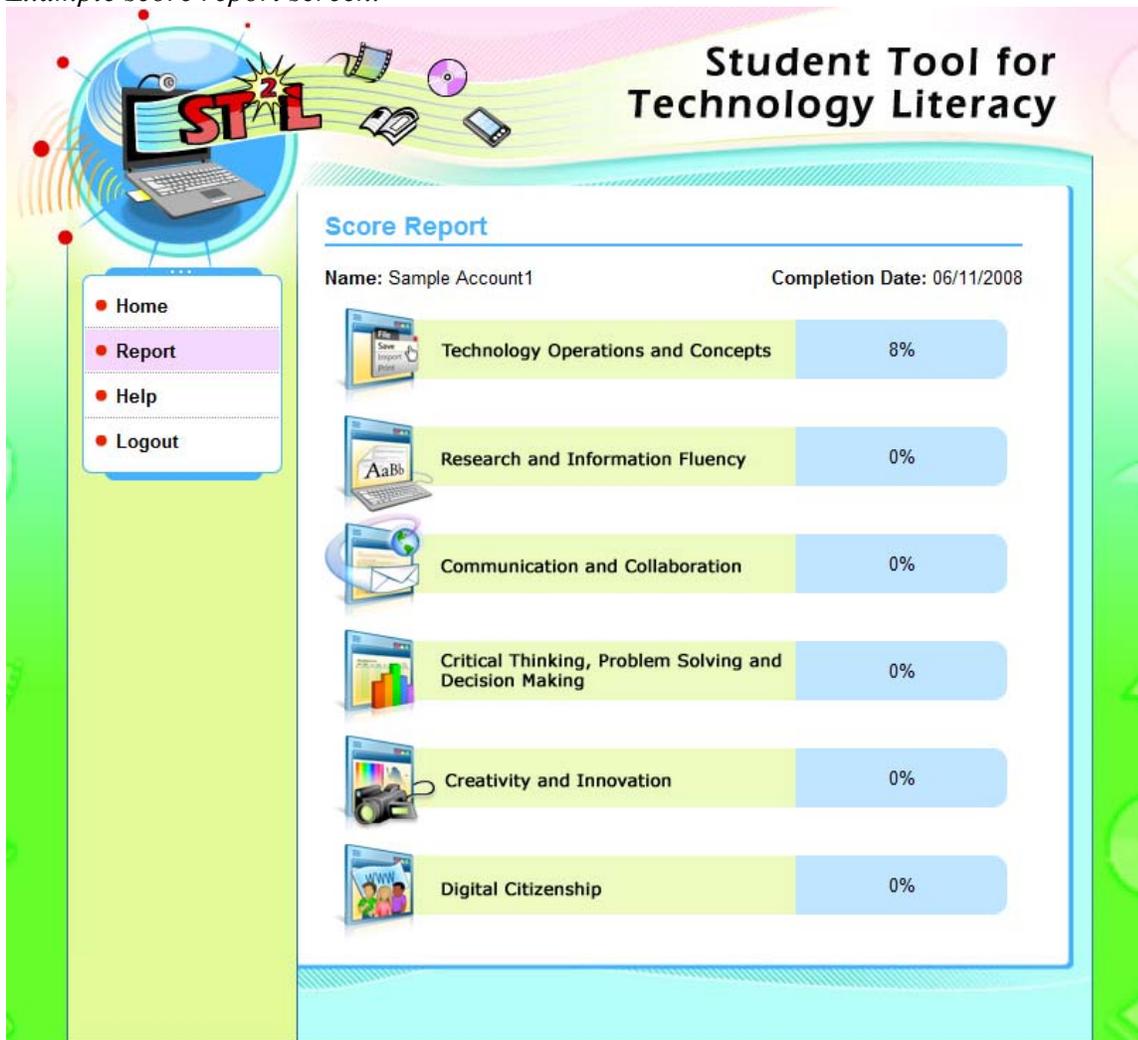*Example screen after completion of first section.*

Student Tool for Technology Literacy

- Home
- Report
- Help
- Logout

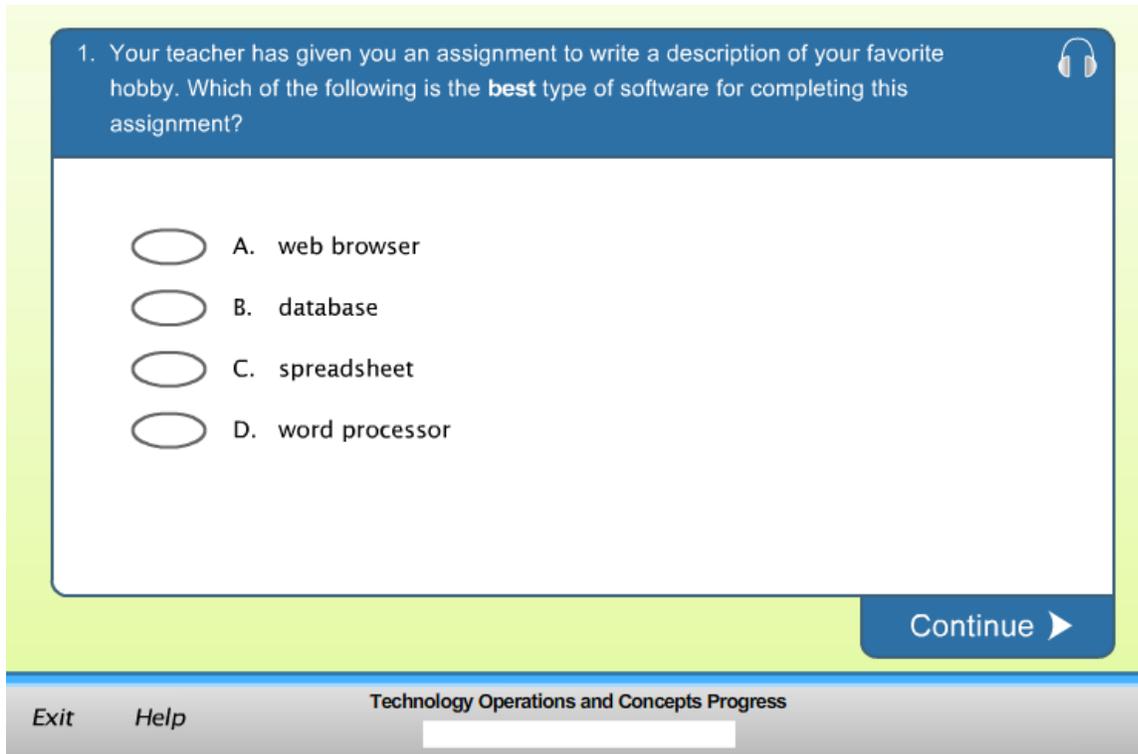**Directions:** Next up, click on Research and Information Fluency.

Technology Operations and Concepts

Research and Information Fluency

Communication and Collaboration

Critical Thinking, Problem Solving and Decision Making

Creativity and Innovation

Digital Citizenship

Example score report screen.

Student Tool for Technology Literacy

Home
Report
Help
Logout

Score Report

Name: Sample Account1   Completion Date: 06/11/2008

	Technology Operations and Concepts	8%
	Research and Information Fluency	0%
	Communication and Collaboration	0%
	Critical Thinking, Problem Solving and Decision Making	0%
	Creativity and Innovation	0%
	Digital Citizenship	0%

Appendix H. Sample Items.

Example multiple-choice item.

Example performance-based task item.

Description: For this task you will use a word processor to edit a document.

Task Progress

Current Step

Complete this step in the screen to the right.

1. Change the page margin of the report to 2 inches on the top.

Skip this step

2. Select/Highlight the text in the first paragraph.

3. Change the line spacing to be double-spaced.

4. Select/Highlight the shuttle names: Columbia, Challenger, Discovery

5. Using formatting tools, bullet

WordTech 4 - Shuttle Report

File   Edit   View   Insert   Format

Arial   8   B *I* U

The space shuttle is made up of the Orbiter, Solid Rocket Boosters, and External Tank. The astronauts ride inside the Orbiter. The Solid Rocket Boosters are the rockets that launch the shuttle into the air, and the External Tank is the fuel tank. The space shuttle program is operated by The National Aeronautics and Space Administration, NASA.

Three of the shuttles are named:
Columbia
Challenger
Discovery

The shuttle program has been operational since 1981. The first three launch dates were:
April 12, 1981
November 12, 1981
March 22, 1982

Exit   Help

Research and Information Fluency Progress