

Final Evaluation Report

Teacher Technology Literacy Inventory Development and Field Test

Enhancing Education Through Technology (EETT) Part II Competitive Opportunity

Dr. Cynthia G. Parshall
Dr. J. Christine Harmes
Gianna Rendina-Gobioff
Peggy K. Jones

December 31, 2004

Table of Contents

Executive Summary	3
Background.....	6
Indicators.....	7
Development of the Inventory	8
Field Test Design	12
Results.....	14
Participants.....	14
Descriptive Data.....	18
Test Quality.....	20
Item Analysis	20
Reliability.....	30
Validity	32
Summary of Evaluation	34
Recommendations.....	36
Limitations	38
Conclusion	40
References.....	41
Appendices.....	42
Appendix A: Performance Indicators	42
Appendix B: Sample Simulation Item	45
Appendix C: Structured Interview Questions.....	46
Appendix D: Field Test Letter	47
Appendix E: Sample Score Reports.....	50
Appendix F: Technology Uses and Attitudes Survey.....	51

Executive Summary

The focus of this project was to design a set of performance indicators for teacher technology literacy, and to design a tool to measure performance on these indicators. The purpose of the measurement tool is to help districts evaluate technology literacy for No Child Left Behind (NCLB) reporting requirements, and to provide teachers with an objective measure for targeting their professional development needs in technology.

Development

The teacher technology literacy domain of content and behavior was delineated by the following areas: basic operations; productivity; communications; research; planning, management, and instruction; and social, ethical, legal, and human issues (all of which were based on the National Educational Technology Standards (NETS)). Based on the stated low-stakes test purpose, an inventory was designed and items were written following appropriate test development processes. Performance indicators were carefully developed with attention to state and national technology standards, including integration within the STaR System. Indicators were reviewed online by technology supervisors and classroom teachers. Items were written to measure the final set of 69 indicators. Item types included multiple choice and simulated performance tasks.

During the inventory development, focus group evaluation sessions were held in Orlando, Pensacola, and Miami, and pilot tests were conducted in Leon and Pasco Counties. In addition, extensive usability analyses were conducted to ensure that the user interface was as clear and intuitive as possible. The final version of the inventory was field tested according to a sound research design and sampling plan.

Field Test Design

For the field test, the researchers sought a target sample that would be representative of the state. First, districts were categorized as small, medium, or large; and districts were selected from each of the five regions of the state. Participants came from 13 districts,

yielding a final sample size of 534. Examination of demographic variables (including computer experience, district size, and school level) found the final sample to be satisfactorily representative of the distribution of teachers across the state. Along with completion of the inventory, participants were asked to fill out a survey on their uses of and attitudes toward technology. District personnel were also asked to provide a technology experience rating (Beginner, Intermediate, or Advanced) for each participant.

Results

The inventory was found to be a sound assessment tool for the intended purpose. Test quality was evaluated by considering item analyses, reliability, and validity. On average, the item difficulty values showed the inventory to be quite easy. Like many basic skills assessments, this inventory is functioning like a mastery test, in which most examinees are expected to do quite well. In support of this interpretation, the item difficulty indices are lower (i.e., harder) for those examinees who were rated as being at a Beginner level of computer experience. Since the inventory was administered in unsecured, unproctored settings (as it will be in operational use), examinees may have received help as they took the assessment, which would also contribute to the high scores. Internal consistency reliability was estimated for the inventory for each subsection and overall. All reliability estimates were found to be positive, but relatively modest. Two validity studies, designed to address content and construct validity, were conducted. First, performance on each of the six inventory subsections was correlated with a rating of each examinee's level of computer experience. Second, the performance on each of the six inventory subsections was correlated with responses from the survey on technology attitudes and uses. The consistent pattern of different results for the six subsections indicates strongly that the inventory is measuring very distinct aspects of teacher technology literacy.

Recommendations

Relatively few modifications should be necessary before using the inventory operationally. Recommendations include deleting or modifying one item, adding items to

measure two database indicators, investigating the source of anomalous missing data issues on seven items, and using subsection scores independently for any reporting or interpretation of mastery.

Limitations

As specified, this inventory 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 (c) the implementation of additional data analyses including Confirmatory Factor Analysis and Differential Item Functioning.

Conclusions

The *Teacher Technology Literacy Inventory* was developed well, and according to low-stakes test purposes as specified in the grant. Additional steps would be needed to use this inventory for high-stakes purposes. But for low-stakes applications, the *Teacher Technology Literacy Inventory* is clearly satisfactory. An extensive, thorough process was followed for defining the indicators, which were developed within the framework of the NETS. The inventory items are well mapped to these indicators, and provide measurement of them in innovative, relevant, performance-based ways. Finally, test quality criteria show reasonable item analysis, reliability, and validity results for a relatively short criterion-referenced test (CRT) or mastery test. With few modifications, the inventory should be ready for a low-stakes implementation, and should become a useful tool for districts for reporting aggregate data for NCLB purposes and helping teachers and districts to target technology-related professional development needs.

Background

As part of the Enhancing Education Through Technology (EETT) section of the No Child Left Behind (NCLB) Act, districts will need to show evidence of teacher technology literacy achievement by 2006. To help districts meet these reporting requirements, the Office of Educational Technology (OET), Florida Department of Education (FLDOE), created a competitive grant to fund development of an objective assessment tool for measuring teacher technology literacy (FLDOE, 2003b). The purpose of this tool is to help districts evaluate technology literacy for NCLB reporting requirements, and to provide teachers with an objective measure for targeting their professional development needs in technology. Leon County School District was awarded the grant, and worked with the Florida Center for Interactive Media and measurement consultants to create a performance measurement system for assessing teacher technology literacy. This task required first establishing a set of performance indicators, then designing and creating an instrument to objectively measure accomplishment of these indicators.

Alignment with National and State Technology Standards

A specific goal of this project was the development of performance indicators for basic technology literacy that were aligned with state and national technology standards. The OET has recently revised the Educator Accomplished Practice #12: Technology (FLDOE, 2003a). Wording from this document was used in developing the performance indicators. Several national groups have developed technology-related standards for teachers, administrators, and students, such as the International Society for Technology in Education (ISTE), National Educational Technology Standards (NETS; International Society for Technology in Education, 2002), and the Partnership for 21st Century Skills. To ensure alignment with these national standards, the framework within which the indicators were developed was based on the ISTE-NETS.

Integration into the School Technology and Readiness (STaR) Chart

In order to align reporting at various levels, it was important to ensure that results from the tool developed in this project dovetail into the existing STaR Chart framework currently in use by OET. The technology standards developed in this project were created as elements to be integrated within the STaR System. Under the Accountability strand of the STaR Chart, in the Teacher Technology Standards category, an element has been added to reflect reporting of teacher technology literacy. Aggregated data from teacher use of this tool can flow directly into the reporting framework to satisfy fulfillment of this element.

Indicators

In a typical test development process, the first step is to decide on the test purpose, and then define the construct. In the Request for Proposal (RFP) for this competitive grant, the key OET initiative with which this project aligned was the “development of practical technology integration tracking tools/instruments” (FLDOE, 2003b). Technology literacy was viewed as a part of technology integration. The emphasis on a practical tool for monitoring sets the framework for a low-stakes test purpose. To further the low-stakes emphasis, the word “inventory” was used in the assessment title as a way to minimize test anxiety.

In this project, the construct definition phase was accomplished by first gathering materials from various districts that had already begun the process of investigating methods for reporting teacher technology literacy. Among these materials were indicators, performance rubrics, and teacher self-assessment tools. Using these documents as a starting point, a team made up of educators from across the state drafted a set of measurable indicators for teacher technology literacy.

The teacher technology literacy domain of content and behavior was delineated by the following areas: basic operations; productivity; communications; research; planning, management, and instruction; and social, ethical, legal, and human issues. These areas were identified based upon the NETS. The panel of content experts produced an initial set of 81

indicators, primarily based on the NETS. Extensive use was made of operational descriptions of a number of the standards from self-assessment instruments that had been developed by several districts within the state. This initial set of indicators was evaluated by the members of the Florida Council of Instructional Technology Leaders (FCITL) and the Florida Association for Computers in Education (FACE) using a web-based review process. Feedback was received from nearly 300 individuals, representing both district level personnel and classroom teachers. Overall feedback was positive, and most respondents thought that the set of indicators was a good representation of basic technology literacy. Using this feedback, the panel was reconvened, and the set of indicators was revised. Several additional indicators were added, some were deleted, and others were modified to yield a final set of 69 indicators. The complete set of indicators is provided in Appendix A, and a sample is given in Table 1 below.

Table 1

Sample of Indicators from Teacher Technology Literacy Inventory

Rename files and folders

Edit data in existing spreadsheets (cut, copy, paste cell contents)

Resize graphical objects in a document

Attach a file to an email message

Add text to a slide within a presentation

Properly cite digital resources

Development of the Inventory

A team of measurement experts next took the set of indicators and developed a test plan to ensure that each indicator was phrased in measurable terms. A decision was made to maintain the content emphasis provided in the set of indicators. This meant that each indicator was weighted equally; if a given indicator was measured by more than one item, then those multiple items were each given proportionally less weight.

After creating the test plan, the measurement team began to develop a set of item types and to write items. A primary goal for this inventory 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 was developed for the test, including several different selected-response item types as well as performance-based tasks. Altogether, the inventory consists of 105 items, of which 54 are selected-response items and 51 are steps within performance-based tasks. The selected-response item types are: text-based multiple choice items, 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 involve a variety of simulations of computerized tasks in which the examinee is guided through a series of integrated steps.

A number of decisions needed to be made about the construction of the simulated tasks. For example, various decisions were made about the layout of the screen and about the elements to be displayed (e.g., the description of the task, the current step, and a list of all required steps within the task). An example of a simulated task, and how these elements were resolved, is provided in Appendix B. In order to maintain the content weighting established by the indicators, the measurement panel continued to treat each indicator as a basic unit of assessment. In some instances, however, it was necessary to phrase a given indicator as a series of steps within a task. For example, one indicator reads, “cut, copy and paste text”. Within the context of a performance-based task, an indicator such as this might become a series of steps, such as: select a specified text phrase, cut or copy the selected phrase, move to a new location, and paste the selected text in the new location. In an instance such as this, each of the four steps would be weighted as one quarter of the indicator; thus, all the indicators continued to be of equivalent value.

Another measurement decision the test developers made was to arrange for each step within a task to be independent of every other step. The reason for this decision was to ensure that even if an examinee did not know how to do one step, he or she would not be penalized or prevented from having a full opportunity to respond correctly to a later step. To accomplish this, several decisions were implemented. First, while examinees were able to look throughout the simulated portion of the screen, once an examinee clicked on a portion of the screen or pressed *Enter*, a “Submit Step?” dialogue box appeared. The examinee was asked to confirm, or cancel, the action just taken. For any step, the examinee was also free to choose “Omit Step” if he or she did not know how to respond to the item and did not choose

to provide a response that was simply a guess. Whether the examinee submitted a correct answer, submitted an incorrect answer, or omitted a step, the program was designed to immediately score the step accordingly and then display the correct action. This enabled the examinee to undertake the next step correctly, even if the previous step was answered incorrectly or omitted.

The inventory was programmed in *Flash* and designed to be administered over the Internet in unproctored settings. Brief instructions and a few practice items were provided to give the examinee basic information on how to take the test. Examinees were required to complete the practice items before beginning the actual inventory. The inventory was designed so that examinees would be free to take the six subsections in any order and at different points in time. Furthermore, an examinee would be free to pause the inventory even while within a subsection, although item review was not permitted. No score was provided for any of the subsections until all subsections had been completed. At that point, a percent correct score on each of the six subsections would be made available to the examinee (see Appendix E for an example).

Expert Review

Concurrent with the development of the test specifications and content, usability testing was conducted. Expert reviews were conducted on very early versions of the programmed inventory. The first iteration of usability testing consisted of an initial, prototype screen for item presentation. This provided a general design of an item screen, including item instructions and navigation at both the item and general test levels. This prototype was evaluated by internal expert reviewers and, based on their comments, the first revisions were made. Next, a prototype was developed for each item type in the exam; these item types underwent expert review early in their development, when they had basic, but not full, functionality. Several additional iterations of screen design followed by expert review were conducted as the inventory and items were more fully developed.

User Tests

As the actual item text was incorporated into the developing program and user interface, several additional iterations, this time of user tests, were conducted. User tests were conducted in two ways, working with both focus groups and individual users.

The focus group user testing was conducted in three regional areas: Orlando, Pensacola, and Miami. For each iteration, between 6 and 13 teachers in the surrounding areas met as a group; teachers had varied experience with computers from entry-level to advanced skills. The research team provided an overview of the *Teacher Technology Literacy Inventory* and reviewed the directions for registration and completion of items. All groups completed the *Basic Operations* subsection of the inventory. Other subsections were randomly assigned to ensure feedback on each subsection; in most cases, the entire inventory was completed. Teachers were able to ask questions of or make comments to the research team while taking the inventory. Additionally, a feedback button was available onscreen allowing participants the option of recording their feedback on specific items. Once the group had completed the assigned inventory subsections, the group was asked a series of questions using a structured format (see Appendix C). Both content and interface issues were addressed.

The second method of user testing was carried out with four individual school personnel from two districts: Manatee and Pinellas. The participants were two teachers and two technology supervisors, with a range of technology abilities from inexperienced to very experienced. In their work environment, each of the participants was asked to take the inventory while speaking aloud their thoughts and comments. In order to reduce participants' anxiety and discomfort, the researchers emphasized that participants' performance on the inventory was not being assessed; rather the performance of the inventory was being analyzed. As the participants went through the inventory, one researcher elicited comments and another videotaped the session. The research team reviewed the resulting feedback and prioritized a set of revisions to be made before pilot testing.

Pilot Tests

After the content and interface of the inventory had been tested repeatedly in the iterative fashion described previously and appropriate modifications had been made, a pilot testing phase was implemented. The pilot testing was conducted in Pasco and Leon Counties. In each case, approximately 20 to 45 teachers attempted to complete the entire inventory. Unlike earlier user tests, the pilot test was intended to be as similar to realistic conditions as possible. For this reason, examinees were asked to work quietly, independently, and with minimal assistance from any development team personnel. Nevertheless, measurement and programming personnel observed the pilot tests and made note of remaining issues, which were then addressed before full-scale field testing was undertaken.

Field Test Design

A field test was conducted in the fall of 2004 with 613 teachers from 13 districts. Districts in the five regions of the state, as defined by the FLDOE Office of School Improvement, were represented in the sample. They were further categorized to ensure representation from small, medium, and large districts. The researchers solicited a sample designed to represent teachers from a variety of computer experience levels, school levels (e.g., elementary, middle, and high), and teaching assignments (e.g., basic, exceptional education, fine arts).

A series of communication venues was used to recruit districts to participate in the field test. An overview of the inventory development was presented to district technology representatives at FCITL meetings. A follow-up letter (see Appendix D) was available on the inventory website with instructions to submit contact information for participation in the field test. A more detailed letter was electronically sent to district contacts with instructions for recruiting a representative sample of teacher participants.

Districts who volunteered were asked to recruit 25-200 volunteer instructional staff, depending on district size. The researchers collected information about potential teacher participants from the districts who volunteered. Two key components provided by each district about the participants were: email contact information and level of computer

experience. Email addresses were used to create user accounts for completion of the inventory and to inform participants of the availability of the inventory. The second component, level of computer experience, was submitted by the district, and defined as Beginner (B), Intermediate (I), or Advanced (A).

As participants in the field test, teachers completed a survey in conjunction with the inventory. The survey consisted of items related to demographics as well as uses of and attitudes toward technology (see Appendix F). This survey was adapted from a larger instrument that has been tested and employed in several previous studies (see Barron, Kemker, Harmes, & Kalaydjian, 2003; Hogarty, Lang, & Kromrey, 2003). Following the completion of this survey, participants were able to complete the six subsections of the inventory in any order, and were able to complete the inventory over a two-week window. A subsection could be paused and completed at a later time. Participants completed the inventory in an unproctored environment. Upon completion of the inventory, participants received a percent correct score for each subsection.

For the field test, the researchers sought a target sample that would be representative of the state. First, districts were categorized as small (under 40,000 students), medium (40,000-99,999 students), or large (100,000 or more students). Six small districts volunteered, representing 22% of the teachers in the state. Four medium districts volunteered, representing 25% of the teachers in the state. Five large districts volunteered, representing 53% of the teachers in the state. The districts that volunteered were mapped by region as defined by the FLDOE Office of School Improvement, and every region was represented. Figure 1 illustrates locations of the participating districts, each of which is marked with a star (Broward, Clay, Holmes, Lee, Manatee, Miami-Dade, Orange, Palm Beach, Pasco, Pinellas, Polk, St. Johns, and Walton).

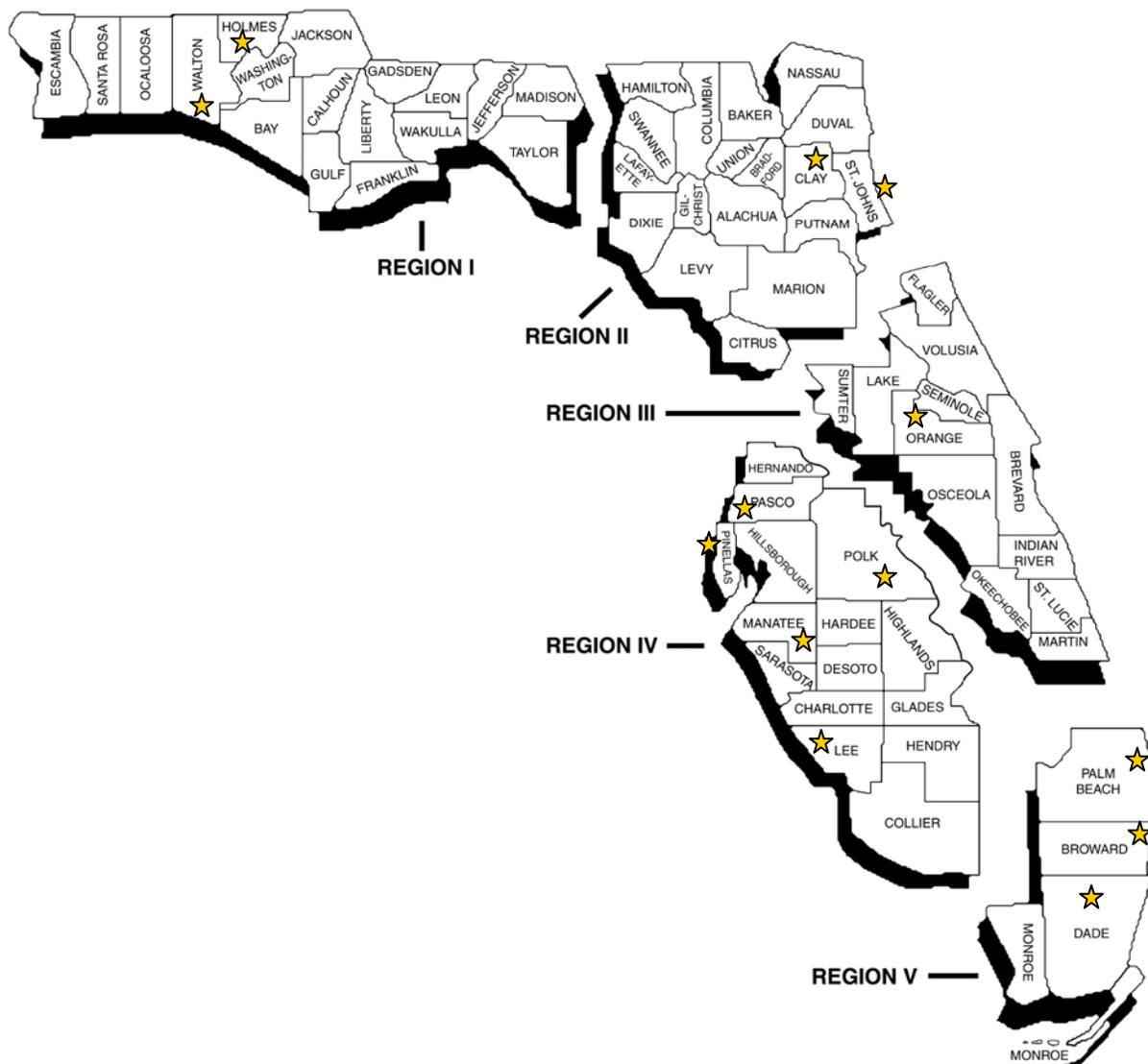


Figure 1. State map designating participating districts by region.

Results

Participants

In order to determine the degree of success in obtaining a sample representative of the state in relation to district size, the number and percent of teachers were compared across three categories: state teachers, target sample, and final sample (see Table 2). For example,

53.54% of teachers within the state teach in large districts. The sampling plan for the field test targeted 53.06% of teacher participants to represent large districts. The actual sample obtained was 44.90% of teacher participants representing large districts. This pattern was similar for medium and small districts, yielding an actual sample that was satisfactorily representative of the state.

Table 2

State Teachers, Target Sample (Requested Participation), and Final Sample (Participated) by District Size

	<u>State Teachers</u>		<u>Target Sample</u>		<u>Final Sample</u>	
	Number of Teachers	Percent of Teachers	Number of Teachers	Percent of Teachers	Number of Teachers	Percent of Teachers
Small Districts	35,986	22.39%	275	22.45%	96	18.00%
Medium Districts	38,667	24.06%	300	24.49%	166	31.10%
Large Districts	86,041	53.54%	650	53.06%	240	44.90%
Alternative District*					32	5.99%
Total	160,694		1,225		534	

*Note: Alternative districts volunteered in addition to the targeted sample.

Although 613 participants started the inventory, not all of them completed all items, and some completed very few items. As a result of this missing data, the researchers chose to restrict the analysis to participants who clearly responded in a motivated fashion. Participants had to complete at least 95% of the total number of items to be included in the analysis. This meant that a participant could be missing up to 5 items overall and still be included. This method resulted in a final sample size of 534 participants. The following descriptions compare the obtained sample of 613 participants with the final sample of 534 participants to illustrate the similarity between the two samples across selected demographic variables.

Table 3 illustrates the number of participants from each of the district size categories. Comparing the percentage within each district size, the final sample was very similar to the obtained sample. For example, the percentage of participants from large districts in the obtained sample is 45.84%, which is similar to the final sample of 44.94%. This data supports the use of the final sample for field test data analyses. Most participants were from the large districts (44.94%), followed by medium size districts (31.09%), and then small districts (17.98%).

Table 3
Comparison of Obtained Sample with Final Sample by District Size

	<u>Obtained Sample</u>		<u>Final Sample</u>	
	<u>(N=613)</u>		<u>(N=534)</u>	
	N	%	N	%
District Size				
Small	120	19.58	96	17.98
Medium	180	29.36	166	31.09
Large	281	45.84	240	44.94
Alternative	32	5.22	32	5.99

Table 4 illustrates the level of computer experience of participants as reported by their district. The majority of participants (41.95%) were classified as having an Intermediate level of computer experience by their district representative, 35.21% were rated as having an Advanced level of computer experience, and 22.66% were rated as having a Beginner level of computer experience. It seems likely that the Beginner group is the smallest since teachers at a Beginner level of computer experience may be less willing to volunteer to take a test of technology literacy. Overall, the three groups were sufficiently representative of the three levels of computer experience. Again, note that the percentage in each computer experience category does not change markedly when the sample is reduced due to missing data.

Table 4
Comparison of Obtained Sample with Final Sample by Computer Experience

	<u>Obtained Sample</u>		<u>Final Sample</u>	
	<u>(N=613)</u>		<u>(N=534)</u>	
	N	%	N	%
Computer Experience				
Beginner	144	23.49	121	22.66
Intermediate	250	40.78	224	41.95
Advanced	218	35.56	188	35.21

Table 5 illustrates the number of participants when classified by level of students taught (i.e., school level). The majority (47.94%) of participants were teachers of elementary students (Kindergarten – 5th grade). Teachers of middle school (6-8th grade) students represented 17.79% of the sample. Teachers of high school (9-12th grade) students

represented 23.03% of the sample. Teachers of grades that overlap in school level were classified as Combination level (e.g., grades 6-10) and made up 2.25% of the sample. Participants classified as Other (8.99%) reported teaching assignments such as resource teacher (grade level not defined) or district office. Statewide and within districts, there are more elementary schools than secondary (e.g., middle or high); therefore, it is reasonable that a majority of teachers of elementary school students would be included in this sample. Once again, the percentages in the final sample are highly similar to those in the initial, obtained sample.

Table 5
Comparison of Obtained Sample with Final Sample by School Level

School Level	<u>Obtained Sample</u> (N=613)		<u>Final Sample</u> (N=534)	
	N	%	N	%
Elementary School	283	46.17	256	47.94
Middle School	125	20.39	95	17.79
High School	140	22.84	123	23.03
Combination	13	2.12	12	2.25
Other	52	8.48	48	8.99

Table 6 illustrates the number of participants grouped by gender and ethnicity. The majority of participants were female (79.59%). This is not surprising since the majority of teachers are female, particularly at the elementary school level which represented 47.94% of the sample. Ethnicity representation was not diverse. The sample was made up mainly of teachers classified as white (86.14%).

Table 6
Comparison of Obtained Sample with Final Sample by Gender and Ethnicity

	<u>Obtained Sample</u>		<u>Final Sample</u>	
	<u>(N=613)</u>		<u>(N=534)</u>	
	N	%	N	%
Gender				
Male	130	21.21	109	20.41
Female	483	78.79	425	79.59
Ethnicity				
White/Non-Hispanic	521	84.99	460	86.14
African American	30	4.89	28	5.24
Hispanic	45	7.34	31	5.81
Native American	4	0.65	4	0.75
Asian/Pacific Islander	5	0.82	4	0.75
Other Ethnicity	8	1.31	7	1.31

These analyses verify that the final sample maintains a very similar distribution across the demographic variables; thus, the remaining analyses were conducted with the final sample (N=534).

Descriptive Data

Mean percent correct scores for each subsection are presented by demographic variables in Table 7. Overall, the mean percent correct scores did not vary a great deal across the demographic variables. Males and females differed by less than 3 percentage points overall. Males scored an average of 4.3% higher than females in the *Basic Operations* subsection. Scores did vary considerably across ethnicity subgroups, although care should be taken in interpreting these values, as several ethnicity subgroups had very few participants. Among ethnicity subgroups, the Native American subgroup received the lowest total mean percent correct (82.8%); however, this subgroup consisted of only 4 participants. The largest subsection score difference was seen for *Basic Operations* between Native Americans (76.9%) and Asian/Pacific Islanders (95.2%). In the school level demographic group, the largest difference (4.7%) was seen on the *Productivity* subsection. In general, the participants who scored the highest were categorized as Other (91.0%), closely followed by teachers of high school students (89.3%), teachers of middle school students (89.1%), and

teachers of elementary school students (88.0%). Mean percent correct values were similar for groups by district size, with the largest difference (8%) occurring on the *Basic Operations* subsection between participants from small districts (85.8%) and participants from alternative districts (93.8%). On most subsections, participants from small districts scored the lowest, and participants from alternative districts scored the highest. It should be noted that most teachers in the alternative group provide online instruction. However, on the *Social, Ethical, Legal, & Human Issues* subsection, participants from small districts scored an average of 2% higher than participants from alternative districts. Finally, as expected, the participants ranked as having Advanced computer experience received the highest mean percent correct scores on all subsections, followed by those participants ranked as having Intermediate skills, and those participants ranked as having Beginner skills. There were large differences (over 10%) between those in the Beginner and Advanced groups on *Basic Operations, Productivity, and Research*. All scores among demographic groups are reasonable for this type of criterion-referenced test (CRT) measuring basic literacy where mastery is expected.

Overall, each of the demographic groups found *Basic Operations* and *Research* to be the most challenging subsections. In contrast, the subsections *Communications* and *Planning, Management, & Instruction* resulted in the highest average scores for all groups. This could indicate that participants are familiar with using tools such as email to communicate with others, and are aware of various forms of technology for planning instruction.

Table 7
Means of Subsection and Total Inventory by Demographic Variables

	<u>Basic</u> <u>Operations</u>	<u>Productivity</u>	<u>Communication</u>	<u>Research</u>	<u>Planning</u>	<u>Social</u> <u>/Ethical</u>	<u>Total</u> <u>Mean</u>	
	Mean %	Mean %	Mean %	Mean %	Mean %	Mean %	Mean %	
	Correct	Correct	Correct	Correct	Correct	Correct	Correct	N
Gender								
Male	92.0%	93.3%	98.4%	89.9%	96.9%	92.5%	90.8%	109
Female	87.7%	90.7%	96.8%	87.2%	95.4%	92.0%	88.3%	425
Ethnicity								
White/Non-								
Hispanic	89.0%	91.5%	97.3%	88.4%	95.8%	92.3%	89.1%	460
African								
American	84.9%	88.3%	95.2%	82.9%	95.1%	87.5%	85.7%	28
Hispanic	86.1%	89.7%	97.2%	83.3%	94.7%	91.7%	87.0%	31

	<u>Basic</u>					<u>Social</u>		
	<u>Operations</u>	<u>Productivity</u>	<u>Communication</u>	<u>Research</u>	<u>Planning</u>	<u>/Ethical</u>	<u>Total</u>	
	Mean	Mean	Mean	Mean	Mean	Mean	Mean	
	%	%	%	%	%	%	%	
	Correct	Correct	Correct	Correct	Correct	Correct	Correct	N
Asian/Pacific								
Islander	95.2%	95.6%	99.6%	88.6%	95.5%	98.0%	92.1%	4
Native								
American	76.9%	90.8%	86.4%	75.0%	93.2%	95.8%	82.8%	4
Other	85.2%	91.7%	93.3%	89.3%	97.4%	92.3%	88.3%	7
District Size								
Small	85.8%	90.3%	96.4%	86.4%	95.0%	92.3%	87.6%	96
Medium	91.0%	92.2%	97.7%	89.2%	96.1%	93.7%	90.0%	166
Large	87.2%	90.5%	97.0%	86.9%	95.5%	91.3%	88.1%	240
Alternative	93.8%	94.8%	98.2%	90.7%	97.5%	90.2%	91.5%	32
School Level								
Elementary	87.8%	90.5%	96.9%	85.8%	95.4%	91.8%	88.0%	256
Middle	87.8%	91.8%	97.9%	89.2%	95.0%	93.7%	89.1%	95
High	89.3%	91.6%	96.4%	89.5%	96.1%	91.8%	89.3%	123
Combination	89.1%	89.0%	99.6%	88.9%	96.2%	90.0%	88.7%	12
Other	92.0%	93.7%	98.4%	90.7%	97.5%	91.5%	91.0%	48
Computer Experience								
Beginner	80.3%	85.0%	94.8%	81.5%	93.3%	89.8%	83.6%	121
Intermediate	88.5%	90.8%	97.6%	86.8%	95.0%	91.8%	88.4%	224
Advanced	93.8%	95.8%	98.1%	92.9%	98.2%	93.8%	92.6%	188

Test Quality

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.

Item Analysis

In classical test theory, item analysis traditionally includes calculations of individual item difficulty (denoted by p) and discrimination (denoted by D). Item difficulty refers to the proportion of examinees giving a correct response. It is reported as a percentage value out of a possible 100. The lower the p value, the harder the item. Item discrimination refers to the ability of a test item to separate those who performed well on the exam overall from those who performed poorly on the exam overall. One method for calculating discrimination is to

correlate item performance with overall test score. Values for item discrimination range from 0 to 1, denoting least to most discriminating.

Item statistics will be reported for two groups; first for the entire group of examinees (Table 8), and second for just the subgroup classified as Beginners by district technology personnel (Table 9). Since the test in question will likely be used primarily with teachers from the Beginner group, it is important to evaluate item performance for this group separately. In both tables, an additional breakdown is provided for item difficulty, separating examinees into quartiles by their overall performance on the test (i.e., lowest performers are in Quartile 1, increasing to the highest performers in Quartile 4). Difficulty values are provided for each item separately by the group of examinees in each quartile.

Table 8
Item Difficulty, Discrimination, and Difficulty by Quartile for Entire Group (N=534)

	Seq. #	Item Type	Item Difficulty (<i>p</i>)	Item Discrim. (<i>D</i>)	Quartile 1 % Correct	Quartile 2 % Correct	Quartile 3 % Correct	Quartile 4 % Correct
Basic Operations								
	1	MC	99.25%	0.24	97.00%	100.00%	100.00%	100.00%
	2	MC	97.00%	0.23	91.10%	98.70%	98.20%	100.00%
	3	Sim	98.50%	0.20	95.60%	98.70%	100.00%	100.00%
	4	Sim	92.32%	0.45	78.50%	94.00%	97.30%	100.00%
	5	Sim	86.39%	0.38	71.00%	83.30%	92.80%	99.30%
	6	Sim	77.19%	0.57	36.90%	86.50%	88.30%	96.40%
	7	Sim	99.10%	0.27	95.70%	100.00%	100.00%	100.00%
	8	Sim	80.34%	0.62	39.10%	89.20%	92.80%	99.30%
	9	Sim	96.57%	0.11	94.50%	94.60%	98.20%	99.30%
	10	Sim	93.94%	0.46	79.60%	95.30%	97.30%	100.00%
	11	Sim	79.23%	0.42	58.00%	73.60%	84.70%	96.40%
	12	Sim	91.00%	0.49	67.50%	97.30%	99.10%	99.30%
	13	Sim	95.56%	0.39	81.80%	97.30%	100.00%	100.00%
	14	Sim	89.64%	0.29	77.30%	87.40%	98.20%	97.10%
Productivity								
	15	MC	98.69%	0.18	96.30%	100.00%	99.10%	99.30%
	16	MC	99.44%	0.10	98.50%	100.00%	99.10%	100.00%
	17	MC	99.25%	0.26	97.00%	100.00%	100.00%	100.00%
	18	MC	98.50%	0.1	95.50%	99.30%	100.00%	99.30%

Seq. #	Item Type	Item Difficulty (<i>p</i>)	Item Discrim. (<i>D</i>)	Quartile 1 % Correct	Quartile 2 % Correct	Quartile 3 % Correct	Quartile 4 % Correct
19	MC	97.75%	0.15	94.00%	97.40%	100.00%	100.00%
20	MC	100.00%		100.00%	100.00%	100.00%	100.00%
21	MC	70.60%	0.29	51.10%	60.30%	82.90%	91.20%
22	MC	85.96%	0.41	65.90%	82.10%	99.10%	99.30%
23	Sim	97.00%	0.33	91.10%	97.40%	100.00%	100.00%
24	Sim	96.44%	0.39	89.60%	98.00%	99.10%	99.30%
25	Sim	97.19%	0.33	91.10%	98.70%	99.10%	100.00%
26	Sim	97.94%	0.36	92.60%	100.00%	100.00%	99.30%
27	Sim	91.95%	0.5	68.90%	100.00%	99.10%	100.00%
28	Sim	92.31%	0.51	69.40%	100.00%	100.00%	100.00%
29	Sim	80.64%	0.23	69.40%	78.70%	82.00%	92.70%
30	Sim	84.40%	0.36	70.10%	81.30%	88.30%	98.50%
31	Sim	92.51%	0.43	71.90%	100.00%	98.20%	100.00%
32	Sim	30.34%	0.23	12.60%	32.50%	28.80%	46.70%
33	Sim	88.76%	0.39	67.40%	92.70%	97.30%	98.50%
34	Sim	86.49%	0.34	66.40%	88.10%	94.60%	97.80%
35	Sim	94.00%	0.44	76.10%	100.00%	100.00%	100.00%
36	Sim	85.55%	0.44	61.90%	85.40%	96.40%	100.00%
37	Sim	86.87%	0.55	59.70%	90.10%	99.10%	100.00%
38	Sim	94.48%	0.43	81.50%	96.50%	100.00%	99.30%
39	Sim	96.83%	0.41	87.70%	99.30%	100.00%	100.00%
40	Sim	87.62%	0.48	61.90%	93.40%	95.50%	100.00%
Communications							
41	MC	99.81%	0.03	99.30%	100.00%	100.00%	100.00%
42	MC	98.31%	0.15	95.60%	99.30%	100.00%	98.50%
43	Sim	98.69%	0.09	97.00%	98.70%	99.10%	100.00%
44	Sim	94.94%	0.05	95.60%	92.10%	94.60%	97.80%
45	Sim	97.37%	0.18	92.50%	97.40%	100.00%	100.00%
46	Sim	98.50%	0.25	95.50%	99.30%	99.10%	100.00%
47	Sim	96.63%	0.07	96.30%	96.70%	95.50%	97.80%
48	Sim	99.25%	0.15	97.80%	99.30%	100.00%	100.00%
49	Sim	99.25%	0.13	97.80%	100.00%	99.10%	100.00%
50	Sim	99.06%	0.25	97.80%	100.00%	98.20%	100.00%
51	Sim	95.69%	0.21	88.10%	96.70%	98.20%	100.00%
52	Sim	99.63%	0.19	98.50%	100.00%	100.00%	100.00%
53	Sim	99.44%	0.03	98.50%	99.30%	100.00%	100.00%
54	Sim	89.70%	0.24	80.70%	90.70%	90.10%	97.10%
55	Sim	99.62%	0.19	98.50%	100.00%	100.00%	100.00%
56	Sim	94.94%	0.32	85.90%	97.40%	96.40%	100.00%
57	Sim	97.38%	0.29	91.10%	98.70%	100.00%	100.00%

Seq. #	Item Type	Item Difficulty (p)	Item Discrim. (D)	Quartile 1 % Correct	Quartile 2 % Correct	Quartile 3 % Correct	Quartile 4 % Correct
Research							
58	Sim	93.43%	0.33	85.10%	94.70%	93.70%	100.00%
59	MC	96.80%	0.29	90.20%	97.40%	100.00%	100.00%
60	MC	98.69%	0	97.80%	98.70%	100.00%	98.50%
61	MC	91.95%	0.25	84.40%	87.40%	97.30%	100.00%
62	MC	46.44%	0.38	22.20%	35.80%	53.20%	76.60%
63	MC	82.40%	0.28	68.90%	76.20%	88.30%	97.80%
64	MC	93.43%	0.32	83.70%	92.70%	99.10%	99.30%
65	MC	98.13%	0.34	93.30%	99.30%	100.00%	100.00%
66	MC	97.37%	0.39	91.10%	100.00%	98.20%	100.00%
67	MC	96.82%	0.16	92.60%	96.00%	99.10%	100.00%
68	MC	97.00%	0.23	91.80%	97.40%	99.10%	100.00%
69	Sim	94.01%	0.22	88.10%	95.40%	93.70%	98.50%
70	Sim	87.05%	0.41	65.70%	88.70%	96.40%	98.50%
71	Sim	70.92%	0.43	44.00%	63.60%	84.70%	94.20%
Planning, Management, & Instruction							
72	MC	91.95%	0.17	85.20%	94.00%	95.50%	93.40%
73	MC	98.50%	0.19	95.60%	98.70%	100.00%	100.00%
74	MC	99.44%	0.16	98.50%	100.00%	99.10%	100.00%
75	MC	95.50%	0.35	85.90%	97.30%	99.10%	100.00%
76	MC	99.25%	0.11	98.50%	99.30%	99.10%	100.00%
77	MC	94.18%	0.2	91.10%	92.00%	94.60%	99.30%
78	MC	94.74%	0.22	90.40%	96.00%	93.70%	98.50%
79	MC	94.19%	0.21	88.10%	93.40%	95.50%	100.00%
80	MC	89.14%	0.2	81.50%	84.80%	94.60%	97.10%
81	MC	98.50%	0.11	97.00%	98.00%	99.10%	100.00%
82	MC	98.13%	0.21	93.30%	99.30%	100.00%	100.00%
Social, Ethical, Legal, & Human Issues							
83	MC	100.00%		100.00%	100.00%	100.00%	100.00%
84	MC	93.63%	0.08	91.90%	90.70%	98.20%	94.90%
85	MC	99.06%	0.03	99.30%	98.70%	98.20%	100.00%
86	MC	99.63%	0.13	98.50%	100.00%	100.00%	100.00%
87	MC	99.81%	-0.03	100.00%	100.00%	99.10%	100.00%
88	MC	99.25%	0.06	98.50%	100.00%	99.10%	99.30%
89	MC	97.57%	0.09	96.30%	94.70%	100.00%	100.00%
90	MC	98.88%	0.08	97.00%	98.70%	100.00%	100.00%
91	MC	96.44%	0.20	88.90%	98.00%	99.10%	100.00%
92	MC	99.25%	0.20	97.00%	100.00%	100.00%	100.00%
93	MC	98.50%	0.09	97.00%	98.70%	99.10%	99.30%
94	MC	90.45%	0.23	82.20%	87.40%	94.60%	98.50%

Seq. #	Item Type	Item Difficulty (<i>p</i>)	Item Discrim. (D)	Quartile 1 % Correct	Quartile 2 % Correct	Quartile 3 % Correct	Quartile 4 % Correct
95	MC	94.94%	-0.01	96.30%	92.10%	96.40%	95.60%
96	MC	87.45%	0.09	83.70%	85.40%	85.60%	94.90%
97	MC	92.88%	0.20	86.70%	92.10%	94.60%	98.50%
98	MC	81.09%	0.04	83.70%	80.10%	78.40%	81.80%
99	MC	98.88%	0.15	97.00%	98.70%	100.00%	100.00%
100	MC	98.50%	0.17	97.00%	98.00%	99.10%	100.00%
101	MC	99.44%	0.16	98.50%	99.30%	100.00%	100.00%
102	MC	97.57%	0.28	94.10%	96.70%	100.00%	100.00%
103	MC	71.91%	0.24	59.30%	62.30%	75.70%	92.00%

The second column, headed “Item Type”, indicates whether each item was a selected-response multiple choice item (MC) or a step in a performance-based simulation task (Sim).

In the item statistics presented here, the missing examinee responses were not scored as incorrect, but rather were dropped from the analyses. This could artificially inflate the *p* values, i.e., making some items appear easier than they actually were. Further discussion of these missing values is provided in the summary section.

For the entire group of examinees, difficulty values ranged from 30.34 % to 100%. The bulk of the items (78) had *p* values ranging between 90% and 99%. Of the remaining items, seven items had *p* values below 80%, 16 items had *p* values from 80% to 89%, and two items had *p* values of 100%. This indicates that for the entire group of examinees, the test was, overall, easy. When looking at item analyses for just those examinees designated as Beginners, *p* values in general dropped, indicating that the items were more difficult for this group. For example, there were 18 items with *p* values below 80% (see Table 9).

Two items of concern, as identified by item analyses, were #32 and #62, with *p* values of 30.34% and 46.44%, respectively. Upon investigation, #32 is a simulation item asking examinees to edit data in an existing spreadsheet; specifically, highlighting or selecting a set of data. This was not intended to be a difficult task, but it appears that something is causing it to function that way. This item should be investigated further and should be modified. Item #62 is a multiple choice item asking about research strategies. Examination of the discrimination index for this item shows that it was very effective in separating the high and low scoring examinees. These results suggest that the item is simply

difficult, and may cover an appropriate concept that many teachers nevertheless find challenging. [Distractor analysis reveals that all options were chosen by some examinees, and they appear to be functioning acceptably.]

Discrimination values ranged from -0.01 to 0.57. A number of the items had low D values, including two that had values just below zero. However, their extremely small values are of little concern, since they accompany very high p values. When almost an entire set of examinees gets an item right (or wrong) an item is not able to show a great deal of discrimination. In general, the high p values and low D values seen in this inventory 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 more high D values would be expected.

Table 9
Item Difficulty, Discrimination, and Difficulty by Quartile for Beginners Only (N=121)

Seq. #	Item Type	Item Difficulty (p)	Item Discrim. (D)	Quartile 1 % Correct	Quartile 2 % Correct	Quartile 3 % Correct	Quartile 4 % Correct
Basic Operations							
1	MC	98.35%	0.34	93.30%	100.00%	100.00%	100.00%
2	MC	92.56%	0.24	83.30%	89.70%	97.20%	100.00%
3	Sim	96.69%	0.17	93.30%	96.60%	97.20%	100.00%
4	Sim	83.47%	0.53	60.00%	82.80%	91.70%	100.00%
5	Sim	76.92%	0.43	58.60%	66.70%	85.70%	96.20%
6	Sim	69.83%	0.58	28.60%	66.70%	85.70%	96.20%
7	Sim	97.87%	0.34	89.50%	100.00%	100.00%	100.00%
8	Sim	63.79%	0.63	14.30%	51.90%	88.60%	96.20%
9	Sim	92.24%	0.1	89.30%	88.90%	91.40%	100.00%
10	Sim	86.36%	0.51	59.10%	84.60%	94.40%	100.00%
11	Sim	62.50%	0.48	25.00%	61.50%	72.20%	84.60%
12	Sim	82.35%	0.46	50.00%	79.30%	97.20%	100.00%
13	Sim	87.39%	0.43	60.90%	84.60%	97.20%	100.00%
14	Sim	78.99%	0.32	67.90%	72.40%	77.80%	100.00%
Productivity							
15	MC	96.69%	0.22	93.30%	93.10%	100.00%	100.00%
16	MC	99.17%	0.18	96.70%	100.00%	100.00%	100.00%
17	MC	97.50%	0.26	93.10%	96.60%	100.00%	100.00%
18	MC	96.64%	0.12	92.90%	93.10%	100.00%	100.00%

Seq. #	Item Type	Item Difficulty (<i>p</i>)	Item Discrim. (<i>D</i>)	Quartile 1 % Correct	Quartile 2 % Correct	Quartile 3 % Correct	Quartile 4 % Correct
19	MC	95.04%	0.16	90.00%	96.60%	94.40%	100.00%
20	MC	100.00%		100.00%	100.00%	100.00%	100.00%
21	MC	59.50%	0.19	50.00%	51.70%	52.80%	88.50%
22	MC	70.25%	0.37	50.00%	62.10%	72.20%	100.00%
23	Sim	91.74%	0.36	80.00%	93.10%	94.40%	100.00%
24	Sim	93.39%	0.51	80.00%	93.10%	100.00%	100.00%
25	Sim	92.56%	0.45	76.70%	96.60%	97.20%	100.00%
26	Sim	95.87%	0.33	83.30%	100.00%	100.00%	100.00%
27	Sim	87.60%	0.47	60.00%	89.70%	100.00%	100.00%
28	Sim	89.17%	0.43	65.50%	89.70%	100.00%	100.00%
29	Sim	70.59%	0.26	55.20%	65.50%	74.30%	88.50%
30	Sim	76.47%	0.44	48.30%	79.30%	85.70%	92.30%
31	Sim	90.08%	0.29	73.30%	86.20%	100.00%	100.00%
32	Sim	23.14%	0.17	10.00%	17.20%	38.90%	23.10%
33	Sim	86.78%	0.24	73.30%	82.80%	94.40%	96.20%
34	Sim	82.64%	0.21	73.30%	75.90%	88.90%	92.30%
35	Sim	90.91%	0.38	70.00%	93.10%	100.00%	100.00%
36	Sim	75.21%	0.51	43.30%	72.40%	86.10%	100.00%
37	Sim	66.94%	0.6	26.70%	58.60%	83.30%	100.00%
38	Sim	85.32%	0.57	53.80%	91.70%	93.90%	100.00%
39	Sim	90.83%	0.52	65.40%	96.30%	100.00%	100.00%
40	Sim	66.12%	0.47	30.00%	58.60%	80.60%	96.20%
Communications							
41	MC	100.00%		100.00%	100.00%	100.00%	100.00%
42	MC	95.87%	0.27	90.00%	93.10%	100.00%	100.00%
43	Sim	98.35%	0.12	93.30%	100.00%	100.00%	100.00%
44	Sim	92.56%	-0.01	96.70%	93.10%	88.90%	92.30%
45	Sim	95.87%	0.06	96.70%	86.20%	100.00%	100.00%
46	Sim	98.35%	0.36	93.30%	100.00%	100.00%	100.00%
47	Sim	96.69%	0.16	96.70%	89.70%	100.00%	100.00%
48	Sim	99.17%	0.21	96.70%	100.00%	100.00%	100.00%
49	Sim	99.17%	-0.01	100.00%	96.60%	100.00%	100.00%
50	Sim	97.52%	0.26	93.30%	100.00%	100.00%	96.20%
51	Sim	91.74%	0.14	86.70%	86.20%	97.20%	96.20%
52	Sim	99.17%	0.21	96.70%	100.00%	100.00%	100.00%
53	Sim	99.17%	0.01	100.00%	96.60%	100.00%	100.00%
54	Sim	85.12%	0.28	73.30%	82.80%	94.40%	88.50%
55	Sim	99.17%	0.21	96.70%	100.00%	100.00%	100.00%
56	Sim	89.26%	0.45	66.70%	93.10%	97.20%	100.00%
57	Sim	93.39%	0.28	86.70%	89.70%	97.20%	100.00%

Seq. #	Item Type	Item Difficulty (<i>p</i>)	Item Discrim. (<i>D</i>)	Quartile 1 % Correct	Quartile 2 % Correct	Quartile 3 % Correct	Quartile 4 % Correct
58	Sim	86.78%	0.39	63.30%	96.60%	88.90%	100.00%
Research							
59	MC	93.28%	0.38	78.60%	93.10%	100.00%	100.00%
60	MC	100.00%		100.00%	100.00%	100.00%	100.00%
61	MC	85.95%	0.35	73.30%	79.30%	91.70%	100.00%
62	MC	35.54%	0.47	6.67%	20.70%	44.40%	73.10%
63	MC	78.51%	0.18	70.00%	75.90%	75.00%	96.20%
64	MC	89.26%	0.39	70.00%	93.10%	94.40%	100.00%
65	MC	93.39%	0.36	83.30%	93.10%	97.20%	100.00%
66	MC	91.74%	0.42	76.70%	93.10%	100.00%	96.20%
67	MC	93.39%	0.17	86.70%	89.70%	97.20%	100.00%
68	MC	95.00%	0.26	89.70%	93.10%	100.00%	96.20%
69	Sim	92.56%	0.31	80.00%	93.10%	97.20%	100.00%
70	Sim	76.03%	0.34	50.00%	72.40%	88.90%	92.30%
71	Sim	48.76%	0.36	23.30%	41.40%	52.80%	80.80%
Planning, Management, & Instruction							
72	MC	89.26%	0.24	80.00%	82.80%	97.20%	96.20%
73	MC	97.52%	0.35	93.30%	96.60%	100.00%	100.00%
74	MC	98.35%	0.28	93.30%	100.00%	100.00%	100.00%
75	MC	89.26%	0.31	73.30%	86.20%	97.20%	100.00%
76	MC	98.35%	0.15	96.70%	100.00%	100.00%	96.20%
77	MC	92.56%	0.19	86.70%	96.60%	94.40%	92.30%
78	MC	92.56%	0.29	83.30%	93.10%	100.00%	92.30%
79	MC	90.08%	0.14	86.70%	86.20%	94.40%	92.30%
80	MC	84.30%	0.22	70.00%	82.80%	88.90%	96.20%
81	MC	97.52%	0.2	90.00%	100.00%	100.00%	100.00%
82	MC	95.87%	0.18	90.00%	93.10%	100.00%	100.00%
Social, Ethical, Legal, & Human Issues							
83	MC	100.00%		100.00%	100.00%	100.00%	100.00%
84	MC	92.56%	0.02	93.30%	89.70%	94.40%	92.30%
85	MC	97.52%	-0.01	96.70%	100.00%	97.20%	96.20%
86	MC	98.35%	0.15	96.70%	96.60%	100.00%	100.00%
87	MC	100.00%		100.00%	100.00%	100.00%	100.00%
88	MC	97.52%	0.07	93.30%	100.00%	100.00%	96.20%
89	MC	97.52%	0.04	100.00%	93.10%	97.20%	100.00%
90	MC	98.35%	0.05	96.70%	96.60%	100.00%	100.00%
91	MC	94.21%	0.06	96.70%	86.20%	97.20%	96.20%
92	MC	99.17%	0.1	96.70%	100.00%	100.00%	100.00%
93	MC	99.17%	0.02	100.00%	96.60%	100.00%	100.00%
94	MC	86.78%	0.21	80.00%	82.80%	88.90%	96.20%

Seq. #	Item Type	Item Difficulty (p)	Item Discrim. (D)	Quartile 1 % Correct	Quartile 2 % Correct	Quartile 3 % Correct	Quartile 4 % Correct
95	MC	94.21%	-0.09	96.70%	100.00%	86.10%	96.20%
96	MC	81.82%	0.03	76.70%	79.30%	86.10%	84.60%
97	MC	88.43%	0.29	73.30%	86.20%	97.20%	96.20%
98	MC	85.12%	0.29	63.30%	89.70%	91.70%	96.20%
99	MC	99.17%	0.3	96.70%	100.00%	100.00%	100.00%
100	MC	97.52%	0.35	90.00%	100.00%	100.00%	100.00%
101	MC	99.17%	0.14	96.70%	100.00%	100.00%	100.00%
102	MC	92.56%	0.39	76.70%	100.00%	94.40%	100.00%
103	MC	64.46%	0.31	50.00%	55.20%	61.10%	96.20%

Average difficulty by subsection is shown in Table 10. For the entire group, the most difficult subsection was *Research*, followed by: *Productivity*; *Basic Operations*; *Social, Ethical, Legal, & Human Issues*; *Planning, Management, & Instruction*; and finally, *Communications*. In all cases, difficulty values were lower for the Beginner group, indicating that the items were harder for these less skilled examinees. The ranking across subsections was the same for the whole group and Beginner group for all except *Planning, Management, & Instruction*, and *Social, Ethical, Legal, & Human Issues*. The Beginner group found *Social, Ethical, Legal, & Human Issues* to be slightly more difficult than *Planning, Management, & Instruction*.

Table 10
Average Item Difficulty by Subsection

Subsection	Entire Group	Item Difficulty (p) Beginners Only
Basic Operations	91.15%	83.52%
Productivity	89.67%	83.47%
Communications	97.31%	95.41%
Research	88.54%	82.57%
Planning, Management, & Instruction	95.77%	93.24%
Social, Ethical, Legal, & Human Issues	95.01%	93.51%

An important issue in the design of the inventory was an awareness of the amount of time that would be required of a teacher to complete the assessment. By allowing subsections to be completed independently, examinees were freed from having to commit a large block of continuous time in order to finish. Based on focus group feedback, a maximum length of 15-20 minutes per subsection was targeted. As shown in Table 11, the median time per subsection was well within the targeted time for all subsections. The median time for completion of the entire inventory was 36.25 minutes.

The measure of central tendency reported here is the median, as it is least affected by extreme values, such as some of those appearing here. Completion times for the entire inventory ranged from 15.63 minutes to 137 hours. This extremely high maximum value is probably due to an examinee leaving the computer for a number of hours without logging out of the inventory. The subsection with the shortest median completion time was *Planning, Management, & Instruction*, at 1.93 minutes. This subsection consisted of only 11 multiple choice items, and should have required the least amount of time. However, this timing data does indicate that more items, especially more difficult items, could be added to this subsection without causing any time problems. *Social, Ethical, Legal, & Human Issues* took the most time on average, at 11.22 minutes. This subsection requires more reading than the others, as well as a larger number of binary (two-option) items. The remaining subsections had median values from approximately 4 to 7 minutes.

Table 11
Timing Data by Subsection

	Median	Minimum	Maximum	Std. Dev.
	<i>(all values expressed in minutes)</i>			
Total Inventory	36.25	15.63	8220.28	355.03
Basic Operations	4.61	1.83	118.78	7.00
Productivity	6.87	2.83	205.73	14.54
Communications	3.87	1.73	226.60	14.05
Research	4.82	1.17	8190.98	354.27
Planning, Management, & Instruction	1.93	0.88	87.40	4.22
Social, Ethical, Legal, & Human Issues	11.22	3.33	227.85	14.01

Reliability.

Following item analysis, a further measure of test quality is reliability, or consistency of scores. In this case, reliability was estimated using a measure of internal consistency known as Cronbach's alpha. For an assessment in which subsection scores are reported, it is necessary to estimate reliability of the scores at both the subsection and entire test levels. As shown in Table 12, the reliability estimate for scores on the entire inventory is .79. This is quite good, especially considering the low levels of correlation between the subsections (see Table 13). At the subsection level, score reliability estimates range from .33 for *Planning, Management, & Instruction* to .75 for *Productivity*. Relatively low reliability estimates for subsections such as *Communications* and *Planning, Management, & Instruction* can be attributed to the fact that both subsections have a small number of items, and there was little variability in performance within each subsection, i.e., most examinees did quite well on all items in these subsections. The modest reliability estimates for *Research, Basic Operations*, and *Social, Ethical, Legal, & Human Issues* are likely due as well to the small number of items per subsection, and the lack of variability of item performance within a subsection. In general, subsection reliability estimates are lower than the reliability estimate of the entire test, simply because they are, by definition, shorter. As discussed earlier, this assessment is performing 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.

Table 12
Internal Consistency Reliability Estimates

	N	Non-Standardized Cronbach Alpha
Basic Operations	404	0.6197
Productivity	474	0.7458
Communications	530	0.4804
Research	527	0.5249
Planning, Management, & Instruction	532	0.3291
Social, Ethical, Legal, & Human Issues	534	0.6302
Total Inventory	361	0.7865

Note: Participants were deleted from analyses if they were missing any one of the items in the calculation of the alpha (hence the varying values for N).

Inter-section correlations are presented in Table 13. As Thompson (2003) points out, variability is a key to reliability, including co-variability. The degree to which subsection scores correlate is an indication of the cohesiveness of the construct. For the *Teacher Technology Literacy Inventory*, inter-section correlations show positive relationships between subsection scores and with scores on the total inventory. Correlations of individual subsections with the total inventory range from .36 (*Social, Ethical, Legal, & Human Issues*) to .85 (*Productivity*). Those subsections showing the strongest relationship to the total are *Productivity, Basic Operations, and Research*. The fact that *Social, Ethical, Legal, & Human Issues* has the weakest relationship to the total score is probably due to the fact that it measures a somewhat different construct from the other subsections; although an important part of technology literacy, this construct is not necessarily related to high achievement in the other areas. Correlations between scores on the individual subsections are generally positive but modest. As expected, *Social, Ethical, Legal, & Human Issues* has the lowest relationships to the other subsections, for reasons described previously.

In addition to providing explanations for some of the lower subsection reliability estimates, these modest inter-section correlations highlight the fact that reporting a single test score, or using only an aggregate test score for classification decisions, would be ill-advised.

Table 13

Correlations Between Scores On Inventory Subsections (N=534)

	Basic Operations	Productivity	Communi- cations	Research	Planning	Social /Ethical	Total
Basic Operations	1.0000	0.6500	0.3505	0.4906	0.3271	0.1506	0.8431
Productivity	0.6500	1.0000	0.3730	0.5610	0.3583	0.2242	0.8521
Communications	0.3505	0.3730	1.0000	0.2994	0.3518	0.2219	0.5265
Research	0.4906	0.5610	0.2994	1.0000	0.3627	0.2607	0.7675
Planning	0.3271	0.3583	0.3518	0.3627	1.0000	0.1606	0.5567
Social/Ethical	0.1506	0.2242	0.2219	0.2607	0.1606	1.0000	0.3582
Total	0.8431	0.8521	0.5265	0.7675	0.5567	0.3582	1.0000

Validity.

To estimate construct validity of scores from this assessment, two approaches were undertaken. The first was to investigate the relationship between scores on the *Teacher Technology Literacy Inventory* and the rating of computer experience provided by district personnel. The second was to examine the degree of association between scores on the inventory and scores from the survey designed to measure technology uses and attitudes.

Correlation of subsection total scores with district computer experience level ratings revealed positive relationships, indicating that those scoring higher on the inventory were rated as having a higher computer experience level (see Table 14). The strongest relationship is seen with the *Productivity* subsection. This is probably due to the fact that this subsection measures knowledge and skills that are those most likely to be observed by others. The lowest correlation, with *Social, Ethical, Legal, & Human Issues*, provides further support for the hypothesis that this subsection measures something very different from the others; a construct that is not strongly related to computer skills or experience.

Table 14
Correlation between Computer Experience (As Rated by District Personnel) and Score on Inventory (N=533)*

	Computer Experience Level (as rated by district personnel)
Basic Operations	0.3248
Productivity	0.4457
Communications	0.1916
Research	0.3671
Planning, Management, & Instruction	0.2607
Social, Ethical, Legal, & Human Issues	0.1694
Total Inventory	0.4517

*Note: One participant did not have an experience level rating.

As part of the field test, teachers were asked to complete a survey that included questions related to their uses of and attitudes toward technology (see Barron et al., 2003; Hogarty et al., 2003). The survey is organized in six sections: (a) Confidence and Comfort Using Computers, (b) Attitudes Towards Computer Use, (c) Teacher Confidence Level, and (d) Integration of Computers into the Classroom. Teachers are asked to respond to prompts on a five-point scale to denote agreement (strongly agree to strongly disagree) or frequency of use (every day to not at all), as appropriate. The first section, referred to as *Comfort*,

contains items regarding a teacher's comfort level and perceived effectiveness with technology use. The next section, *Attitudes*, includes statements regarding perceptions of computers and their use in the school environment, as well as their use in general. The *Confidence* section contains a list of technology tasks that a teacher might perform on a regular basis. Teachers are asked to respond as to their level of confidence in completing these tasks. The final section, *Classroom*, includes descriptions of specific technology uses in the classroom, and asks teachers to respond as to how frequently they use technology in each of these ways.

Correlations of inventory subsection scores with scores on the various sections of the survey are provided in Table 15. The strongest relationship is between the *Productivity* subsection of the inventory and the *Confidence* section of the survey (.57). This section contained items that asked teachers to rate their comfort level with regard to performing various technology-related tasks. These tasks were very similar to those measured through simulations in the *Productivity* subsection of the inventory. The weakest relationships were with the *Social, Ethical, Legal, & Human Issues* and *Communications* subsections of the inventory. Since correlation is dependent upon variability, the similarity of performance on the *Communications* subsection (i.e., nearly all examinees did similarly well), would cause any correlation coefficients to be suppressed. Again, as the *Social, Ethical, Legal, and Human Issues* subsection appears to be tapping a different part of the technology literacy construct than the other subsections, as well as any items on the survey, its weaker relationship to the survey sections is to be expected and should not be interpreted as a weakness of that subsection.

Table 15
Correlation between the Inventory and the Technology Attitudes and Uses Survey for Total and Subsections (N=534)

	Basic Operations	Productivity	Communi- cations	Research	Planning	Social /Ethical	Inventory Total
Comfort	0.2259	0.3882	0.1582	0.2938	0.1818	0.0702	0.3479
Attitude	0.2576	0.3906	0.2142	0.2815	0.3080	0.1157	0.3884
Confidence	0.3598	0.5675	0.2056	0.4419	0.2360	0.1866	0.5251
Classroom	0.1565	0.3143	0.0789	0.2107	0.1517	0.0908	0.2627
Survey Total	0.2828	0.4880	0.1715	0.3550	0.2430	0.1416	0.4379

Summary of Evaluation

An assessment tool or instrument is always developed according to a particular purpose. In this case, according to the RFP, the purpose of the inventory was related to creating a practical tool for monitoring technology literacy. In other words, the purpose of this inventory was determined in advance to be low-stakes applications. Appropriate low-stakes applications include the identification of professional development needs by individual teachers who are seeking to improve their skills, the identification by a district of the professional development needs of their staff, and reporting for monitoring purposes in which no high-stakes consequences are attached.

Based on this stated test purpose, appropriate test development processes were followed throughout the project. For the initial stage of construct definition, the indicators were carefully developed with attention to state and national technology standards. Furthermore, the inventory was designed to be integrated within the STaR System. Sound test development procedures were followed for item writing and review, and, extensive usability analyses were conducted to ensure that the user interface, particularly of the novel performance-based tasks, was as clear and intuitive as possible. Finally, the field test was conducted according to a sound research design and sampling plan.

In addition, the results discussed earlier in this paper show the final product of this grant, the inventory, to be a sound assessment tool. Test quality was evaluated by considering item analyses, test reliability, and test validity. On average, the item p values showed the inventory to be quite easy. Like many basic skills assessments, this inventory is functioning like a CRT or a mastery test. That is, most examinees will do quite well overall. This is not a flaw in the test; however, when the majority of examinees perform very similarly on an item (in this case, by most examinees performing well) this results in low item discriminations.

In support of this interpretation, the item differential indices are lower (i.e., harder) for those examinees who were rated as being at a Beginner level of computer experience. However, there is another possible explanation for the overall high performance of examinees of the inventory. Given that the inventory was administered in unsecured, unproctored settings, examinees may have received help as they took the assessment. Again, this is in keeping with the stated low-stakes test purpose, which included a specification that teachers be able to complete the inventory at any time or at any place.

Reliability was estimated for the inventory for each subsection and for the total score, using internal consistency, or Cronbach's alpha reliability. All reliability estimates were found to be positive, but relatively modest. A similar range of performance on items, as seen in the item analysis, can reduce the value of reliability estimates. The modest reliability estimates reported may also be attributed to the low number of items in each subsection. Total score reliability is computed on the full set of items and is thus higher than any subsection reliability. In the case of the inventory, total score reliability may have increased even more, if the inventory subsections had been correlated more strongly with each other.

Two validity studies designed to address content and construct validity were conducted. First, performance on the six inventory subsections was correlated with a rating on each examinee's level of computer experience. The highest correlation with computer experience level was found for *Productivity* and the lowest for *Social, Ethical, Legal, and Human Issues*. Next, correlations were computed between the six subsections and each examinee's response to a self-report measure of technology attitudes and uses. All of these correlations show positive but modest relationships. Once again, the strongest correlations

were found for the *Productivity* subsection; in most cases, the lowest correlations were found for *Social, Ethical, Legal, and Human Issues*.

These results confirm the findings for the inter-section correlations, where *Productivity* was most strongly related to the other five subsections of the inventory, and *Social, Ethical, Legal, and Human Issues* was least related. In fact, the consistent pattern of different results for the six subsections indicates strongly that the inventory is measuring very distinct aspects of teacher technology literacy.

Recommendations

Results from the field test bring to light several key recommendations. The first is in regard to specific items that should either be deleted or modified. Based on item analysis data, item # 32 in the *Productivity* subsection should be very carefully examined before being used in an operational setting, due to its p value of 30.34, indicating that only 30% of the examinees responded correctly to this item. This is a simulation item that asks examinees to select a set of scores in a spreadsheet. Given that the item was not intended to be this difficult, it should be investigated more thoroughly. First of all, the scoring algorithm should be examined to verify that it is functioning properly. If it is, then the incorrect responses given by the examinees should be evaluated to determine what types of actions are being taken. If both of these are in order, then the step should be revised, possibly by making directions more explicit, or allowing for a larger acceptable selection area.

The second recommendation is in regard to a topic where items need to be added. The two indicators pertaining to databases were designed to be measured through completing the registration process for the inventory (i.e., entering data into a database and editing data in an existing database). For the field test, participants were registered by their districts, and portions of the survey were used to satisfy these indicators. Since this survey will not be administered along with the inventory when it becomes operational, items should be written to address these indicators. Furthermore, if an item bank is constructed for this inventory, it may be desirable to add some more difficult items to the *Communications and Planning, Management, & Instruction* subsections, in an effort to decrease the average p values of these subsections closer to an 80% level for the Beginner group. However, it is important to note

that any policy decisions that are made regarding the setting of mastery levels or “cut scores” should be done after any changes are made that would affect subsection and test difficulties.

The third recommendation involves issues of missing data in instances where it was not anticipated. The inventory was designed to require an examinee to either submit an answer or action, or to specifically choose to omit an item or step. If an examinee saw an item but chose not to respond (by omitting), the item was marked as incorrect, and upon the next login, the examinee was automatically advanced to the next item. Examinees were required to complete an entire subsection before being able to begin another subsection. Due to these design constraints, there should not have been any instances of missing data. However, as Table 16 shows, there were several items for which large numbers of examinees did not have a recorded response (either correct/incorrect or omit). In these cases, the examinees did not simply quit the subsection or inventory altogether, in which case missing values would be expected, but instead had recorded responses for items subsequent to those in the table.

Table 16
Items with Large Numbers of Missing Values

Item Number	Item Code	Item Description	Number of Missing Values (out of 534)
7	BOPS5	Save a document	89
10	BOFM1	Create a new folder	39
11	BOFM2	Rename a file	38
12	BOFM3	Move a file into a folder	12
13	BOFM4	Rename a folder	39
38	PROWP21	Add a graphic	27
39	PROWP22	Resize a graphic	29

While these items require further investigation at the programming level, several causes seem possible. One is that examinees paused on an item and closed the window without answering. Second is that there might potentially be a keystroke or action that allowed the examinee to move on, but did not log a response for the item. For example, the worst case of missingness occurred on the step which asked examinees to save a file to the desktop. It is possible that examinees attempted some kind of computer action on their own desktops as opposed to looking at the simulated computer desktop. An additional reason for

considering these hypotheses is that several of the other items that had many missing values were all four of the steps in the file management simulation. There could be some action that examinees are taking that test developers did not anticipate and thus did not account for. These items should be investigated for possible technology issues before operational use.

A final recommendation pertains to the use of scores from this assessment to determine mastery. Use of a single aggregated score would be misrepresentative of actual examinee knowledge. Different patterns of knowledge would be obscured by such a practice, given the varying levels of performance on the subsections and their relatively modest correlations with each other. The most informative approach would be to report the results of this inventory 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 devised using an accepted standard setting method (see Cizek, 2001) and should treat subsection scores individually.

Limitations

Professional standards for testing require certain actions on the part of test developers. For example, “Test developers have a responsibility to provide evidence regarding reliability and validity for stated testing purposes... They also need to anticipate how their tests will be used and misused, to do research that helps distinguish proper from improper uses, and to design tests and accompanying materials in ways that promote proper uses” (APA, 1985). This statement of test development standards makes 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 *Teacher Technology Literacy Inventory* was specifically developed for low-stakes test purposes; that is, the inventory is appropriate for uses such as identifying teachers’ professional development needs. However, since the test was developed for this test purpose, the inventory in its present form is *not* appropriate for use in high-stakes applications such as evaluating job performance or computing school grades.

If it becomes desirable in the future to use this assessment for high-stakes purposes, several additional steps should be undertaken. These steps include: (a) utilizing standardized

administration procedures; (b) developing multiple, parallel test forms; (c) developing an appropriate scoring model; and, (d) conducting additional data analyses. Each of these test development steps is discussed in greater detail next.

The current inventory has a very open test administration procedure, in fitting with its low-stakes test purpose. There are no constraints placed on where an examinee can take this Internet-based assessment, and very open scheduling is provided (e.g., for the field test, examinees could take the inventory in their own homes and they had two weeks during which they could complete the various subsections). In high-stakes applications, however, standardized procedures are necessary for fairness reasons as well as to ensure test security. High-stakes, standardized test administration procedures typically require a proctor to be present, monitoring all examinees and ensuring that the items are not printed, photographed, or otherwise exposed. Standardized time limits are usually required as well.

Another element of high-stakes testing is the provision of multiple, parallel test forms, all of which are designed to match the original test plan in terms of content, and all of which are assembled so that the item difficulty is highly similar across forms. As with the development of standardized procedures, this characteristic of high-stakes tests helps keep the test items secure and ensures fair conditions across examinees. The challenges of developing parallel forms extend to the need for comparable groups of field test examinees in terms of ability and other relevant examinee characteristics. High quality effort is needed at each stage of the test development process: item writing, item review, test form assembly and field testing, data analyses, and test form equating, as well as all other stages.

The inventory in its current state provides a simple, yet informative, approach to scoring. Each teacher is given his or her percent correct score on each of the six subsections of the inventory. If the inventory were to be used for higher-stakes purposes, it would be necessary to undertake an appropriate, formal standard setting process (see Cizek, 2001). Given the content and statistical differences across the six subsections of the inventory, this standard setting process should carefully consider appropriate means of standard setting that incorporate an examinee's profile of performance.

Furthermore, all of the data analyses conducted following the field test of the inventory would need to be conducted on each additional test form developed and on any new examinee group that utilized the assessment. Reliability estimates, validity estimates,

and item analyses such as those reported here would need to be conducted. Both internal consistency reliability and parallel forms reliability would be recommended. It might also be desirable to increase the reliability of the subsections; a primary means of doing this would be to increase the number of items in each subsection. In addition, it would be critical that the same level of quality item development process was followed; test reliability cannot be maintained or increased if additional items are of lower quality. Validity studies could go beyond those reported here. One recommended analysis is a Confirmatory Factor Analysis, as this could provide substantial additional content validity evidence and help in further defining the construct. Finally, greater attention to diversity should be incorporated into any future development processes. There should ideally be diverse representation in all stages of the test development process, from item writing through field testing and analysis. High-stakes exams should also include statistical analyses of potential item bias, such as Differential Item Functioning (DIF).

Conclusion

The *Teacher Technology Literacy Inventory* was developed well, and according to low-stakes test purposes as specified in the grant. Additional steps would be needed to use this inventory for high-stakes purposes, and are initially specified above. But for low-stakes applications, the *Teacher Technology Literacy Inventory* is clearly satisfactory. An extensive, thorough process was followed for defining the indicators, which were developed within the framework of the ISTE-NETS. The inventory items are well mapped to these indicators, and provide measurement of them in innovative, relevant, performance-based ways. Finally, test quality criteria show reasonable item analysis, reliability, and validity results for a relatively short CRT or mastery test. With few modifications, the inventory should be ready for a low-stakes implementation, and should become a useful tool for districts for reporting aggregate data for NCLB purposes and for helping teachers and districts to target technology-related professional development needs.

References

- AERA, APA, and NCME (1985). *Standards for educational and psychological testing*. Washington, D.C.: American Psychological Association.
- Barron, A. E., Kemker, K., Harmes, C., & Kalaydjian, K. (2003, Summer). Large-scale research study on technology in K-12 schools: Technology integration as it relates to the National Technology Standards. *Journal of Research on Technology in Education* 35(4), 489-507.
- Cizek, G. J. (Ed.). (2001). *Setting performance standards; Concepts, methods, and perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Florida Department of Education, Office of Educational Technology. (2003a). *Educator Accomplished Practice #12: Technology*. Retrieved December 30, 2004, from <http://www.doe.firn.edu/edtech/sr/downloads/EAP12New.pdf>
- Florida Department of Education, Office of Educational Technology. (2003b). *Request for Proposal. Enhancing Education Through Technology (EETT). Part II competitive opportunity*. Retrieved December 30, 2004, from <http://www.doe.firn.edu/edtech/it/eett/downloads/EETT-RFP-DOE900-PARTII.doc>
- Hogarty, K.Y., Lang, T. R., & Kromrey, J. D. (2003). Another look at technology use in classrooms: The development and validation of an instrument to measure teachers' perceptions. *Educational and Psychological Measurement* 63(1), 137-160.
- International Society for Technology in Education. (2002). *National Educational Technology Standards for Teachers: Preparing teachers to use technology*. Eugene, OR: Author.
- Thompson, B. (Ed.). (2003). *Score reliability: Contemporary thinking on reliability issues*. Thousand Oaks, CA: Sage.

Appendix A Performance Indicators

1. Basic Operations & Concepts

- Locate and open a file
- Adjust the volume on the computer
- Save a file to a specific location
- Save a file under a new name using "Save As"
- "Print preview" a document
- Print a file to a specific printer
- Locate and open applications
- Quit applications
- Create and name folders
- Organize files and folders
- Rename files and folders
- Check cables for proper attachment to computer and peripherals
- Recover from a software failure

2. Productivity

Word Processing

- Recognize a word processing program
- Create a new document
- Enter and format text (fonts, size, emphasis)
- Cut, copy, and paste text

Spreadsheets

- Recognize a spreadsheet program
- Identify elements of a spreadsheet
- Create a spreadsheet
- Enter data into a spreadsheet
- Edit data in an existing spreadsheet
- Create a chart (graph) with the data in a spreadsheet

Databases

- Recognize a database program
- Enter data into an existing database
- Edit data in an existing database

Graphics

- Recognize a graphics program
- Recognize graphics file formats
- Resize graphical objects in a document
- Identify tools of graphics software
- Add a graphic (i.e. pictures, photos or clip art) to a document

3. Communication

Email

- Compose and send an email message
- Reply to an email message (reply to sender; reply to all)
- Open an email attachment
- Attach a file to an email message
- Recognize an email program

Video

- Play a video clip on a computer

Presentations

- Recognize a slide show presentation
- Open and play an existing slide show presentation
- Add text to a slide within a presentation

4. Research

Research Skills

- Search online databases and reference software
- Recognize search strategies
- Identify appropriate search strategies
- Evaluate search results
- Evaluate the accuracy and credibility of information obtained through the Internet

Browser Use

- Recognize a browser
- Identify the elements of a browser
- Identify the elements of a web page
- Type a URL into the address bar
- Bookmark a site (using bookmark or Favorite)
- Recognize links on a web page
- Open a new browser window

5. Planning, Management, & Instruction

Planning and Management

- Identify basic uses of word processing programs
- Identify basic uses of spreadsheet programs
- Identify basic uses of database programs
- Identify basic uses of graphics programs
- Identify basic uses of communication tools

Instruction

- Identify basic instructional uses of word processing programs
- Identify basic instructional uses of spreadsheet programs
- Identify basic instructional uses of database programs
- Identify basic instructional uses of graphics programs
- Identify basic instructional uses of communication tools
- Identify basic instructional uses of concept mapping programs

6. Social, Ethical, Legal, and Human Issues

- Evaluate teacher use of technology in the classroom with regard to an Acceptable Use Policy for employees
- Evaluate student use of technology in the classroom with regard to an Acceptable Use Policy for students
- Judge appropriateness of posting student information on the Internet
- Identify safety and security issues with regard to technology and the Internet
- Identify fair use practices/copyright policy
- Properly cite digital resources

Appendix B
Sample Simulation Item

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

Task Progress

1. Create a new document.

Current Step

Complete this step in the screen to the right.

2. Type the following text and press ENTER:
FCITL is good. FCITL is great.

Omit this step

3. Select/highlight the word "good".
4. Change the style of "good" to italics.
5. Change the font of "good" to Courier.

WordTech 4 - Untitled

File Edit View Insert Format

Times New Roman 8 B / U

FCITL is good. FC|

Exit Help Sample Status

Appendix C Structured Interview Questions

Focus Group Structured Interview Questions

Now we want to ask you to give us your feedback and reactions to the parts of the inventory you've seen so far. Remember, we are interested in your negative feedback as well as any positive feedback. Your reactions will help us make the inventory better.

DIRECTIONS

1. From the Main Menu, was it clear how you were to proceed through the sections and how to select a section?
 - a. Did you have any difficulty reading the text on the screen?
 - b. Did the directions seem clear?
 - c. Did you have any difficulty knowing what to do on any screen?
2. Within each *section*, was it clear what you were supposed to do and how you were to respond to each question?
 - a. For the *multiple choice* items, was it clear what you were supposed to do and how you were supposed to respond to each question?
 - b. For the *simulation* items, was it clear how you were supposed to proceed through the steps and how you were supposed to respond to each question?

CONTENT

1. For each item, was the content of the item:
 - accurate
 - appropriate
 - unbiased (fair, sensitive)
 - at the right level of difficulty
2. Can you tell us about any questions that stood out because they were particularly: easy, difficult, interesting, confusing, or possibly flawed?
3. Do you have any comments about any wording choices, format or layout elements, or other aspects of any writing on screen?
4. Did the items and the tasks seem to be the right sorts of things we should be asking people to do, to measure basic technology literacy?
5. Did the items, and the overall inventory, seem to be about at the right difficulty level?
6. Did the items seem fair (unbiased, sensitive)?

OTHER

1. Do you think each section took a reasonable amount of time?
2. If you had to prepare for a test like this, what would you do?
How would you recommend a friend prepare for this kind of test?
3. Do you have any suggestions for features we could add that would help you in taking this kind of instrument?
4. Are there any other specific things you did or did not like about the inventory?

Appendix D Field Test Letter

October 29, 2004

Dear _____ :

Thank you for volunteering! Your district has been selected to participate in the field test of the Teacher Technology Literacy Inventory. The following information provides instructions and specifications regarding your participation in the field test.

Timeframe

The field test inventory will be available beginning November 8, 2004. Teachers will have 2 weeks to complete all six sections of the inventory along with a short survey regarding their use of and attitudes toward technology. Each teacher may access the inventory at any time during the 2-week window and may take each section separately (e.g., at a different time). Each section of the inventory is estimated to take about 10-20 minutes to complete. Along with registration and completion of the survey, participation in the field test will probably take a teacher about 1.5-2 hours. Once a teacher has completed the six sections of the inventory, he or she will have access to a score report that will provide the percent correct for the teacher in each of the sections. After the field test has been completed, you will be provided with results for all of the teachers from your district who participated.

Selecting Teachers

- Please select (**specific number**) teachers according to the following criteria:
 - Computer Experience
 - 33% should be Beginners
 - 33% should be Intermediate
 - 33% should be Advanced
 - Teaching Level - Where possible, please select teachers to represent:
 - Elementary School
 - Middle School
 - High School
 - Teaching Assignment
 - Teachers should represent various teaching assignments including (where possible):
 - General/Basic (including Elementary, Language Arts, Reading, Math, Science, Social Studies)
 - Vocational
 - Exceptional Student Education
 - Special Areas (including Music, Art, P.E)
 - Resource (including Media, Technology, Limited English Proficient, Guidance, Reading/Literacy Specialist)

Registering Teachers

To register your selected teachers, we will need the following information in an Excel spreadsheet: First name, Last Name, Email address, Computer Experience Level (Beginner, Intermediate, Advanced)

Please use the spreadsheet, which was attached to the same email as this document, as a template. Send this information to fltechlit@fcim.org. Please provide this information by November 4th, 2004. User lists received after the 4th will be accepted, however it will take 1-2 business days before they will be able to login to the inventory.

Computers

For all browser/operating system combinations, the Flash Player version 7.0 or later is required. The Flash Player is a free download, and is available at <http://www.macromedia.com/go/getflashplayer>

Please use the following compatibility matrix to make sure your teachers will have access to the configurations required to access the inventory. In this matrix, X denotes that the inventory can be used with that browser/operating system combination.

	Windows 98/Me/2000/XP	Mac OS 9	Mac OS 10
Internet Explorer 5 or later	X	Not compatible	X
Mozilla 1.0.2 or later	X	X	X
Netscape 6.0 or later	X	X	X
Opera 6.0	X	X	X
Opera 7.0 or later	X	NA	X
Safari 1.0 or later	NA	NA	X

All of the above browsers (with the exception of Opera) are available as free downloads on the Internet

Internet Explorer for Windows
Generally installed through windows update service
<http://windowsupdate.microsoft.com>

Internet Explorer for Mac
<http://www.microsoft.com/mac>

Netscape

<http://www.netscape.com/download>

Safari

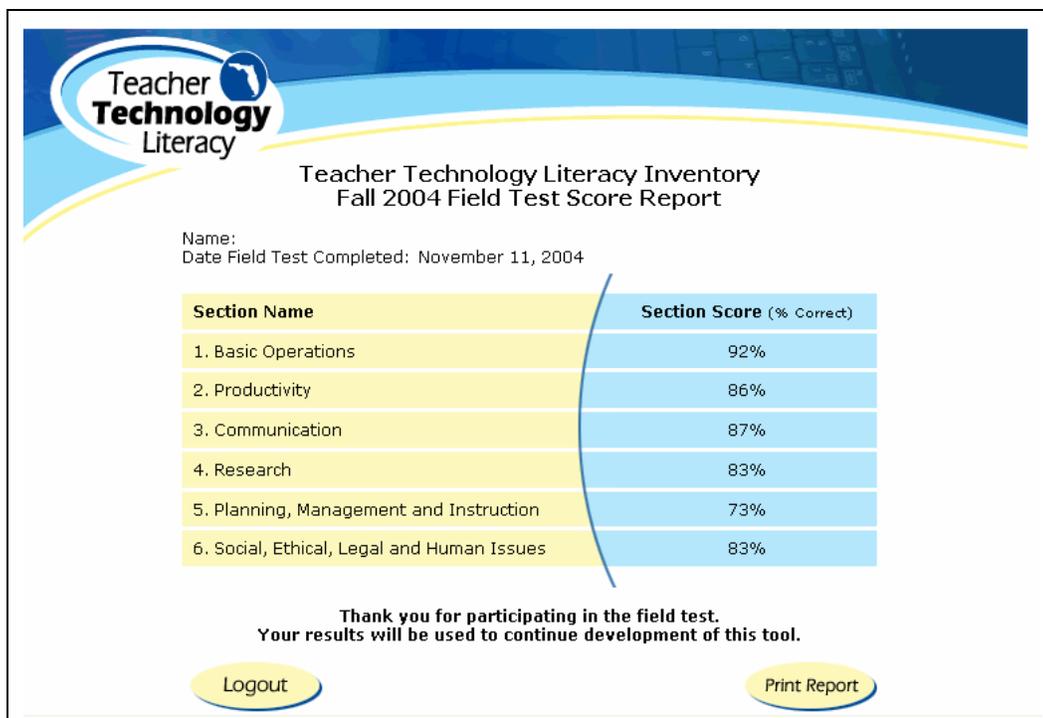
<http://www.apple.com/safari>

Questions?

Thank you for your interest in participating in the Fall 2004 Field Test. The results from the field test will be used to further the development of this tool. If you have any questions, please email: fltechlit@fcim.org or call 850-645-1759.

Appendix E Sample Score Reports

Online Score Report



Teacher Technology Literacy

Teacher Technology Literacy Inventory Fall 2004 Field Test Score Report

Name:
Date Field Test Completed: November 11, 2004

Section Name	Section Score (% Correct)
1. Basic Operations	92%
2. Productivity	86%
3. Communication	87%
4. Research	83%
5. Planning, Management and Instruction	73%
6. Social, Ethical, Legal and Human Issues	83%

**Thank you for participating in the field test.
Your results will be used to continue development of this tool.**

[Logout](#)
[Print Report](#)

Printable Score Report

Teacher Technology Literacy Inventory Fall 2004 Field Test Score Report

Name:
Date Field Test Completed: November 11, 2004

Section Name	Section Score (% Correct)
1. Basic Operations	92%
2. Productivity	84%
3. Communication	87%
4. Research	83%
5. Planning, Management and Instruction	73%
6. Social, Ethical, Legal and Human Issues	83%

*Note: These score reports are not from actual examinees; they are samples only

Appendix F Technology Uses and Attitudes Survey



**Teacher
Technology
Literacy**

Technology Uses and Attitudes

Page 1 of 5

Name of your school:



Gender: Male Female

Race/Ethnicity:

Native American /American Indian

Asian/Pacific Islander

African American

Hispanic

White/ Non-Hispanic

Other, please specify:

Highest degree earned:

Bachelors

Masters

Specialist (Ed.S)

Doctorate

Other, please specify:

What subject area(s) do you teach? (Check all that apply)

<input type="checkbox"/> Art / Music	<input type="checkbox"/> English	<input type="checkbox"/> Math
<input type="checkbox"/> Media / Technology specialist	<input type="checkbox"/> Physical Education	<input type="checkbox"/> Reading
<input type="checkbox"/> Science	<input type="checkbox"/> Social Studies	<input type="checkbox"/> Special Education
<input type="checkbox"/> Vocational Education	<input type="checkbox"/> Other, please specify: <input type="text"/>	

Total teaching experience in years:

What grade level(s) do you currently teach:

Average number of students per class:

Number of computers in your classroom used for instruction:

How many years have you been using computers in your classroom for instruction?

Do you have access to a computer lab? Yes No

If yes, how many hours each week do your students use the lab?



Technology Uses and Attitudes

Confidence and Comfort Using Computers

Directions: Please read the following statements and select the one response that best reflects your level of agreement.

Key
1=strongly disagree
2=disagree
3=neutral
4=agree
5=strongly agree

Key
1=strongly disagree
2=disagree
3=neutral
4=agree
5=strongly agree

1. I have had adequate training in using computers.	1	2	3	4	5
2. I use computers effectively in my classroom.	1	2	3	4	5
3. I am comfortable giving computer assignments to my students.	1	2	3	4	5
4. The computer enhances my teaching.	1	2	3	4	5
5. I am comfortable using computers during classroom instruction.	1	2	3	4	5
6. My use of computer technology enhances student performance.	1	2	3	4	5
7. My incorporation of multi-media into lessons enhances my teaching.	1	2	3	4	5
8. I am comfortable with computer terminology.	1	2	3	4	5
9. I am developing expertise in the uses of technology in the classroom.	1	2	3	4	5

Continue >

Technology Uses and Attitudes

Page 3 of 5

Attitudes Towards Computer Use

Directions: The following statements address general attitudes towards computer use. Please select the one answer that best reflects your level of agreement.

Key

1=strongly disagree
2=disagree
3=neutral
4=agree
5=strongly agree

Key

1=strongly disagree
2=disagree
3=neutral
4=agree
5=strongly agree

1. I would like every student in my classes to have access to a computer.	1	2	3	4	5
2. Computer skills are essential to my students.	1	2	3	4	5
3. I feel tense when people start talking about computers.	1	2	3	4	5
4. I feel pressure from others to integrate the computer more into my classroom.	1	2	3	4	5
5. I would like my students to be able to use the computer more.	1	2	3	4	5
6. Computers are dehumanizing.	1	2	3	4	5
7. I avoid the computer whenever possible.	1	2	3	4	5
8. Computer instruction is just another fad.	1	2	3	4	5
9. The use of computers should be confined to computer courses.	1	2	3	4	5
10. I like using the computer to solve complex problems.	1	2	3	4	5
11. More training would increase my use of the computer in the classroom.	1	2	3	4	5
12. Computers diminish my role as a teacher.	1	2	3	4	5
13. Computers should be incorporated into the classroom curriculum.	1	2	3	4	5
14. Computers make my job easier.	1	2	3	4	5
15. Computers further the gap between students along socio-economic lines.	1	2	3	4	5
16. Computer skills will help me as a professional.	1	2	3	4	5
17. Learning computers makes high demands on my professional time.	1	2	3	4	5
18. Computers change my role as a teacher.	1	2	3	4	5
19. I can help others solve computer problems.	1	2	3	4	5
20. Computers enhance classroom instruction.	1	2	3	4	5

[Continue >](#)



Technology Uses and Attitudes

Page 4 of 5

Teacher Confidence Level

Directions: The following were designed to capture your confidence and comfort with various aspects of technology. Using the response scale provided, please circle the one that best reflects your level of confidence.

Key

1=strongly disagree
2=disagree
3=neutral
4=agree
5=strongly agree

Key

1=strongly disagree
2=disagree
3=neutral
4=agree
5=strongly agree

1. Generate student grades using available software programs.	1	2	3	4	5
2. Create charts in a spreadsheet program.	1	2	3	4	5
3. Create a presentation on the computer.	1	2	3	4	5
4. Create a lesson or unit that incorporates technology.	1	2	3	4	5
5. Create newsletters using word processing.	1	2	3	4	5
6. Enter information in a database.	1	2	3	4	5
7. Create a concept map using a graphic organizer (e.g., Inspiration).	1	2	3	4	5
8. Use a scanner to place an image in a document.	1	2	3	4	5
9. Select a printer on the network.	1	2	3	4	5
10. Take a picture using digital camera.	1	2	3	4	5
11. Connect a digital camera to the computer to download the pictures.	1	2	3	4	5
12. Edit and manipulate digital photos.	1	2	3	4	5
13. Create a movie on the computer.	1	2	3	4	5
14. Access and retrieve information on the server.	1	2	3	4	5
15. Research and gather information from the Internet.	1	2	3	4	5
16. Design and create a web page.	1	2	3	4	5
17. Back up files on my hard drive.	1	2	3	4	5
18. Send an email with an attachment.	1	2	3	4	5
19. Install software on the computer.	1	2	3	4	5
20. Connect to the Internet wirelessly.	1	2	3	4	5
21. Recognize the unethical uses of technology.	1	2	3	4	5

Continue >

Technology Uses and Attitudes

Page 5 of 5

Integration of Computers Into the Classroom

Directions: Listed below are modes in which computers may be used. Indicate how often you and your students use computers in each mode.

Key

- 1=not at all
2=once a month or less
3=once a week
4=several times a week
5=every day

Key

- 1=not at all
2=once a month or less
3=once a week
4=several times a week
5=every day

1. My students use word processing as a tool for writing.	1	2	3	4	5
2. I use word processing for lesson plans, newsletters, and letters to parents.	1	2	3	4	5
3. My students use spreadsheets to present mathematical concepts.	1	2	3	4	5
4. I use spreadsheets to create charts and graphs for students.	1	2	3	4	5
5. My students use e-mail or chats to gather information from experts.	1	2	3	4	5
6. I use the Internet to participate in discussion groups for professional development.	1	2	3	4	5
7. My students create electronic books on the computer using a variety of media, e.g. text, graphics, audio or video.	1	2	3	4	5
8. I use digital images to create picture books for the students.	1	2	3	4	5
9. My students use a variety of media to create electronic portfolios.	1	2	3	4	5
10. I use the computer for classroom management, e.g. seating charts, attendance, and assignments.	1	2	3	4	5
11. My students create multimedia presentations using a variety of media.	1	2	3	4	5
12. I use the presentation station to present student's work.	1	2	3	4	5
13. My students use graphic organizers, e.g. Inspiration, to develop concept maps.	1	2	3	4	5
14. I use graphic organizers to present concepts to the class.	1	2	3	4	5
15. My students search for and evaluate information on the Internet.	1	2	3	4	5
16. I search for and evaluate information on the Internet when I prepare my lessons.	1	2	3	4	5
17. My students use authoring software to develop web pages.	1	2	3	4	5
18. I create web pages for instruction.	1	2	3	4	5
19. My students create desktop movies to communicate thoughts, ideas, and stories.	1	2	3	4	5
20. I use multimedia presentations to introduce learning activities to students.	1	2	3	4	5
21. Small group instruction	1	2	3	4	5
22. Individual instruction	1	2	3	4	5
23. Cooperative groups	1	2	3	4	5
24. As a reward	1	2	3	4	5
25. Independent learning	1	2	3	4	5
26. To tutor	1	2	3	4	5
27. To promote student centered learning	1	2	3	4	5
28. As a research tool for students	1	2	3	4	5
29. As a problem solving/decision making tool	1	2	3	4	5
30. As a productivity tool (to create charts, reports or other products)	1	2	3	4	5
31. As a classroom presentation tool	1	2	3	4	5
32. As a communication tool (e.g., email, electronic discussion)	1	2	3	4	5

[Finish >](#)