

# **FCAT Computer Accommodations Pilot Study Final Report**

Prepared by:

Nimble Assessment Systems, Inc. and the National Center for Educational Outcomes

November 11, 2008

## **Overview**

This report presents findings from a study that examined the use of NimbleTools to deliver the FCAT on computer with accommodations provided to students with disabilities and special needs. Commissioned by the state of Florida and funded through state legislative action, the study focused on the feasibility, usability, and effect of using NimbleTools to deliver multiple sections of a test composed of released FCAT items to students in grades 6 and 9.

As described more thoroughly below, NimbleTools is an innovative, state-of-the-art computer-based test delivery system. NimbleTools embraces principles of Universal Design such that several accessibility tools are built into the test delivery system. Based on each individual student's need, NimbleTools flexibly tailors the accessibility tools available for each student. Flexibly tailoring the availability of accessibility tools enables the test to be delivered to all students across the state using a single computer-based test delivery system with a common interface. Beyond assuring adequate access to appropriate and high-quality accessibility tools, the use of a universally designed test delivery system can significantly decrease the costs of student assessment by eliminating the need to develop, distribute, and monitor the use of multiple test forms in order to provide test accommodations to the approximately 12% of students who require them.

As described in detail below, the study was intended to examine the use of NimbleTools by 150 students from schools across Florida. In actuality, the study included 444 students, of which 385 completed both tests, from 21 schools located in southern Florida, Northeastern Florida, Central Florida, the Gulf Coast, and the Panhandle. Although the time frame for the study required the test to be administered the third week of school, school test coordinators reported no problems preparing to deliver the tests to their students. Test coordinators, however, did note that a longer time frame for the study would have allowed them more time to make informed decisions about which accommodations were most appropriate for each of their students. Some test coordinators also noted that the tight time frame for the study also

required the two tests to be administered back-to-back, causing fatigue for their students. Even though students had limited time to become accustomed to using NimbleTools prior to performing the tests, students overwhelmingly reported that they found NimbleTools easy to use and that they would prefer using NimbleTools to perform the FCAT in the future. Finally, although the study focused on just thirty-two test items, the findings indicate that the use of the accessibility tools built into NimbleTools had a differential effect on the performance of students who require test accommodations compared to students who did not require accommodations. Collectively, this study provides evidence that the use of NimbleTools to deliver the FCAT via computer, with or without accommodations, is feasible and effective.

The remainder of this report presents a detailed description of NimbleTools, the study design, the recruitment procedures, the sample of students who participated, findings from statistical analyses, and policy implications for the future of computer-based testing in Florida. Note that to avoid bias in conducting analyses, all inferential statistical analyses were conducted by researchers at the National Center for Educational Outcomes, a non-profit research center with a strong national reputation for examining and monitoring test accommodations.

## **Computer-Assisted Accommodations**

Computer-based testing holds great promise to increase the efficiency and utility of student testing, while also decreasing the cost of distributing, delivering, scoring, and returning results for state tests. A major problem confronting computer-based testing, however, is the need to provide test accommodations for students with disabilities and special needs. For paper-based tests, accessibility and accommodations are typically provided by: a) producing alternate forms of test materials (e.g., large print or Braille copies); or b) altering the test administration conditions (e.g., extending test administration time, allowing students to listen to soothing background sounds, or having a proctor work with a student to read parts or all of a test aloud and/or to read back written responses). While these same techniques can be used for a computer-based test, they do not fully capitalize on the power of computer-based technologies to adapt to individual needs and to standardized procedures across users. As an example, current read aloud accommodations provided by human proctors are problematic for several reasons, including:

- they are expensive for schools to provide
- read aloud accommodations are often provided to groups of students instead of individual students
- students are often reluctant to ask a human proctor to read a section of an item multiple times
- human readers do not always pronounce or read items correctly or the same across students
- unintentional and intentional clues as to the correct answer or a portion of a phrase that is important are provided
- mathematical functions, diagrams, and other non-text based materials are presented in an unstandardized manner

Each of these problems present threats to the validity of inferences about student achievement made based on test scores.

## Shortcomings of Current Computer-based Test Delivery Options

Some current computer-based testing systems are unable to directly support any accommodations. Other systems attempt to provide a few accommodations by allowing students to use commercial software that operates outside of or on top of the test delivery system. This approach, while less costly for the test developer, places a burden on schools to purchase and install commercial software and to then configure their computers to enable the software to work in conjunction with the test delivery software. In addition, the use of external commercial software to provide accommodations leads to unpredictable provision of accommodations. As an example, commercial text-to-speech software applications have difficulty presenting mathematical equations such as  $X^2 + Y^3 = 17$ . Many text-to-speech programs will present this item as "X two plus Y three equals seventeen." In other cases, words like "triangle" will be pronounced as "tri-aingal." While the difference in how some words are pronounced by commercial text-to-speech software is minor, these differences can have an adverse impact on a student's understanding of the question, which in turn presents a potential threat to test validity.

Similarly, the use of screen magnification tools that adjust the size of text presented on the screen can have two adverse effects. First, those tools that adjust font size also alter the layout of an item. For a reading test that references specific lines within a passage, enlarging font size and in turn altering the number of words on a line results in inaccurate references to passage lines. Second, magnifying text without also enlarging graphic elements that accompany that text, such as graphs and diagrams, does not provide students with low vision with adequate access to the test content. In both cases, these shortcomings present threats to test validity.

Another challenge that arises when external commercial software is used to provide accommodations occurs when a student requires multiple accommodations. Because commercial software is designed to work independent of other applications, conflicts can arise when multiple software applications are being run to provide more than one accommodation. As an example, the combined use of Microsoft's screen magnification tools and a commercial text-to-speech tool can lead to the reading of text that appears outside the viewable area of the screen. For some students, the reading of text that is not visible on the screen can cause confusion and poses a threat to test validity.

To decrease the burden placed on schools to provide commercial software, increase the standardization with which accommodations are provided, improve the quality with which accommodations are provided, and to ensure equity across all students in need of a given accommodation, principles of Universal Design must be applied when developing computer-based tests.

## Universal Design and Student Assessment

Spurred by the Americans with Disabilities Act, universal design is a direct response to design flaws in buildings – staircases, narrow entrances, escalators, high sinks, etc. – that made it difficult for people with physical disabilities to access buildings or use facilities within those buildings. Prior to the act, people with physical disabilities often required assistance from others to ascend staircases or maneuver through tight entry ways. In other cases, they were required to use a separate entrance, such as loading docks or doors located on the side or back of a building. In the worst cases, a person was removed from a wheel chair and carried into a building. Universal design overcomes these obstacles by purposefully designing buildings so

that they provide a choice of convenient and appropriate options when accessing a building or using specific facilities.

Applying the principal of universal design to a high-stakes test assures that all students have access to the tools they need to demonstrate what they actually know and can do. Just as it is no longer acceptable to design and build a structure that requires a person with a physical disability to use a doorway located in the back of the building, a high-stakes test should not require students with disabilities and special needs to take a test that is separate and distinct from that which is used by all other students. Just as it is unreasonable to require people with a physical disability to bring their own ramp to ascend a staircase, a testing program should not require students or their school to purchase special software in order to take a test. Instead, a universally designed test should be usable by all students, regardless of their disability or special need.

## **NimbleTools and Universal Design**

NimbleTools is a universally designed test delivery system that embeds several different accessibility and accommodation tools within a single system. The tools available to students during testing are flexibly adjusted based on the students needs. For students who do not require a test accommodation, NimbleTools delivers the test using a standard computer-based test delivery interface. For students who need a given accommodation or set of accommodations, a test proctor/teacher settings tool is used to customize the tools available for each student. As the student performs a test, s/he is able to use available tools as needed. This flexibility allows testing programs to customize the delivery interface to meet the specific needs of each student and for the student to then use specific tools as needed for each item on the test. In addition, NimbleTools collects information about the tools used for each item, which provides state testing programs with more accurate data on the use of accommodations during testing.

By integrating the accessibility and accommodation tools into a universally designed test delivery system, all students across the state use the same standard interface when performing a test, yet all students also have access to the same high quality tools and accommodations that are delivered in a controlled, standardized, equitable manner. This approach of building all accessibility and accommodation tools into the architecture of the test delivery system and presenting all students with a common, core interface upon which tools are layered contrasts sharply with other current computer-based test systems. These other systems employ completely different interfaces for students who require specific accommodations as compared to those students who do not require accommodations. The use of separate interfaces for different student groups increases the cost of computer-based testing, causes confusion for students and schools who must be trained on using different interfaces, introduces additional construct irrelevant variance to test scores, and results in a different test experience for a student who is assigned an accommodation prior to testing but opts not to use that accommodation, as compared to the student who is not assigned an accommodation.

As a final note, when applied to student assessment, universal design strives to provide students with the same experience despite the computer or operating system they are using. Unlike some systems that require a student who needs accommodations to work on a Windows-based system, NimbleTools was carefully designed to appear and function the same on Windows, Macintosh, and Linux operating systems. In addition, recognizing that the

technological capacity differs among schools, NimbleTools is designed to be used to deliver a test and accompanying accommodations in three modes, namely secure computer-assisted delivery, secure local delivery, and secure on-line delivery. While each mode varies whether the test resides on a local computer or an on-line server during delivery, and how responses are recorded, students' experiences are identical for all three delivery modes.

## **Brief Description of NimbleTools**

NimbleTools is a fully integrated universally designed computer-based test delivery system that includes several access and test accommodation tools. NimbleTools enables students to experiment with the specific accessibility and accommodation tools while performing practice tests. Following practice tests, teachers and special educators are able to activate one or more access and accommodation tool for a state test. The tools are set separately for each student for each state test (e.g., mathematics, reading, science). While taking a state test, the test delivery interface is customized for each student based on the tools that have been activated for him or her. Students then have the option of using a tool on an item-by-item basis. That is, a student can opt to use a given tool for one test item and decide not to use the tool for another item. As a student performs items, his/her answers are recorded and information about each of the tools used for that item is also recorded. When a student completes a test, his/her responses and all information about the use of the tool are recorded in a secure manner. It is important to note that NimbleTools will allow a student to use more than one access or accommodation tool at a time, depending upon their needs (e.g., magnification and read aloud or music calming and masking, etc.).

In its current form, NimbleTools includes the following accessibility and accommodation features:

- Read aloud of text with choice of human or digital reader (the student has the choice)
- Accessibility by Intellikeys keyboard with custom keyboard overlays
- Tab/Enter navigation of the interface (allowing keyboard or switch access)
- Presentation of signed text in American Sign Language or Pidgin Signed English
- Magnification of text and images for students with moderate visual impairments
- Magnification of text and images for students with low vision
- Masking of test items
- Masking of answers
- Auditory Calming
- Reverse contrast with selection of contrast color
- Color overlays with selection of overlay color
- Reading assistant with option of color overlay and/or magnification
- Read back of open-ended responses
- Alternate language text and read aloud (e.g., Spanish)
- Heritage language dictionary
- Talking Calculator
- Talking formula sheets and periodic table
- Extended Time

## **NimbleTools Computer-Assisted Accommodation Pilot Study**

As described above, NimbleTools is capable of providing a large number of test accommodations. For the FCAT Pilot Study, the following subset of accommodations were examined:

- Read Aloud of text and tags using high quality digital recordings of human voice
- Alternate contrast (changing background and font colors)
- Magnification of Test Items
- Auditory Calming

The Read Aloud Tool contained several unique features designed to meet the various needs of students who require assistance accessing text and other information contained in a test item. These features included:

- a) Automatic or select reading. Students who require a high level of reading support could opt to have items read automatically as the item appeared on the screen. Students could then click on any block of text to have it read aloud as many times as needed. Students who did not need all text read aloud could turn the automatic reader off and select only the blocks of text they wanted to be read as needed.
- b) Multiple Controls. Students were provided access to a read aloud player that functioned similarly to sound controls on a CD player. Students could also click directly on a block of text to have it read. Students could also tab through blocks of text and press the Enter key to have a block of text read aloud. Finally, students could stop reading at any time by pressing the “s” key on the keyboard or by click on a “Stop” button.
- c) Button Reading. Students who require a very high level of reading support could activate a button reader. The button reader provided descriptions of a given button that were designed to remind students of the buttons functionality.
- d) Tabbing. Students with low vision were provided the option of having verbal descriptions of their current location presented as they tab through the test taking environment.
- e) Graphic Descriptions. Students with low vision were provided the option of entering a graphic description mode. This mode built on the Tabbing option by automatically magnifying the interface, providing verbal descriptions of the student’s current location, and included verbal descriptions of graphic elements associated with a test item (e.g., picture, graph, table, etc.).
- f) Coordinated magnification and reading. For students who activated the magnifier and the read aloud tool, NimbleTools automatically placed the text being read in the upper left portion of the screen. This assures that the text being read to the student was also visible to the student.
- g) Coordinated reading and auditory calming. For students who activated the auditory calming tool and the read aloud tool, NimbleTools automatically decreased the volume of the auditory calming sound file when a block of text was read aloud. This assures that the two sound files do not compete for the student’s attention.

Magnification was provided to students in two forms:

- a) **Enlarger.** For students who only needed the content of a test item enlarges occasionally, the enlarger enabled students to magnify a test item up to 8 fold, but did not affect the magnification of the rest of the test environment (e.g., options panel and navigation buttons). The Enlarger also provided visual cues when information was moved off the screen due to the level of magnification. These cues are designed to help students access the entire test content.
- b) **Magnifier.** For students who needed the entire test environment enlarged, the Magnifier allows students to magnify everything presented on the screen up to 8 fold. When in magnified view, the student could adjust what was displayed on the screen using the mouse or the arrow keys. The level of magnification could be adjusted at any time using a slider or by using the "+" and "-" keys on the keyboard.

The auditory calming tool allowed students to select from among a set of sound loops and to have the loops played softly in the background as they performed the test. Students were provided with several music loops and three white noise tracks.

For students who benefit from alternate contrast, the Color Chooser allowed students to alter the background and foreground color. The Color Chooser presented students with a wide range of color combinations and allowed students to adjust their selection at any time.

It should be noted that all of these accessibility tools interacted with each other allowing a student to have an item displayed with alternate color contrast, magnified, read aloud, and auditory calming all functioning in coordinated manner.

In addition to these accessibility and accommodation tools, NimbleTools provided students participating in the pilot study access to:

- an embedded calculator that was able to read aloud when the read aloud tool was activated
- a formula sheet and periodic table of elements that read aloud formulas and text when the read aloud tool is activated
- direct input of responses using a mouse, tab-entering, touch screen, or switch mechanisms

The purpose of the pilot study was three fold:

1. To examine the feasibility of providing test accommodations to students with specific disabilities and special needs via computer.
2. To examine the usability of the computer assisted accommodations for students who are eligible for the accommodations.
3. To examine the efficacy of the computer assisted accommodations.

We describe the evaluation methods used to address each of these purposes separately below.

### **Examining Feasibility**

For the purposes of this research, feasibility was defined as the ease with which a participating school is able to implement and deliver a computer-assisted test using NimbleTools. To examine the feasibility of using NimbleTools to provide the test accommodations specified

above, to the extent possible the same distribution, installation, and administration procedures were followed in all schools for the efficacy study (described below) that would be used for an actual administration of the FCAT. These procedures include:

- A CD and a web site containing the NimbleTools Installer and instructions for installation were provided to each school participating in the efficacy studies.
- Within each participating school, one person was assigned as the Test Coordinator and was charged with installing NimbleTools on the computers to be used for the efficacy study.
- The Test Coordinator provided information about the names of the students who participated in the efficacy study.
- Teachers of the participating students or whomever the Test Coordinator determined was most appropriate for making decisions about the tools that should be enabled for each individual student and which tools should be blocked entered that information into a student preferences data base.
- At the appropriate time, students launched the Test Player, logged in to NimbleTools, and performed the assigned tests.

To evaluate the feasibility of employing NimbleTools for test administration of the FCAT, Test Coordinators were asked to complete a survey that focuses on the feasibility of using NimbleTools to provide accommodations for the FCAT.

### **Examining Usability**

For the purposes of this research, we define usability as the ease with which students are able to: a) use the tutorial to learn about NimbleTools' accessibility and accommodation tools; b) use the practice tests to become familiar with the test delivery interface and the accessibility and accommodation tools; and c) use the appropriate accessibility and accommodation tools while performing a test under simulated conditions.

To examine usability, two sources of data were collected. First, as described above, responses to specific questions in the Test Coordinators Survey that focus on the Test Coordinator's observations of the student use of NimbleTools were examined. Second, a student survey was administered to all students immediately following the completion of the last test included in the efficacy study (see below). The student survey contained 11 Likert scale items and two open-response items.

As noted above, each student was provided an opportunity to complete the survey immediately after they finished the second test used for the efficacy study. The survey was administered on computer and incorporated all appropriate accommodation tools included in the efficacy study (i.e., the survey was able to be read aloud, magnified, and/or presented with alternate contrast).

Together, data from the Test Coordinator survey and the student survey was used to evaluate the usability of NimbleTools for performing test items employed for the FCAT.

### **Examining Efficacy**

Given that this pilot study focused on the use of a computer-based test delivery system that is able to provide multiple accommodations for students with disabilities and special needs, efficacy was clearly defined in the context of the study's purpose. In educational research,

efficacy typically refers to the effect that an intervention has on student learning. An intervention is typically described as efficacious if its use leads to improved learning. In the context of this pilot study, however, the intervention (computer assisted accommodations) was not intended to have an effect on learning. Instead, the intervention was intended to:

- A) increase access to quality test accommodations;
- B) decrease the burden placed on the state and its schools for providing quality test accommodations; and
- C) increase the validity of inferences about student achievement based on test scores.

With respect to Goal C, Thurlow, McGrew, Tindal, Thompson, Ysseldyke and Elliott (2000) emphasize that before any test accommodation is deemed valid, it must be established that the accommodation or accessibility tool has a differential effect on students who are in need of the accommodation as compared to students who do not require the accommodation. Given time and budget constraints, the efficacy study focused on:

- The extent to which students who are eligible for a given accommodation indicate that they would use NimbleTools to take a future test
- The extent to which the use of NimbleTools has a differential effect on the performance of students eligible for the accommodation when NimbleTools accommodation tool is used versus when they are not used
- The extent to which the use of NimbleTools improves attitudes toward taking tests for students with disabilities and special needs who are eligible for NimbleTools provided accommodations

## Study Design

The three efficacy questions were examined separately for Grade 5 mathematics and Grade 8 science. Since the study was conducted during the first weeks of the school year, students who participated in the study were in Grade 6 and Grade 9. The decision to focus participation in these grade levels was made in order to assure that students had an adequate opportunity to be exposed to the content covered by each test.

For each grade level, a repeated measures design was employed in which students performed two sets of items matched by content and item. Each test contained 17 released FCAT items. The test forms were matched by content. In addition, the two forms were assembled so that the overall difficulties were within .02 for mathematics and .01 for science.

For both the mathematics and science studies, the same test administration protocol was employed. This protocol entailed eight steps:

1. Students performed a short tutorial that demonstrated how to use the standard NimbleTools interface.
2. Students performed a 3 item practice test to become accustomed to using NimbleTools.
3. Students performed Form 1 without access to the NimbleTools accessibility tools. Form 1 contained 17 FCAT items.
4. Students took a break. Breaks ranged from a few minutes to 2 days.

5. Students performed a short tutorial that demonstrated how to use the NimbleTools read aloud, magnification, auditory calming, and color contrast tools.
6. Students performed a 3 item practice test to become familiar with using the NimbleTools accessibility tools.
7. Students performed Form 2. Accommodated students had access to the accessibility tools assigned by their teacher. General Education students had access to all accessibility tools. Form 2 contained 17 different FCAT items.
8. Students completed a 13 item survey. Accessibility tools were available when completing the survey.

As described below, the original sampling plan included only those students who were eligible for one or more of the accommodations that were the focus of this study. When recruiting schools to participate in the study, however, several schools expressed a desire to have some of their general education students participate in the study. To better enable the examination of a differential effect, the study allowed general education students to participate. Like students eligible for the targeted accommodations, general education students performed the first item set using NimbleTools without any accessibility tools activated. For the second item set, the read aloud, magnification, color contrast, and auditory calming tools were made available to the general education students.

### **Alternate Study Design**

It is important to note that the design employed for this study required all students to first perform 17 items without access to any accommodation or accessibility tools. Students were then supposed to have 1–3 days before performing a second set of 17 items with access to accommodation and accessibility tools. For students who were identified by their teacher or Test Coordinator as needing specific accommodations (Accommodated Students), only those accessibility tools that were aligned with their need were made available. For students whose teachers indicated that no accommodations were required (Non-Accommodated Students), all tools (except read aloud for low vision) were made available. This design was adopted to examine the differential effect hypothesis that maintained that the tools would benefit students in need of the tool and would not have a smaller, no, or even negative effect on students who did not need the tool.

While this study design was practical given the limited time frame for this study, it has two major flaws. First, because all students performed the same 17 items first and then a different 17 items second, differences in the difficulty of the two item sets and the effects of fatigue were not controlled. To limit the effects of differences in item difficulty, the two sets of 17 items were selected so that the percent of students responding correctly to the items when administered to all students across Florida were very similar. To limit the effects of fatigue, teachers were asked to wait 1-3 days before administering the second set of items. Due to time constraints and limited availability of computer labs, many teachers were unable to provide this break and instead administered the tests back-to-back. Second, because all non-accommodated students were provided access to the accommodation tools for the second test administration, it is not possible to separate the effect that the tools had on student performance from other factors such as fatigue or test difficulty. Thus, any change in scores for the non-accommodated students may be due to differences in test difficulty, fatigue, the accommodation tools, or any combination of these factors.

To overcome these shortcomings in the study design, an alternate design was proposed and briefly considered prior to the study (Table 1). In this alternate design, half of the students would first perform one 17 item set without accommodations while the remaining students would first perform the other 17 item set without accommodations. Next, students for whom their teacher identified specific accommodations would perform the other 17 item set with the assigned accommodations. Students who did not require accommodations would be randomly divided into two sub-groups. The first sub-group would perform the second 17 item set without any accommodations. The second sub-group would perform the second 17 item set with accommodations. This design would have enabled analyses to separate the effects of fatigue, test difficulty, and accommodations. Due to the complexity of the design and the limited time frame for collecting data necessary to assign students to conditions, the research team was advised not to employ this design.

**Table 1: Alternate Counter-Balanced Study Design**

	Test Administration 1	Test Administration 2
<b>Accommodated Students</b>		
Group 1	Item Set 1 No Accommodations	Item Set 2 Accommodations
Group 2	Item Set 2 No Accommodations	Item Set 1 Accommodations
<b>½ Non-Accommodated Students</b>		
Group 1	Item Set 1 No Accommodations	Item Set 2 No Accommodations
Group 2	Item Set 2 No Accommodations	Item Set 1 No Accommodations
<b>½ Non-Accommodated Students</b>		
Group 1	Item Set 1 No Accommodations	Item Set 2 All Accommodations
Group 2	Item Set 2 No Accommodations	Item Set 1 All Accommodations

### Sampling

The sampling procedures employed for this study differed from those originally proposed. The original plan was as follows:

1. A sub-set of districts and schools serving students in sixth or ninth grade within the districts will be selected by the Florida Department of Education.
2. Selected districts will provide a list of students who meet the targeted disabilities within the selected schools.
3. The list of students provided by the district will be merged with State 2008 FCAT data and students who used the targeted accommodations during the 2008 test will be identified. These students will be eligible for the efficacy.
4. Eligible students within the selected schools will be selected to participate in the efficacy study. When selecting students, the number of students selected will be proportional to the number of students who used a given accommodation and who

had a given primary disability designation. As an example, if 20% of all students eligible to participate in the study had a Presentation accommodation and had a SLD primary designation, then 20% of the sample will be comprised of students who used a Presentation accommodation and designated SLD.

5. Selected schools will be asked to indicate whether the selected student actually entered the school.
6. From the list of eligible students who are attending the selected school, a sample of 180 students will be selected to participate in the efficacy study. In selecting 180 students, we anticipate that up to 30 students will decline or will not be able to participate in the efficacy study.

Due to limited time and resources, the Department of Education did not have access to sufficient data to complete the first step of the original sampling plan. Instead, Nimble worked with the department to contact district and school leaders to solicit schools that would be willing to participate in the study. Specifically, the following activities were performed to recruit participating schools.

First, in May 2008 the Florida Department of Education distributed information about the study to a set of schools/districts. Four schools responded and indicated interest in having their students participate in the study.

Second, between May and August 2008, Nimble Assessment Systems solicited participation of schools by:

- 1) Calling ESE district leaders
- 2) Emailing principals
- 3) Calling principals in the following districts:  
Miami/Dade County, Broward County, Orange County, Seminole County, Hillsborough County, Sarasota County, Duval County, Brevard County, Okeechobee County, Gulf County, and Santa Rosa County.
- 4) Obtaining permission from district technology personnel to install NimbleTools on school computers in Okeechobee, Santa Rosa, and Duval Counties.

The following schools agreed to participate in the study:

**6th Grade Mathematics Study**

- Osceola Middle School, Okeechobee County
- Yearling Middle School, Okeechobee County
- Woodlawn Beach Middle School, Santa Rosa County
- Gulf Breeze Middle School, Santa Rosa County
- Jackson Middle School, Orange County
- Port St. Joe Middle School, Gulf County
- Wewahitchka Middle School, Gulf County
- Andover Middle School, Dade County
- Carol City Middle School, Dade County
- Enterprise Elementary School, Brevard County

Mayport Middle School, Duval County  
 Lavilla Middle School for Arts, Duval County

### 9th Grade Science Study

Rosenwald School, Seminole County  
 Cyprus Creek High School, Orange County  
 Port St. Joe High School, Gulf County  
 Wewahitchka High School, Gulf County  
 Blake High School, Hillsborough County  
 Venice High School, Sarasota County  
 Ed White High School, Duval County  
 Cocoa High School, Brevard County

As of September 15, 2008, 118 students had completed both of the grade 5 mathematics tests and 204 students had completed both grade 8 science tests. Of these students, 103 grade 6 students and 163 grade 9 students opted to complete the post-test survey. It should be noted that an additional 45 grade 6 students and 11 grade 9 students completed the tests after the cut off date. Due to time constraints, data for these students are not included in the analyses reported here.

## Test Instruments

The tests used for the efficacy study each contained 17 items. Table 2 displays the composition of each test form. Specifically, the grade 5 mathematics tests primarily contained multiple-choice items, a small number of short answer/grid-in items, and one open-response item. The grade 8 science tests contained 16 multiple-choice items and one open-response item. It should be noted that the open-response items were included only to demonstrate that NimbleTools could be used to present and record student responses to open-responses items. The open-response items, however, were not scored and are not reported here. Thus, the data used for the analyses presented below is based on student responses to the multiple-choice and grid-in items, only.

**Table 2: Composition of Test Forms**

	G5 Math Form 1	G5 Math Form 2	G8 Science Form 1	G8 Science Form 2
Multiple-choice	12	11	15	15
Short-answer/Grid	4	5	1	1
Open-Response	1	1	1	1

For each item, a read aloud script was developed. The scripts prescribed how the test content, formulas, periodic table, and description of images (i.e., tables, graphs, charts, and figures) were to be read aloud. All scripts were reviewed and approved by content experts within the Department of Education. Once scripts were approved, voice recordings were created. All items, voice recordings, and associated graphic files were placed into NimbleTools and were

prepared for delivery. During this process, a beta version of each test was shared with and approved by the Department of Education. Final modifications were made prior to use for this pilot study.

## **Overview of Analyses**

Three types of analyses were conducted. To examine the feasibility and usability of a universally designed computer-based test system that can flexibility adjust to provide test accommodations for students with disabilities and special needs, survey data from students, test coordinators, and test administrators was analyzed. Survey analyses examined the distribution of responses to Likert-scale items and patterns in responses to open-ended items.

To examine the efficacy of using a universally designed computer-based test system that can flexibility adjust to provide test accommodations, descriptive statistics and inferential statistical tests were conducted. Finally, to examine whether the use of a universally designed computer-based test system that can flexibility adjust to provide test accommodations had a differential effect on the performance of Accommodation and Non-Accommodation students, a time series visual analyses is presented.

## **Student Survey Results**

Tables 3 and 4 display summary data for each survey item. Overall, students responded positively to using NimbleTools to perform a simulated FCAT test. Eight-seven percent of sixth graders and 90% of ninth graders agreed or strongly agreed that it was easy to perform the test on computer after using NimbleTools. After using NimbleTools, 78% of sixth graders and 88% of ninth graders stated that they liked taking a test on computer. The vast majority of students also reported that the NimbleTools tutorial was easy to use and that the practice tests helped them learn how to take a test on computer. Of those students who used the Read Aloud and Magnification tools, the vast majority reported that they were easy to use and that they were helpful during the test. Finally, and perhaps most importantly, after using NimbleTools, 82% of sixth graders and 85% of ninth graders agreed or strongly agreed that they wanted to take tests in the future using a computer. Overall, student responses provide evidence that NimbleTools was highly usable and was perceived by students to be an effective computer-based test delivery interface that provided them with accessibility tools that were easy to use and helpful during testing.

**Table 3: Summary Statistics for Grade 6 Student Survey**

n = 103	Does Not Apply	Strongly Disagree	Disagree	Agree	Strongly Agree
It was easy to perform the test on a computer.		.05	.08	.62	.25
The tutorial was easy to use.		.01	.13	.57	.30
After using the tutorial, I still wasn't sure how to take a test on computer.		.34	.41	.17	.09
The practice tests helped me learn how to take a test on computer.		.05	.13	.52	.29
The practice test helped me learn how to use the read aloud, magnification, and other tools.		.07	.13	.56	.24
The read aloud tools were easy to use.	.19	.03	.06	.57	.34
The read aloud tools were helpful during the test.	.13	.06	.12	.49	.33
The magnification tools were easy to use.	.22	.09	.10	.52	.29
The magnification tools were helpful during the test.	.20	.03	.18	.54	.25
I liked taking the test on computer.		.09	.13	.40	.38
I want to use a computer to take tests in the future.		.08	.10	.39	.43

**Table 4: Summary Statistics for Grade 9 Student Survey**

n = 163	Does Not Apply	Strongly Disagree	Disagree	Agree	Strongly Agree
It was easy to perform the test on a computer.		.04	.06	.59	.31
The tutorial was easy to use.		.09	.12	.50	.28
After using the tutorial, I still wasn't sure how to take a test on computer.		.39	.35	.20	.06
The practice tests helped me learn how to take a test on computer.		.12	.15	.52	.22
The practice test helped me learn how to use the read aloud, magnification, and other tools.		.07	.17	.50	.25
The read aloud tools were easy to use.	.25	.08	.04	.56	.33
The read aloud tools were helpful during the test.	.26	.11	.12	.51	.26
The magnification tools were easy to use.	.29	.07	.08	.61	.24
The magnification tools were helpful during the test.	.33	.06	.19	.52	.21
I liked taking the test on computer.		.05	.08	.51	.37
I want to use a computer to take tests in the future.		.04	.11	.47	.38

In addition to the eleven Likert-scale items presented in Tables 3 and 4, students were also asked two open-ended items. The first open-ended item asked students to report any problems they experienced while taking a test with NimbleTools. Table 5 displays a summary of issues students raised in response to this item, along with the number of students who mentioned the issue. Clearly, the vast majority of students experienced no problems. In fact, of the students who completed the survey, 124 stated that they did not experience any problems. The next two most frequent problems related to the content of the test being too difficult (which is reflected in the low scores, particularly for the math test) and the tutorial being too long or moving too slowly. Note that in order to assure that all students were exposed to the same information about taking a test with NimbleTools, the tutorial did not allow students to move at their own pace, but instead required students to view the same "movie." Clearly, some students who may have been familiar taking a test wanted to move through the tutorial at a quicker pace.

The remaining issues related to specific elements of NimbleTools or to hardware problems. Four students noted that their computer froze. However, since these students had not indicated that they had completed the test prior to the computer freezing, these students were able to log back in and continue performing their test. Six students reported problems using the color chooser. Four students reported problems using the music (auditory calming).

Interestingly, two of these students were not provided access to the auditory calming tool, yet they must have become aware of the feature and wanted to make use of it during testing. Three students reported difficulty using the read aloud tool, while three others reported that did not like having portions of the test read to them and it took them a while to realize that they could turn it off. Finally, two students reported having problems with the magnifier.

**Table 5: Problems Students Reported Experiencing While Using NimbleTools**

	Number of Students
No problems	124
Items were difficult	13
The tutorial was too long/slow	13
Had difficulty using the color chooser	6
Computer froze or logged them out	4
Could not find music	4
Problems using read aloud	3
Did not want read aloud/was bothered by it	3
Wanted to use scratch paper	2
Problems with the magnifier	2

Table 6 presents a summary of student responses to an open-ended item that asked them to describe any suggestions they have for improving NimbleTools. The most common response was “no suggestions.” Thirty-two students indicated that they would like to have more or better music options. Thirteen students indicated that they would like to either skip the tutorial or be able to move through it more quickly. Seven students indicated that they would like a choice of voices for the read aloud. It is important to note that NimbleTools does provide a choice of voices, but that voice option was not included as a feature in this study. Four students requested a wider array of color choices. Again, NimbleTools has the capacity to allow students to select from a full array of colors, but the number of color options was restricted to 16 for this study. One student also asked for access to an English/Spanish dictionary. NimbleTools contains a heritage language translator that was not included as a feature for this study. One student requested an easier way to find the tools and another asked that the practice items be eliminated. Finally, one student asked for access to a ruler. None of the items included in the study required students to use a ruler, but NimbleTools can be equipped with a digital ruler that can be manipulated with a mouse, with Intellikeys, or with tab-enter devices.

**Table 6: Suggestions For Improving NimbleTools**

	Number of Students
No suggestions	90
More/better music choices	32
Be able to skip or speed up tutorial	13
Change read aloud voice	7
More/different color choices	4
Easier content	3
Access to ruler	1
Access to English/Spanish dictionary	1
Fix problem with color changer	1
Make it easier to find tools?	1
Remove practice items	1

### Test Coordinator and Proctor Survey Results

Table 7 displays summary data for the Likert scale survey items presented to either the test coordinator or test proctor within each participating school. As of September 22, 2008, survey responses were submitted by 11 Test Coordinators and Proctors. Overall, the coordinators and proctors responded positively to their experiences installing and proctoring students as they used NimbleTools to perform a simulated FCAT test. Ninety one percent of the coordinators and proctors agreed that it was easy to prepare computers for students to use NimbleTools for testing. One hundred percent of respondents indicated that Nimble Assessment Systems support team was available and helpful in supporting the FCAT pilot study. Ninety percent of coordinators and proctors agreed that students responded favorably using NimbleTools and 91 percent felt NimbleTools helped them access the test content. Ninety percent of the coordinators also agreed that they wanted their students to use NimbleTools to perform tests in the future.

**Table 7: Summary Statistics for the Test Coordinator and Proctor Survey**

n = 11	Strongly Disagree	Disagree	Agree	Strongly Agree
It was easy to prepare computers for students to use for testing.	.00	.09	.55	.36
Nimble Assessment contacts were available and helpful in preparing and supporting the research.	.00	.00	.36	.64
Many students responded favorably to taking the test using NimbleTools.	.00	.10	.40	.50
NimbleTools helped students appropriately access the test content.	.00	.09	.455	.455
I would like for my students to use NimbleTools in the future.	.00	.10	.30	.60

In addition to the Likert scale items reported in Table 7, the Test Coordinators and Proctors were asked to identify any challenges they or their students encountered and to make suggestions for improving the process of using NimbleTools to deliver a test with accommodations. The three major challenges test coordinators and proctors reported included:

1. Requiring more time to prepare for the pilot study.
2. Students feeling frustrated by the slow pace of the tutorials.
3. Students experiencing technical problems with their computers during testing.

Regarding the first issue, it should be noted that the study was conducted under a very tight time frame during the first three weeks of school. As one person noted, "Turn around time even with out the storm days was too quick. Asking teachers to respond about accommodations is a thoughtful and lengthy process. Also at the start of the school year teachers don't know students well enough to make best case decisions about accommodations. It should either be based on the accommodations on the student's IEP or later in the year so that teachers can give more relevant input." Unfortunately, due to the timing of funding and budget cycles, this was the only time of the year when the study could be conducted. In the future, however, care should be taken to implement studies such as this over a longer time period and during a different period of the school year.

The second issue was the result of a carefully considered decision to have all students view two 3-minute tutorials prior to performing practice tests and then the actual pilot test. To assure that all students were presented with the same information about how to use NimbleTools, a fixed paced tutorial was employed. In the best case scenario, students would have time prior to the test administration to become familiar with a computer-based test interface and would be able to practice using that interface during class time. The short window for collecting data for this study, however, required that the tutorial, practice items, and actual pilot test be delivered during a single block of time.

The third issue was the result of two components. In some schools, students experienced problems with their computers that were unrelated to NimbleTools. Students then had to restart their computers and log back into the test. In one school, problems arose because the

Test Administration Instructions had not been passed on to the test proctor by the school's Test Coordinator. As a result, the proctor was not aware of the procedures for logging students into NimbleTools. Lacking instructions, the proctor and the students had to make informed guesses about the log in procedures. Although they were able to do so, the lack of instructions created tension that would have been avoid if the proctor had been provided the proper information about the login procedures. It should be noted that it was this proctor who tended to rate their overall experience participating in the study lower than respondents from other schools.

Beyond these three issues, the experiences and responses by test coordinators and proctors were overwhelmingly positive. One person stated, "Couldn't have been easier, the teacher felt unnecessary." Another wrote, "Serving as the test administrator is quite simple." A fourth wrote, "My role as the test administrator was simple... Once I received the students' login information, I was able to schedule the students to come in and complete the pilot testing in the computer lab. During test administration, I did not experience any problems." Finally, another person stated, "Make it available to all students, ESE, ESOL."

## Test Results

Results of statistical analyses are presented separately for the Grade 5 mathematics and Grade 8 science tests. Recall that since data was collected during the third week of school, each test was administered to students in the next grade level (Grade 6 and Grade 9).

### Grade 6 Students Performing Grade 5 Mathematics Test

As of September 15, 2008, 232 students had completed Form 1 and 181 completed Form 2. Since the analyses focused on the difference in performance when accommodation and accessibility tools were not available and when they were, those students who did not complete Form 2 were removed from the analyses. Table 8 shows that out of the 181 students who completed both forms, 146 were classified as Accommodation and 35 were classified as Non-Accommodation.

**Table 8: Classification of Participation Students**

	N	Percent
Accommodation	146	80.7
Non-Accommodation	35	19.3
Total	181	100

Table 9 displays the distribution of students across the participating schools. In general, students were fairly well distributed among the participating schools, with only one of the twelve schools contributing more than 10% of students. This one school, however, contributed 30% of the students and contributed 28 of the 35 non-accommodated students.

**Table 9: Distribution of Students Across Participating Schools**

	Percent
School A	5.2
School B	4.7
School C	2.6
School D	1.2
School E	3.9
School F	7.8
School G	8.6
School H	6.5
School I	3.9
School J	6.5
School K	5.2
School L	3.0
School M	30.2

Table 10 displays the percent of Accommodation students who had each of the accessibility tools activated for Form 2. Table 10 also displays the number of Accommodation students who had one, two, three, four, or five accessibility tools activated. The vast majority of students were provided access to the read aloud tool. Approximately half of the students were also provided access to the auditory calming and the color contrast tools. The majority of students also were provided to multiple tools.

**Table 10: Percent of Accommodation Students with Each Tool Activated for Form 2**

Tools	Percent
Standard Read Aloud	92.5
Low Vision Read Aloud	5.5
Magnification	8.2
Auditory Calming	47.9
Color Contrast	42.5
One Tool	39.0
Two Tools	34.2
Three Tools	20.5
Four Tools	3.4
Five Tools	2.7

Table 11 displays the mean score for Form 1 and Form 2 for the total group. As seen in Table 11, students performed slightly better when accommodation tools were activated (mean = .349) compared to when the tools were unavailable (mean = .344). Table 12 presents results of paired sample t-tests, which examine the difference in test scores for Form 1 and Form 2 for the total group of students. Preliminary results using paired samples t-tests indicate that there was not a significant difference between the student performance when accommodation tools were unavailable and when they were available.

**Table 11: Paired Samples Statistics, Grade 6 Math**

	Mean	N	Std. Deviation	Std. Error Mean
<b>Form 1</b>	.349	181	.16	.01
<b>Form 2</b>	.344	181	.16	.01

**Table 12: Paired Samples Test, Grade 6 Math**

Mean Difference	Standard Deviation	Standard Error Mean	T	Significance (2-tailed)
.005	.21	.02	.34	.73

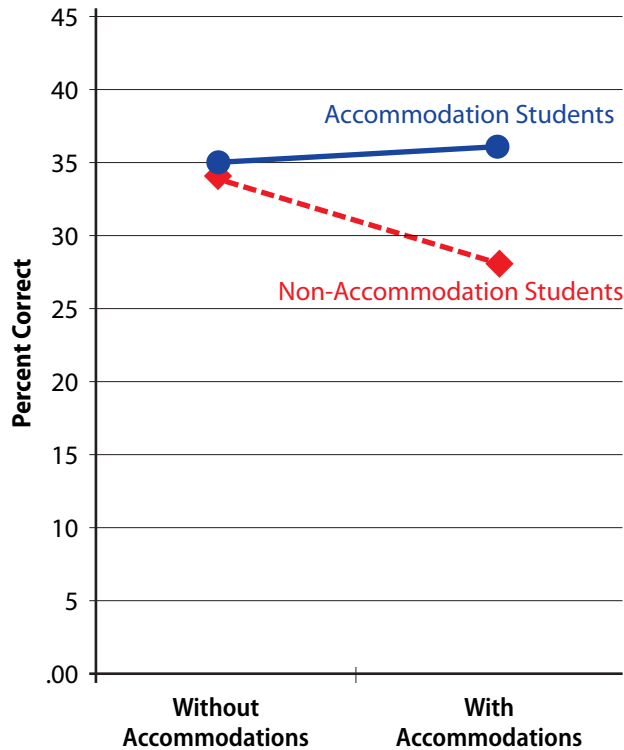
Paired sample t-tests were conducted for the Accommodation and the Non-Accommodated groups to determine if there were differences in test forms for these two groups of student. As seen in Table 13, the Accommodation students performed slightly better when accommodation tools were provided than when they were not. Conversely, the Non-Accommodation students performed approximately six percentage points worse when accommodation tools were provided. Neither of these score differences, however, was statistically significant.

**Table 13: Paired Sample t-Tests**

Group	Form 1 Mean	Form 2 Mean	Dif.	SD	Standard Error	t	Sig. (2 tailed)
Accommodation	.35	.36	.01	.21	.02	.40	.69
Non-Accommodation	.34	.28	.06	.21	.03	1.64	.11

Although statistical significance was not present, there was a notable difference in mean scores by population. As shown in Figure 1, a sample of general education students took this test and fared worse when accommodation tools were provided for them compared to when the tools were not available (a 6% decrease in correctly answered items). Students with disabilities and English language learners, however, had a 1.0% increase in correct items when tested with the accommodation tools available.

**Figure 1: Mean Scores of Accommodation and Non-Accommodation Students**



Next, we sought to determine if there would be a significant difference in scores by built-in accommodations (tools) selected for the assessment population (teachers assigned tools to students based on their need for the second test). Data for read-aloud, magnification, auditory calming, color contrast, and GDX were analyzed for the entire sample. No significant difference in achievement results were found between test forms for read aloud, auditory calming, color contrast, or GDX, but the use of magnification tool was found to have a statistically significant difference in assessment scores ( $p$ -value = .01). We then analyzed the same data, but removed Non-Accommodation students to determine if any particular accommodation had an effect on the achievement of the Accommodation group. We again found that magnification had a significant impact on scores ( $p$ -value = .02) for students with disabilities and English language learners. Because the effect size for the Accommodation group was smaller than it was for the entire population, we can assume that the magnification tool had a positive effect on all students who used it, regardless of disability or English language learner status. This does not represent an accommodation effect (a differential boost for the SPEDELL group), but does provide evidence that magnification may be a tool that may be useful in improving validity for all students.

A second noteworthy result of these data was the built-in accommodations (called “tools” on this assessment) that were assigned to students. Among all the students who took the Grade 6 mathematics assessment, 92.5% were assigned the read aloud tools, 8.2% magnification, 47.9% auditory calming, 42.5% color contrast, and 5.0% GDX (read aloud for blind/low vision students). Students in the Accommodation group were assigned specific tools to use by their teachers, while the general education students were assigned all tools (except GDX) and were allowed to choose their own tools on the assessment. A majority of students used more

than one tool on the accommodated form (Form 2). In order to gauge whether use of multiple accommodations had an impact on student achievement, we examined if there was a statistically significant correlation between the number of tools a student used and their score difference between Form 1 (without access to tools) and Form 2 (with access to tools). We did not find a positive correlation between using more tools and higher scores on Form 2. This result is likely explained by the ineffectual role that some technology tools have on student achievement (e.g., some students may prefer different colored font, but it will not help them to score better on an assessment).

Our final set of analyses attempted to link perceptions of use of assessment tools with improved performance on the technology-enhanced assessment. Our hypothesis for these analyses was that if students feel comfortable with technology tools, the tools will have an impact on their achievement. To a limited extent, our hypotheses were correct. For the overall population, we found a positive correlation between student feedback on the survey question that asked “I like taking the test on the computer” (Pearson Correlation = .352; two-tailed p-value = .048). Liking technology-based tests correlated positively to improved achievement for both the total sample and students with disabilities and English language learners.

Table 14 provides correlation information for selected survey questions. Only information for the total sample is included here, because there were no statistically significant correlations for the Accommodation group. There was a second statistically significant correlation for the total population. A negative correlation existed between improved achievement and responses to the item “After using the tutorial, I still wasn’t sure how to take a test on the computer.” This negative correlation can be interpreted as a positive relationship between tutorial use and improved achievement – students who felt more comfortable after completing the tutorial tended to perform better on the test when accommodation tools were made available.

**Table 14: Correlations Between Survey Items and Score Change**

Question	Pearson Correlation	Sig. (2-tailed)
It was easy to perform the test on a computer.	0.07	0.70
After using the tutorial, I still wasn’t sure how to take a test on computer.	-.35*	0.05
The practice test helped me learn how to take a test on computer.	0.13	0.49
The practice test helped me learn how to use the read aloud, magnification, and other tools.	-0.14	0.45
I liked taking the test on computer.	.35*	0.05
I want to use a computer to take tests in the future.	0.22	0.23

\* Correlation is significant at the 0.05 level (2-tailed).

## Limitations of Grade 6 Mathematics Test Study

Four notable limitations were present in the Grade 6 Mathematics analyses. First, the total N in the data file is 232, but the number of valid data was only 181 because of non-completion of Form 2. The group of students who completed both forms was so heavily dominated by the Accommodation group that meaningful between-group statistics could not be generated. Second, tests were matched by difficulty and item content, but item-level statistics may also be needed for a more reliable match. In future iterations the reliability (or score variance) of each form could be calculated to ensure equivalence. A third limitation, which was operational in nature, is that students took the accommodated version of the test in the third week of school. This prevented students from having sufficient time to meaningfully interact with the assessment tools. Accommodations research has demonstrated that testing accommodations (such as those on the accommodated form) have greater effects when they are familiar to the student throughout the instructional school year. It is possible that greater accommodation effects would have been seen if students had more time to familiarize themselves with tools. Finally, our analyses of selected accommodations (tools) were also limited because students often picked more than one accommodation. We were unable to find sufficient numbers of students within “combination categories” (e.g., students who selected both read aloud and magnification accommodations) to conduct meaningful analyses.

## Grade 9 Students Performing Grade 8 Science Test

As of September 15, 2008, 212 students had completed Form 1 and 204 completed Form 2. Since the analyses focused on the difference in performance when accommodation and accessibility tools were not available and when they were, those students who did not complete Form 2 were removed from the analyses. Table 15 shows that at of the students who completed both forms, 130 were classified as Accommodation and 74 were classified as Non-Accommodation.

**Table 15: Classification of Participation Students**

	N	Percent
Accommodation	130	63.7
Non-Accommodation	74	36.3
Total	204	100

Table 16 displays the distribution of students across the participating schools. Unlike the grade 6 mathematics study, students in the grade 9 science study were over-represented in two schools (B and F) and under-represented in the remaining schools. This occurred, in part, because School B and F had large numbers of Non-Accommodation students participate in the study.

**Table 16: Distribution of Students Across Participating Schools**

	Percent
School A	1.0
School B	23.0
School C	10.8
School D	7.4
School E	5.9
School F	47.5
School G	4.4

Table 17 displays the percent of Accommodation students who had each of the accessibility tools activated for Form 2. Table 17 also displays the number of Accommodation students who had one, two, three, four, or five accessibility tools activated. The vast majority of students were provided access to the read aloud tool. Approximately half of the students were also provided access to the auditory calming and the color contrast tools. The majority of students also were provided to multiple tools.

**Table 17: Percent of Accommodation Students with Each Tool Activated for Form 2**

Tools	Percent
Standard Read Aloud	70.0
Low Vision Read Aloud	3.1
Magnification	4.6
Auditory Calming	51.5
Color Contrast	16.9
One Tool	39.2
Two Tools	39.2
Three Tools	6.9
Four Tools	0
Five Tools	1.5

Unlike the Grade 6 examination, achievement scores were lower for the accommodated form of the Grade 9 assessment (as compared to un-accommodated Form 1). This test also appeared difficult for students. Overall results for the two tests were 50% correct for Form 1 (without accommodation tools) and 46% correct for Form 2 (with accommodation tools). There was a noticeable difference in scores by sample group. Non-Accommodation student scores dropped from .58 to .50 while Accommodation student scores dropped from .45 to .43 from Form 1 to Form 2.

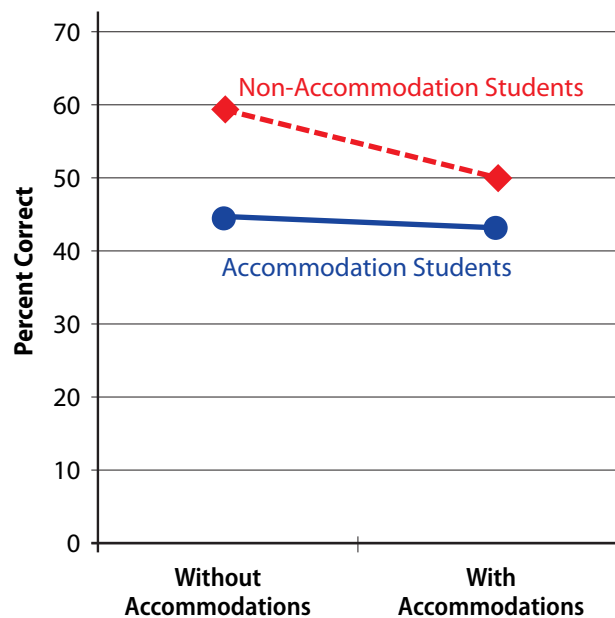
T-test analysis demonstrated that this difference was statistically significant ( $\alpha=.004$ ). For this test, the impact on typically achieving students was more pronounced than students with disabilities and English language learners, although both were statistically significant. Effect sizes of the score decreases were relatively low (Partial eta squared of .056 for typically achieving students and .025 for the SPEDELL group). Table 18 demonstrates score differences by group and size of test form effects.

**Table 18: Tests of Within-Subjects Effects by Form, Grade 9**

Source	Type III Sum of Squares	Degrees of Freedom	Mean Square	F	Sig.	Partial Eta Squared
Form	.23	1	.23	11.55	.001	.06
Form by Accommodation	.10	1	.10	5.07	.03	.03
Error	3.79	195	.02			

In addition to the difference in scores for Form 1 and Form 2 being statistically significant for both groups, there was also a noticeable difference mean scores by population. As shown in Figure 2, the small sample of Non-Accommodation students who performed in this study had a noticeably larger score decrease on Form 2 as compared to the Accommodation students. Specifically, the Non-Accommodation students experienced score decrease of more than 8 percentage points, while Accommodation student scores dropped by less than 2 percentage points.

**Figure 2: Mean Scores of Accommodation and Non-Accommodation Students**



To determine if there would be a significant difference in scores by built-in accommodations (tools) selected for the assessment population, the data was analyzed for read-aloud, magnification, auditory calming, color contrast, and GDX tools for the entire sample. In this study, 86.5% of students used the read aloud tool, 47.5% used magnification, 68.8% used auditory calming, 53.1% used color contrast, and 1.2% used GDX. When analyzing for achievement gains, we found there was no significant difference in achievement results between test forms when students were assigned a particular tool. We then analyzed the same data, but removed Non-Accommodation students to determine if any particular accommodation had an effect on the achievement of the Accommodations group. We again found that there were no statistically significant differences in scores by test format when particular tools were used. We could

not conduct the same analyses for the Non-Accommodations group because every typically achieving student in Grade 9 were assigned read-aloud, magnification, auditory calming, and color contrast for Form 2.

The final set of analyses attempted to link perceptions of use of assessment tools with change in scores from Form 1 to Form 2. The hypothesis tested for these analyses was that if students feel comfortable with technology tools, it will have an impact on their achievement. In this case, we found no significant correlations between student preferences and score gain. In total, 87.5% of students said they “liked” taking the test on the computer and 86.4% said they would like to take a computerized test in the future, but these preferences did not have an impact on achievement, either positively or negatively. Non-significant correlations existed for both the total population and students in the SPEDELL group.

## Summary of Findings

The data collected as part of this study provides evidence that the use of a universally designed computer-based test delivery interface that provides a variety of accessibility tools is feasible for the FCAT or other state testing programs. Despite the very tight time frame for the study, the fact that the software had to be installed during the first two weeks of school, and that data was collected during the third week of school (a time period that is perhaps the busiest and most hectic for schools), Test Coordinators indicate that they experienced no major problems preparing for and administering two tests composed of released FCAT items using NimbleTools.

In addition, only two technical calls were received during the course of the study. One call focused on whether NimbleTools could be re-installed on a computer whose hard drive had been wiped clean by the school’s technology director after NimbleTools was previously been installed. The Test Coordinator was told to re-install NimbleTools, which took approximately 2 minutes. The Test Coordinator then reported no further problems as students used NimbleTools to perform the tests. The second call inquired what to do after a student’s computer crashed. The Test Coordinator was informed that the student could restart the computer, launch NimbleTools, and log in using the same User Name and Ticket Number. After doing so, the student was able to continue working on the test from the point where he was prior to the computer crash. Beyond these two calls, no other technical calls were made during the test administration window.

Following testing, one Test Coordinator noted that some students using Windows became confused when the NimbleTools logo appeared on the screen after starting the application. The Administration Instructions stated that students using some versions of Windows should press the Escape key to access the login screen after startup. While following up with the Test Coordinator, it was revealed that the proctor had not been provided a copy of the Test Administration Instructions and therefore was not familiar with the log in procedures.

Test Coordinators and students also provided ample evidence that NimbleTools, a computer-based test delivery system that is universally designed to provide students with multiple accessibility tools, is usable for the FCAT. Test Coordinators reported that students did not experienced any major problems taking tests using NimbleTools. In addition, the vast majority of students reported that NimbleTools was easy to use. After using NimbleTools to perform two tests, the vast majority of students reported that they wanted to perform tests on computer in the future. Finally, those students who used the NimbleTools accessibility

tools overwhelmingly reported that the tools were easy to use and were helpful during testing. Collectively, evidence provided by the Test Coordinators and by students indicates that NimbleTools and the embedded accessibility/accommodation tools were usable by students when performing FCAT test items.

In addition to evidence from the student survey, results from the test administrations also provide evidence that the use of NimbleTools had differential effect on test performance of students in need of the tools as compared to those were not identified as requiring the tools. While test scores were lower for Form 2 for the limited number of non-Accommodation students who performed the grade 5 mathematics tests, scores for Accommodation students who used the NimbleTools accessibility tools increased slightly. For the grade 9 science test, scores for Non-Accommodation students dropped considerably when they used the NimbleTools accessibility features while scores for the Accommodation students decreased by only a small amount. While a number of factors, including fatigue, may have contributed to these score changes, both studies provide evidence of a differential effect that favored Accommodation students who were provided access to the accessibility/accommodation tools that were aligned with their individual needs. This differential effect provides evidence that NimbleTools is effective for providing computer-based accommodations for the FCAT and similar state tests.

## **Policy Implications**

This study provides evidence that NimbleTools, a universally designed computer-based test delivery interface with several accessibility tools embedded into its system, is feasible, usable, and effective for meeting the accommodation needs of students with disabilities and special needs. This study also provides evidence that the NimbleTools non-accommodated interface is feasible and usable by students who do not require test accommodations. In addition to these research findings, the study has several policy implications. These include readiness of schools, flexible delivery methods, importance of a close working relationship between test coordinator and technology director, future study designs, and wording of future request for proposals/invitations to negotiate. Below, each of these issues is discussed separately.

### **Readiness of Schools**

Although this study was conducted in a small sub-set of Florida schools that was not intended to be representative of all schools in Florida, the study demonstrates that schools similar to the participating schools are ready to deliver the FCAT on computer to students with disabilities and special needs, as well as those without special needs. Given these findings, the first policy implication that results from this study is that the Florida Department of Education use a universally designed computer-based test delivery system to provide test accommodations to students with disabilities and special needs.

### **Flexible Delivery Methods**

This study employed a local installation and delivery method, often known as a fat client approach. A fat client approach decreases Internet traffic during testing, which allows schools with poor Internet connectivity to test more students simultaneously without overburdening their Internet service. As noted above, none of the schools reported Internet problems during testing and none reported that a student was unable to complete testing due to a computer malfunction. Collectively, this evidence suggests that schools with similar technology facilities

across the state are well positioned to implement a fat client approach to computer-based testing. However, some of the schools that participated in the study had more robust Internet connections that would enable them to implement on-line delivery of a test, known as a thin client approach. While a thin client approach places greater demand on the Internet connectivity, it eliminates the need for installing testing software on computers prior to test administration and removing that software after testing is complete. Given the wide range in the technological capacity of schools across the state, the Department of Education should consider a mixed approach to computer-based testing. The mixed approach would enable schools with robust technological capacity to implement a thin client approach while schools with modest technological capacity employ a fat client approach. If a mixed approach is adopted, however, it is important that the test experience, including the log in, interface design, and functionality of accessibility tools, be identical for students across the thin and fat client approaches.

### **Close Working Relationship between Test Coordinator and Technology Director**

Although no major technical problems arose during testing, this study highlights the importance of working closely with schools to implement a computer-based testing program and requiring schools to coordinate the activities of the Test Coordinator and the local technology director. As one example, one school Test Coordinator discovered on the day of testing that the testing software had been removed from a computer by the technology personnel. Had the technology personnel and the Test Coordinator been required to have a closer working relationship, this problem may have been avoided.

As a second example, Test Coordinators were provided with two ways of installing the testing software on school computers. First, they could download the installation package from a web site. Second, they were provided a CD-ROM that contained the installation package. In many districts, downloading software from the Internet is blocked. As a result, the Test Coordinators were not able to capitalize on the more efficient approach to installing the testing software through the Internet, and instead had to use the CD-ROM. This example highlights the importance of providing multiple strategies for preparing computers for testing. It also highlights the importance of requiring the Test Coordinator and technology director to work together so that appropriate privileges are provided to access software through the Internet.

### **Future Study Design**

Due to time constraints and a desire to make the study easy for schools to implement, this study employed a two-group repeated measures design. As noted above, this design has three potential shortcomings. First, it does not permit analyses that examine differences in the difficulty of test forms. Second, it does not control for the potential effects of fatigue. And third, it does not allow one to obtain a direct estimate of the effect that the use of accommodation/accessibility tools has on the performance of Non-Accommodation students.

As described above, an alternate three-group counter-balanced design was discussed as an option for this study. Given time constraints and the difficulty of implementing this more complex design in schools at the start of the year, the research team was encouraged to employ the simpler two-group repeated measures design. In the future, the state and other researchers who wish to examine the efficacy of test accommodations are strongly encouraged to employ the complex three-group counterbalanced design. Doing so will allow for a cleaner estimate of the effect of a given accommodation or set of accommodation.

## **Wording of Future RFP/ITNs**

Given that this study provides evidence that a universally designed computer-based test delivery interface is feasible, usable, and effective, a final policy implication pertains to language in future requests of proposals and invitations to negotiate. During the course of this study, the study director conversed with personnel at the National Center for Educational Outcomes about shortcomings of current language regarding test accommodations in many state RFP and ITNs, including Florida. While it appears many states desire a universal design approach to student assessment, the language employed in state RFP and ITNs does not require respondents to provide universally designed solutions for computer-based test delivery. Specifically, the language employed focuses on and treats each accommodation as a separate entity, and does not require a comprehensive and coordinated strategy for providing accommodations for a computer-based test. As a result, respondents address each accommodation separately, and often offer solutions that are separate and distinct from the solution offered for non-accommodated students, and in some cases separate and distinct solutions for each accommodation. Enabling respondents to provide separate and distinct solutions for general education and accommodated students increases the cost of a testing program, increases the burden on schools and teachers to provide support for different applications, requires students to learn different interfaces depending on their need, and prevents students from altering their decision about the use of an accommodation during testing. Working with the members of the National Center for Educational Outcomes, the following language was drafted and is recommended for consideration and modification for future solicitations.

### ***Universal Design Computer-Based Testing Language Suggested for Use and/or Modification for Future Solicitations***

When applied to computer-based testing, the Principle of Universal Design maintains that the interface students use to access test content, interact with test materials (e.g., manipulatives, calculators, reference sheets, digital rulers, etc.), and record responses should not adversely affect test validity for students with or without disabilities or special needs. A universally designed computer-based test will provide multiple techniques for students to access test content, interact with testing materials, and record responses. Each technique will be designed to meet a specific need or accommodate the manner in which a student is accustomed or prefers to access test content, interact with test materials, or record responses.

A universally designed computer-based test will adjust the accessibility and interaction tools available to students based both on each student's individual need and the construct being measured. This principle requires careful thought about what construct a test item or the test as a whole is intended to measure, and requires that one or more accessibility or interaction tool be made unavailable when the need addressed by the tool overlaps with the measured construct.

As an example, the reading and signing of text by a computer-based test is intended to decrease the effect that the construct of decoding has on student performance when test items contain text. In all cases, except when the test or test item is intended to measure decoding skills, decreasing the effect of decoding on student performance is desirable. However, the use of reading and signing of text would be inappropriate when the test or test item is designed to measure decoding. A universally designed computer-based test will provide

access to accessibility and interaction tools when their use does not overlap with the measured construct, and will make tools unavailable when overlap occurs.

The Bidder's computer-based test delivery system must be based on the principles of universal design described above and must flexibly meet the needs of each individual user. At a minimum, the universally designed solution must:

1. Allow users to navigate and interact with all functional elements of the test delivery system using a standard mouse, a keyboard, touch screen, Intellikeys, switch mechanism, sip-and-puff device, eagle-eyes, and other assistive adaptive communication device. This specification requires the entire test interface, including all elements of an item, navigation buttons, option menus, setting controls, and controls be tab-entered enabled.
2. Allow all text appearing within each test item and on all interactive areas of the screen, including menus and navigation buttons, to be read aloud using a human voice and/or synthesized voices.
3. Allow blind or visually impaired students to access verbal descriptions of zones, menus, options, and buttons as they navigate through the test interface.
4. Provide spoken descriptions of all graphics, drawings, tables, functions, formulas, and other non-text-based elements of an item.
5. Provide multiple magnification tools that allow users to alter the level of magnification at any time during testing. At a minimum, the universally designed solution must include:
  - a. a magnification tool that enlarges all elements of the screen (including all components of an item, all menus, and all navigation buttons);
  - b. a magnification tool that enlarges only the test item and provides visual or auditory clues when a portion of an item appears off the screen when magnified; and
  - c. a magnification tool that is specifically designed for low vision students who prefer to view magnified information in a fixed location.
6. Provide an auditory calming tool that allows students to select from among a list of pre-approved sound files and to have the sound file play softly in the background as the user works on the test. The auditory calming tool must allow users to change their sound selection, alter the volume, and turn the tool on or off at any time during testing. The auditory calming tool must also interact with the read aloud tool so that reading and auditory calming sound files do not compete for the user's attention.
7. Provide signing of all text contained in an item. The signing tool must allow students to select text to be presented in American Sign Language or Signed English. The signing tool must also function identically to the read aloud tool so that text is highlighted on the screen as it is being signed and the student is able to select blocks of text that are then presented in sign. The signing tool must allow students to alter the size of the signing display. To the extent possible, the signing tool will allow students to have signing be presented with or without captioning, mouthing, and speaking of text being signed.

8. Allow users to alter the color contrast of the display. At a minimum, the color contrast tool must allow users to:
  - a. place a colored overlay over the entire screen;
  - b. alter the background color over which information is displayed;
  - c. alter the color of non-graphic elements associated with an item, including text, tables, formulas, and functions; and
  - d. select from a set of predefined color combinations known to increase contrast and visual perceptibility of text.
9. Allow users to mask portions of the screen as they perform an item. At a minimum, masking must allow users to:
  - a. mask all elements of the screen that are not directly associated with the current item, including navigation buttons, menu options, time remaining, etc.;
  - b. view an item prompt in isolation from answer options;
  - c. reveal and mask individual item answer options; and
  - d. mask and reveal any element of an item.
10. Allow users to view an item using an isolated viewer. The isolated viewer must hide all elements of an item that are not contained within it. The isolated viewer must allow user to change the size of the viewer, the level of magnification within the viewer, and the color contrast of information presented within the viewer.
11. Allow users who are Blind or visually impaired to navigate and work within the test delivery system using the same keyboard functions employed when using JAWS.
12. Allow all text associated with an item to be presented on an electronic Braille display attached to the user's computer.
13. Allow all user preferences to be set prior to testing and to have the tools specified for a given user made available when a test is first opened. Depending on the user's preference, the pre-specified settings will be automatically applied when the test is launched or the setting/tools will be accessible through an options panel that has been customized based on those settings.
14. Provide users access to heritage language translation of words or phrases without requiring them to type or otherwise enter words or phrases into a translation dictionary. Preferably, heritage language translation will function similarly to the read aloud tool, allowing users to click on a text element to access the translated version of that element. Heritage language translation must be based on the context in which a word is used in the item prompt, answer option, or other element associated with the item.
15. Provide users access to alternate language presentation of the test interface and test items. Alternate language presentation must also be provided for read aloud of text. Read aloud of alternate language must be possible using either a human or synthesized voice for languages that have synthesized speech profiles.

To meet the requirement for universal design, the solution must:

- A. Function identically on Windows, Macintosh, and Linux operating systems
- B. Function with all elements of the test, including the test items, directions, reference sheets (e.g., periodic table), formula sheets, calculators, and other electronic tools provided for students to use during the test (e.g., protractor, ruler, etc.).
- C. Embed all accessibility tools within the test delivery system and must not require users to employ any software that is external to the test delivery system, including text readers, DVD players/controllers, audio players/controllers, or magnification tools.
- D. Allow multiple tools to be used in conjunction with each other and, to the extent possible, the tools must interact in manner that increases access to test content. As an example, when a user is employing the read aloud and auditory calming tools simultaneously, the system must recognize this and must decrease the volume of the auditory calming tool, or turn it off entirely, based on the user's preference. Similarly, when the user is viewing an item with alternate contrast and then activates an isolated viewer, information presented within the viewer must also be presented with the alternate contrast.
- E. Record the use of each tool for each test item in order to document the use of each accessibility tool for each individual item. This information must be stored along with the user's response to each item.

To improve universal access to test items, the design of all test items must anticipate the use of one or more accessibility tools, and thus each item must be designed in a manner that does not disadvantage a student who makes use of one or more accessibility tool. As an example, an item that contains a colored graphic (e.g., a pie chart) must not reference specific colors in the prompt or answer choices (e.g., referencing the blue portion of the chart may disadvantage a student who is employing a yellow overlay which may cause the blue portion to have a green hue).