Testing on Tablets: Part II of a Series of Usability Studies on the Use of Tablets for K-12 Assessment Programs

White Paper

Laurie Laughlin Davis, Ph.D.
Vice President, Measurement Services
Pearson

Ellen Strain-Seymour, Ph.D.
Director, Digital Content Development
Pearson

Heather Gay
Director of Research
Mediabarn Research Services

ALWAYS LEARNING

PEARSON

About Pearson

Pearson, the global leader in education and education technology, provides innovative print and digital education materials for preK through college, student information systems and learning management systems, teacher licensure testing, teacher professional development, career certification programs, and testing and assessment products that set the standard for the industry. Pearson's other primary businesses include the Financial Times Group and the Penguin Group. For more information about the Assessment & Information group of Pearson, visit http://www.pearsonassessments.com/.

About Pearson's White Papers

Pearson's white paper series shares the perspectives of our assessment experts on selected topics of interest to educators, researchers, policy makers and other stakeholders involved in assessment and instruction. Pearson's publications in .pdf format may be obtained at: http://researchnetwork.pearson.com/library.

Overview

Pearson is dedicated to offering our customers and end-users the broadest range of possible tools and technology for instruction and assessment—whether formative or summative. We have made a commitment to research on tablet devices to inform the design of the user interface for the tablet so we minimize student challenges with this technology, optimize student experience, and enhance measurement.

Today's students use digital devices on a regular basis both in and out of the classroom. This suggests that these devices may provide alternatives for delivering both learning content and assessments. Although it seems obvious that advantages in flexibility and student engagement will be obtained by providing learning and assessment content through a variety of digital devices, it is important to consider the comparability of both instructional efficacy and scores based on assessments delivered across devices. For example, in previous tablet research, Pearson has observed some student interactions that could be a concern like keyboards covering critical content and the complexity of using a fingertip to maneuver within assessment questions. It is important to identify these types of issues and, when possible, design the user interface to facilitate comparability of student experience across delivery platforms.

This report summarizes the findings from a November 2012 research study conducted to evaluate the usability of tablet devices as tools for assessment. Students were given tablet devices and were asked to walk through a test form while sitting with a researcher. The researcher used a "think-aloud" protocol where the student with the tablet talked about their experience as they move through the test form.

The goals of the study were to:

1. Observe student interactions with tools, navigation, and question components on the tablets and identify specific areas where the interactions on tablets could present challenges
2. Understand the ergonomics of short-term tablet usage in order to identify any areas of concern around fatigue or strain
3. Observe the impact of differing screen sizes and device features across a limited range, using two different devices

While the findings of this study are not intended to be definitive on their own, they do suggest areas of improvement for user interfaces for tablet-delivered assessments and areas for more targeted research to understand the potential for comparability across a range of digital devices used for assessment delivery.

Methodology

The study was conducted over three weeks in November 2012 on site at 16 school locations in four states—Maryland, Virginia, Florida, and Texas. In total 63 students participated.

	GRADE 5	HIGH SCHOOL	TOTAL
Gender			
Male	16	15	31
Female	15	17	32
Ethnicity			
African-American	6	6	12
Asian	1	1	2
Caucasian	19	18	37
Hispanic	5	7	12
Familiar with:			
iPad	24	28	52
Nexus 7	5	8	13
TOTAL	**31**	**32**	**63**

Researchers used a "think-aloud" protocol (concurrent verbalization) whereby students narrated their experience as they moved through the test questions on the tablet. As necessary the researcher prompted the student around specific areas of feedback. Before beginning each study session the lead researcher modeled the "think-aloud" behavior for students and used standard usability study language explaining that the purpose of the study was to test the software, not student knowledge.

Students saw test questions on two different tablets—a 10" iPad and a 7" Google Nexus 7. The study design was counterbalanced so that half of the students used the iPad first and half used the Nexus first. Additionally, tablets were labeled as "Tablet I" (iPad) and "Tablet N" (Nexus 7) to avoid preferences due to brand awareness—although many students recognized the iPad in spite of this. Each tablet test form contained 10-15 test questions including a variety of different question types and content areas. Researchers worked one-on-one with students for a total of 45-60 minutes and recorded notes about students' observations on a study protocol. In addition, audio recording of the session was conducted. At the end of the session, students were asked to complete a short survey on their experience with the tablet and level of computer expertise.

Results

Overall, students were very excited about the prospect of taking a test on a tablet device. Of the 63 students in the study, all but one had seen a touch-screen device previously; however, frequency of use and use in an academic setting was more variable. While some interactions generated frustration, students were generally able to navigate within the testing software and interact with questions in a fluid manner. Most students were able to accomplish the test tasks even if they were not familiar with the specific testing software. Key findings are discussed in more detail below.

Screen Size

The Partnership for Readiness for College and Careers (PARCC) and Smarter-Balanced Assessment Consortium (SBAC) have both specified a minimum screen size of 9.5″ for tablets to be used in consortia assessments (PARCC, 2012; SBAC, 2013). While an exhaustive review of screen sizes, aspect ratios, and other screen size issues likely to impact usability was beyond the scope of this study, the inclusion of two screen sizes facilitated observation of the effect of screen size on students' test-taking activities. Students generally found the 10″ form factor of the iPad to be acceptable for viewing and working with test content and found the smaller form factor of the Nexus 7 to be more challenging. Without prompting or direct questioning, most students commented that the Nexus device was "much smaller" than the iPad. Students also remarked on the use of smaller devices for high-stakes testing, with comments such as "[the smaller device] seems less serious" than the iPad and that it "might be okay for taking a quiz, but not a test."

"The font for Tablet I was bigger and easier to read. There was less room for error on the smaller tablet and therefore more difficult to drag, for example, the graph. It was easier on Tablet I."

Specific challenges arose in terms of viewing text and selecting objects with precision on the smaller Nexus device. While most students were able to read and complete the question on the Nexus, many said that they thought they would have difficulty using the device for long reading passages, and many students tried to pinch-and-zoom to enlarge the text (a feature that was not enabled in this study). In addition, some questions were difficult to complete because the screen size was too small to accommodate the students' fingers on the area they were trying to manipulate. Although this was an issue observed with some questions for both large and small form factors, it was exacerbated and observed more widely with the smaller tablet.

Device Positioning & Ergonomics

Although students were given no specific direction about how to position the tablets, the majority of students placed the tablets flat on the table and leaned over them to view the screen. A few students held the device throughout the test (more likely with the smaller Nexus 7 tablet), and others propped the iPad up on its cover (the Nexus 7 cover did not permit this). The limited range of positions observed with students in this study stands in contrast to previous research conducted with adults which indicated that tablet users take advantage of their devices' many potential display positions, changing device position and their bodily position based on the task (Young et al., 2012).
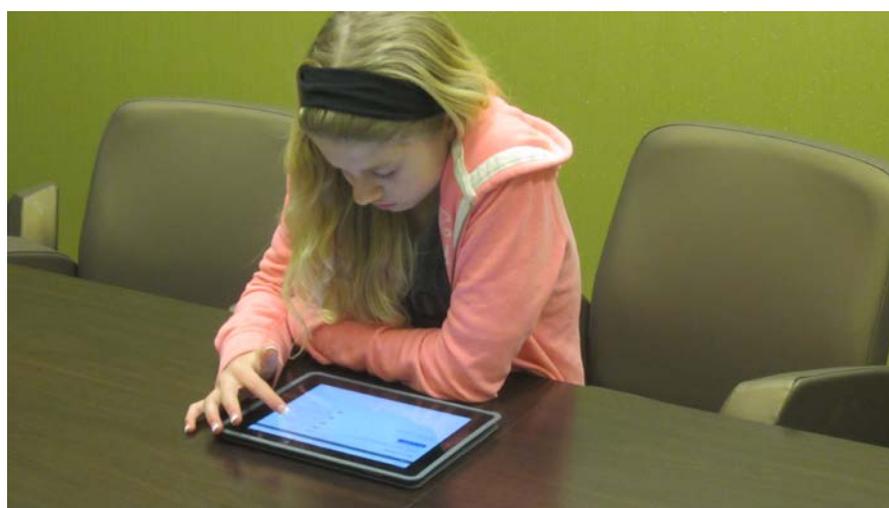

Photo caption: Student leaning over tablet placed flat on the table

Most students were able to complete the study questions on both devices within a total of 30-45 minutes without significant strain or difficulty. While no physical discomfort or strain was explicitly observed, students may have experienced limited discomfort within the session or within their prior experiences with tablets, since some students expressed concern without any prompting or explicit questioning on this topic. Some students mentioned that they would likely suffer some issue such as neck pain, thumb strain, or headache due to holding or viewing the device for a lengthy testing session. A few students also mentioned that the angle of the device in conjunction with overhead lighting or eye glasses may create a glare that would cause eye strain.

Question Types

The following question types were included in this study:

- Multiple-choice
- Drag-and-drop

- Hot spot
- Bar graph
- Point/line graph
- Fill-ins/Grid-ins
- Short-answer
- Extended text-entry

> *"I liked how interactive the questions were and the way it was easy for me to select different options with my fingers."*

A number of question types were included in the study in anticipation that some question types might be an intuitive fit for a touch-screen tablet interface, while others might lend themselves to the use of peripherals (external keyboards or styluses) in order to improve the precision of student input. Additionally, even within question types, some variability was intentionally included. For instance, within a drag-and-drop question type, both the size of objects to be dragged and the rules around how the dragging functionality is configured (e.g. do objects snap to place when they are close enough?) could influence a student's experience.

Multiple-choice

For the most part, students were successful in responding within the familiar multiple-choice question framework. Some students did experience slight precision issues because they had been trained in paper and pencil tests to fill in the circle completely and not go outside of the circle. Therefore, they did not realize that touching anywhere on the text of the response option would select the answer and, as a result, they had slight difficulty pinpointing the "bubble" to make their selection. The roll-over effect of the cursor switching to a pointing finger did not appear on the tablet due to the absence of cursors on a touch-screen.

Drag-and-Drop

Students appeared to capably interact with draggable objects on the tablet, with some commenting favorably on the ability to interact directly using their finger and not having to use a mouse as an intermediary to express their intent. However, when the "target area" for dropping the object was small or close to other target areas, students sometimes struggled to precisely place the object (see High school geometry proof drag-and-drop below). As a result, the objects might "snap back" to their original position rather than sticking to the place where the student intended to drop the object. This snapping effect sometimes caused students to question whether their answer was correct; some thought that the tablet was giving them a hint that their answer was wrong. Conversely, when the target area was sufficiently large or separated from other target areas (see High school history timeline drag-and-drop below) the question was relatively easy for students to interact with.

Construct a proof for the theorem given by arranging the statements and reasons in the table below.

If polygon ABCD is a square, then ABCD is a quadrilateral.

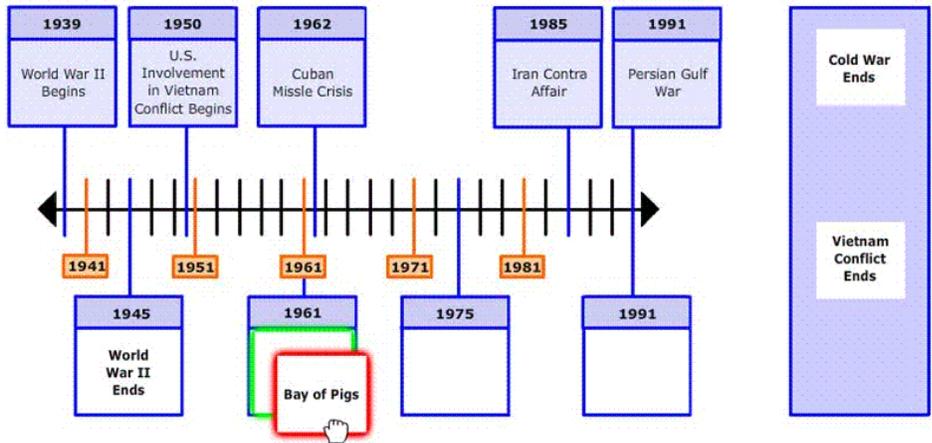| Statement	Reason
1. *ABCD* is a square.	1. Given
ABCD is a parallelogram.	
ABCD is a quadrilateral.	
ABCD is a rhombus.	

Statements

Reasons

If a figure is a rhombus, then it is a parallelogram.

All parallelograms are quadrilaterals.

All squares are rhombuses.

High School Geometry Proof Drag-and-Drop Question

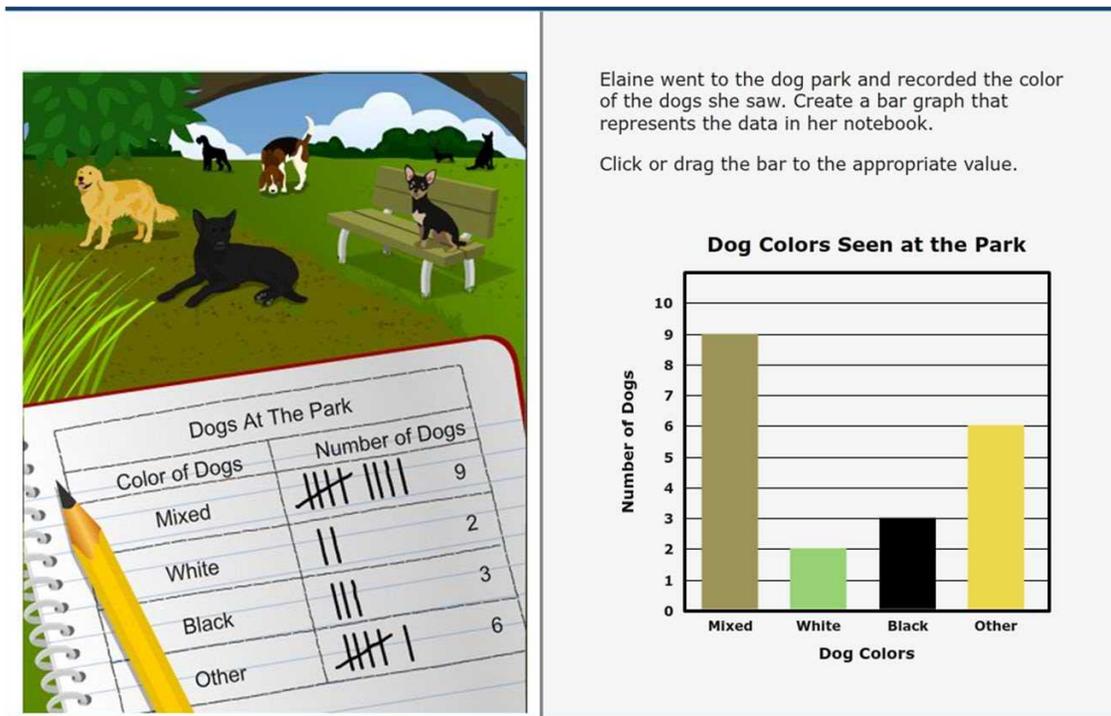Place the events in the blue box onto the correct place along the timeline.

1939	1950	1962	1985	1991
World War II Begins	U.S. Involvement in Vietnam Conflict Begins	Cuban Missle Crisis	Iran Contra Affair	Persian Gulf War

Cold War Ends

1941   1951   1961   1971   1981

Vietnam Conflict Ends

1945	1961	1975	1991
World War II Ends	Bay of Pigs		

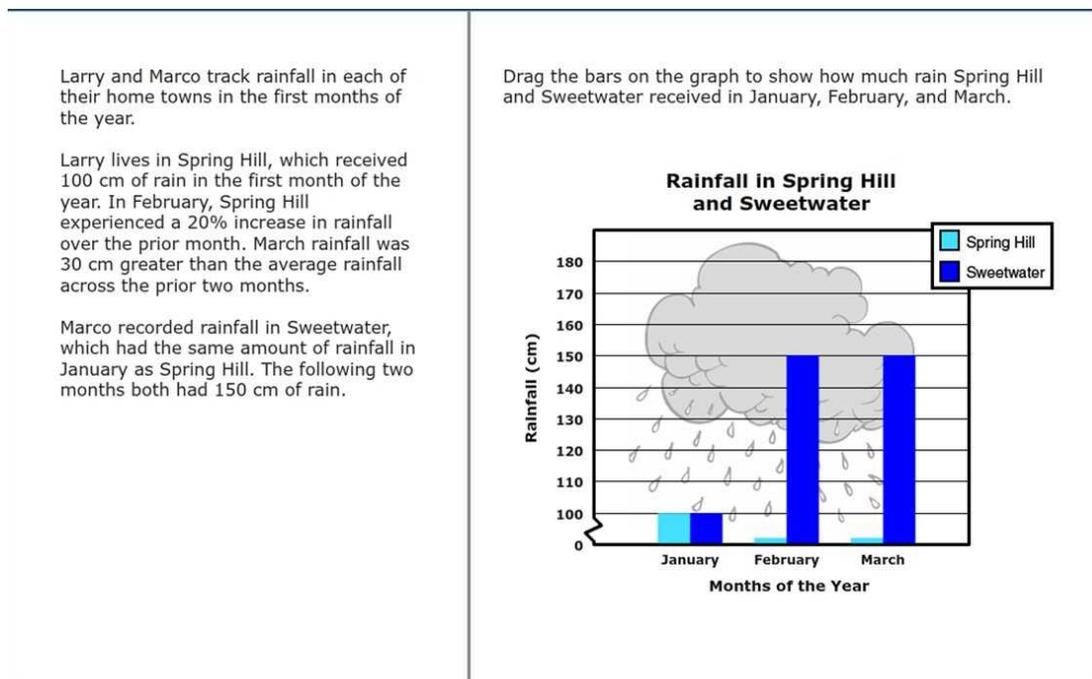**High School History Timeline Drag-and-Drop Question**

Hot spot

Most students experienced no difficulty in placing a mark or selecting an image using their finger within a hot-spot question. However, a few students did not realize that the mark would be placed simply by tapping the screen. These students assumed that they needed to select the dot tool from the toolbar, which placed dots on the screen. However, since marking tools (like the dot tool) are intended to support a student's thought processes but are never scored, this was a problematic issue. On an actual test, the question would be considered unanswered. This usability issue is normally avoided by not including the dot tool on these types of questions, but unfortunately, that standard solution was not available in preparing the usability study materials.

Bar graph

Bar graph questions worked well on tablets so long as the width of the bar was large enough for students to drag with their finger and the bar was sufficiently separated from other bars (see Grade 5 dog park bar graph question below). Thinner bars or bars that were spaced too close together proved challenging for students (see High School rainfall bar graph question below) as they were unable to grab the bar at all or inadvertently grabbed the wrong bar.

Grade 5 Dog Park Bar Graph Question

Larry and Marco track rainfall in each of their home towns in the first months of the year.

Larry lives in Spring Hill, which received 100 cm of rain in the first month of the year. In February, Spring Hill experienced a 20% increase in rainfall over the prior month. March rainfall was 30 cm greater than the average rainfall across the prior two months.

Marco recorded rainfall in Sweetwater, which had the same amount of rainfall in January as Spring Hill. The following two months both had 150 cm of rain.

Drag the bars on the graph to show how much rain Spring Hill and Sweetwater received in January, February, and March.

Rainfall in Spring Hill and Sweetwater

High School Rainfall Bar Graph Question

Point/Line graph

Point or line graph questions requiring students to plot points and lines on a Cartesian coordinate graph were sometimes challenging for students because the student's finger obscured the placement of the point. As a result, some students expended additional effort to move their points to the correct locations. Holz and Baudisch (2010) similarly observed that touch-screen input accuracy may suffer when the finger blocks some part of the graphical interface. Additionally, as with hot spot questions, some students did not realize that they could simply touch the graph to plot a point. Instead they attempted to use the dot tool or line tool to interact with the question.

"My finger is bigger than the dot. It's hard to see where the dot is going."

Fill-ins/Grid-ins

In order to maintain literal comparability with paper-and-pencil assessments, some states have elected to carry the concept of a "grid" over to computer-based testing for answering numerical constructed response math and science questions. This typically takes one of two forms: a series of boxes that each hold a single character or those boxes with a series of "bubbles" (for 0-9 and decimal points) beneath each box in a near duplication of the appearance of a scannable answer document. While the first format allows the student to see the maximum

number of characters allowed in a similar way as in a paper-based test, the second allows for both maximum and allowable characters to be detected by the student. The first of these solutions – individual boxes, but no bubbles – was included in this study and is referenced here by the term "grid-in." Other states have adapted this format for online use by allowing the numerical response to be typed into a short answer, or "fill-in" box that only allows a limited range of characters to be used. This study also included an example using this fill-in format. For the grid-in interface for numeric entry, each individual box had to be tapped separately to enter a number, which students found frustrating. When shown the same question as a fill-in rather than a grid-in, students more easily interacted with the fill-in version and described preferring the fill-in version. For both formats, the onscreen keyboard native to the tablet defaulted to the alphabet even though the answer allowed only numbers. Students had to know or figure out how to navigate to the numeric/symbol keyboard to enter their answer.

Short-answer

In this study, short-answer questions took a variety of forms ranging from typing labels to identify parts of an eyeball to captioning a photo. In all cases students were asked to use the onscreen keyboard to enter their answers. While some students did attempt to use standard keyboarding skills when responding to these questions, these were quickly abandoned in favor of single or double-finger hunt-and-peck style typing. Although it took these students a moment to adapt, most students did not appear to be excessively bothered by this and were able to complete the question with this approach. Many students remarked that they liked the onscreen keyboard though a few students commented that this would be more of a concern if they had to type more than a few words. However, students did seem temporarily dismayed or inconvenienced by the fact that the opening of the onscreen keyboard pushed question content off-screen. Students had to scroll back up to enter their response or to see information they needed to answer the question.

Extended text entry

This study included two longer text entry questions. The first question asked students to briefly describe the difference between a food-web and an ecosystem with an expected response being a couple of sentences. Most students were able to open and use the onscreen keyboard to complete this question. However, as noted with short-answer questions, when the onscreen keyboard opened, it pushed the top of the question off screen and students had to scroll up to see the prompt. Students wanted to be sure they were answering the right question, and several wanted to double-check the spelling of "ecosystem" so the ability to see the prompt while answering the question seemed to be of significance to students.

The second question asked high school students to create an outline for an essay on improving the economy. With the essay outline question, students were asked to begin typing using the onscreen keyboard, but after a few minutes were provided with an external keyboard to complete the question. This question appeared to be problematic for students to complete by using the onscreen keyboard. The keyboard often covered the text entry boxes, making it difficult for students to see what and where they were typing. When students were given the opportunity to type on the external keyboard with the iPad, most typed more text and typed more quickly and accurately (using keyboarding techniques) than they had using the onscreen keyboard. Several students described that it was easier for them to feel the keys on the external keyboard. In addition, the external keyboard allowed them to focus their vision on the screen without the keyboard being in the way of the text entry boxes. Nearly two-thirds of high school students said that they preferred using the external keyboard to the on-screen keyboard.

> *"It's like after you type in the first one you cannot go back and look at the prompt so you are stuck looking at the top and you have to like figure out how to close the keyboard and then go back to your thoughts."*

Discussion

The results of this study were compelling in the area of screen size; the smaller screen sizes that fall below PARCC's and SBAC's recommended sizes may interfere with performance in a high-stakes testing situation. However, with this qualitative research conducted on only two devices, it is not possible to isolate definitively a size below which a student's chances for a successful testing experience are compromised. While the theory that the bigger the screen the better the experience cannot be decisively discounted, it is far more likely that there is an optimal range of screen sizes within which students tend to perform similarly with all other factors being equal. However, this approach of specifying an acceptable range or minimum size would still require discipline around user interface design and question specifications. Any button, control, drag-able/tap-able, textual element, or detailed imagery must be evaluated for acceptability on the minimum screen size. As an added safeguard, controls for enlarging should also be provided either for all students or, at the very least, on tablets where the screen size may be smaller.

The exclusion of screen sizes smaller than the 9" from high-stakes testing does not mean that devices such as the Nexus do not have a place in the classroom. The students' generally positive reaction to the device and ease in navigating some of the assessment materials suggests the following:
- Smaller devices may be acceptable for short quizzes or knowledge checks used for lower-stakes purposes.

- Students' cited familiarity and satisfaction with devices such as the Kindle or Nook and the ease in which students used the Nexus for reading passages suggest that Nexus use as a reading device (ideally with adjustable font size) may be desirable.
- Tests that are designed exclusively for smaller devices and perhaps limited to question types that do not require input precision may be acceptable, assuming that they are low stakes or that comparability and fairness issues are avoided by virtue of all students experiencing the test on a smaller device such as the Nexus or iPad Mini.

This study's findings around ergonomics and gesture support are less definitive. Follow-up studies to further explore these areas might involve a range of research questions:

- Do students experience discomfort, fatigue, or eye strain during long testing sessions with tablets? How does this experience compare to discomfort experience with desktop and laptop computers?
- When students are provided with stands and external keyboards, do students prefer certain device/peripheral positioning based on task (reading, extended text entry, choice selection)? Do they maneuver the tablet into these preferable positions on their own accord or fail to make such position adjustments during the course of a test?
- When a device and the testing software support gestures do students know to use these gestures? What role does device familiarity play? Do any complications come from the variability of what gestures are supported on which devices?
- Do students use gestures like pinch and zoom to counteract decreased input precision? If so, is this done fluidly and ubiquitously or is it limited to some students and/or introduce other issues, such as not being able to see the full context of the question when zoomed in to place a dragger, for instance?

In terms of some of the questions above as well as the usability of some question types on the tablet, one might expect that familiarity with a particular device and with the testing software will maximize students' chances for being successful in using the device and the software on testing day. One might expect that students would use the same devices in an instructional environment as in a testing situation, such that they have a greater likelihood of being familiar with the device, the gestures it supports, peripheral devices that can be used with it, and the ways its position can be manipulated for particular tasks. Practice tests and tutorials provided using the computer-based testing software have long been used to prepare students for online testing. The nuances of how that software works on a tablet provides additional impetus to continue this practice.

Where particular question types or the specificities of a particular question decrease usability on the tablet, a number of solutions may be possible. A few of these are described below.

Question Type	Possible Usability Issue	Possible Solutions
Multiple Choice	Students mistakenly believe that only the select-able area of an answer choice is the radio button and have difficulty selecting this small region with their finger.	• The "hit" area of the radio button can be enlarged. • The selectable answer choice area could be represented as a colored region that looks like and acts as a button. • Tutorial materials can help to inform students that the entire area is selectable.
Hot Spot	Students attempted to use tool marks to indicate selection.	• Configure available tools to avoid confusion on these question types.
Drag-and-Drop	Small draggers can be difficult to place on small target areas. When the dragger snaps back to its original position due to not being fully placed over a target area, students can misinterpret this as the system telling them that their choice was incorrect.	• Question specifications should mandate the minimum size for any dragger and any target area. • Target areas should not butt against one another, so that highly forgiving snap-to logics result in draggers snapping to the target area rather than return to their original position. • Tutorial materials can be used to dispel misperceptions.
Bar Graphs	Bars can be difficult to grab and position on the tablet.	• Bars should have a width that is greater than the average width of the human finger. • When paired bars are used, even greater bar widths should be used to compensate for the proximity of the two bars. • When not using grouped bars, bars should be spaced apart to account for the reduced precision of the finger in comparison to the mouse. • Bars should be able to be positioned by dragging or tapping. • These actions should be possible without being forced to grab the top of the bar where one's finger might occlude the view of the lines behind the bar.
Grid-in	Tapping on each box of a grid-in question with care to select the right box can be tedious.	• Grid-ins can be changed to fill-ins. • Or auto-advance capability should be added.
All Question Types	Since the precision of a finger as an input device is always less than a mouse's precision, compromised performance should always be investigated.	• Question specifications and optimizing question type functionality to work well on a tablet • The use of a stylus should be investigated, particularly for graphing.

Keyboard Input

Text entry is deserving of extended discussion, since adequate solutions may be more difficult to achieve than with other question interactions. Touch-screen keyboards do not allow students to rest their fingers on the keyboard without activating the keys, which makes it difficult for students to use keyboarding skills. While less of an issue for short-answer questions, this phenomenon is cause for concern with longer text-entry questions. For most students, typing speed and accuracy decrease and fatigue increases over longer periods of time, since it takes longer to convey their thoughts. The observations of students in this study confirm those of previous research that students tend to write less under these circumstances than they would with an external keyboard (Strain-Seymour, Craft, Davis, & Elbom, 2012). It should be noted, however, that many students (especially younger ones who have not yet perfected keyboarding skills) expressed that they preferred the onscreen keyboard and some even theorized that it might be faster for them than an external keyboard.

"You can type a letter on the keyboard and the screen shows what you typed in. The Tablet I screen was so big, it was awesome. You just tap the letter instead of pressing down the keys."

In addition to the basic text entry issues, onscreen keyboards represent a different set of challenges than external keyboards.

- First, the keyboards themselves are not ever-present and students must know how to open and close them. While most students in this study had no problems opening the onscreen keyboard, it was not always clear to them how to close the keyboard when they were done typing and ready to move on.
- Second, the onscreen keyboard takes up screen real estate and often pushed content off the screen, forcing students to scroll up to locate information they wanted to reference in answering the question.
- Third, not all keys are visible when the onscreen keyboard opens and students must know how to toggle between alpha and numeric keyboards. This is an issue both when numeric characters are the expected response and when students want to use specific punctuation in their text responses.

At a minimum, students should have an opportunity to work with tablets as part of their daily classroom activities to become familiar with the onscreen keyboards before tablets should be considered for use in a high-stakes assessment.

Conclusions and Next Steps

Regardless of any issues they encountered, students overwhelmingly reported that they would like to take a test on a tablet. They found the experience to be more interactive and hands-on, thought that the graphics were sharper and more clear than they might be if viewed on a computer, and thought it was "more fun" than testing on either computer or with paper-and-pencil. However, the study did reveal several usability issues which could potentially affect student performance, such that student excitement over tablet usage should not be cited as the sole evidence for tablet comparability. It is not known to what degree students' excitement might also have been a reaction to technology-enhanced question types, which might not be used with their current high-stakes tests. Additionally, the students' attraction to the tablet may resemble student survey results within computer-based testing (CBT), when it was relatively novel. Students reported preferring the computer to paper and remarked that it was easier and "more fun," even when they scored similarly or worse than with paper-and-pencil.

"I liked that the graphs were touchable. Everything was hands-on."

It is expected that optimization of the user interface and appropriate use of peripherals such as styluses and external keyboards with tablets will address the vast majority of issues which impact students testing on tablets. However, as a final research step before using tablets in high-stakes testing situations, states should consider to what degree any remaining challenges could potentially affect student scores. Pearson is planning to conduct large scale quantitative studies that will allow the statistical comparison of student performance across a traditional desktop/laptop condition and a tablet condition.

References

Holz, C. & Baudisch, P. (2010). *The Generalized Perceived Input Point Model and How to Double Touch Accuracy by Extracting Fingerprints*. Proc. CHI '10, 581-590.

Partnership for Assessment of Readiness for College and Careers (December, 2012). *Technology Guidelines for PARCC assessments: Version 2.0.* Retrieved on February 28, 2013 from: http://www.parcconline.org/sites/parcc/files/PARCCTechnologyGuidelinesDecember2012.pdf.

Smarter Balanced Assessment Consortium. (January, 2013). *The Smarter Balanced Technology Strategy Framework and System Requirements Specifications.* Retrieved on February 28, 2013 from: http://www.smarterbalanced.org/wordpress/wp-content/uploads/2013/01/Technology-Strategy-Framework_1-11-13.pdf.

Strain-Seymour, E., Craft, J., Davis, L.L, & Elbom, J. (October, 2012). *Moving assessments from desktop computers to tablets: An observational study.* Technical report for Virginia Department of Education.

Young, J. G., Trudeau, M., Odell, D., Marinelli, K., Dennerlein, J. T. (2012). *Touch-screen tablet user configurations and case-supported tilt affect head and neck flexion angles. Work* (41), 81-91. Retrieved from: http://iospress.metapress.com/content/x668002xv6211041/fulltext.pdf