Missouri End-of-Course Assessments

Cognitive Labs Report for Tablet Testing

Presented to the
Missouri Department of Elementary and Secondary Education
by
Questar Assessment, Inc.

August 22, 2013

5550 Upper 147th Street West
Apple Valley, MN 55124
(952) 997-2700
www.questarAI.com

Table of Contents

List of Tables

List of Figures

EXECUTIVE SUMMARY

In collaboration with Missouri's Department of Elementary and Secondary Education (DESE) and Missouri school districts, Questar Assessment, Inc. (Questar) conducted cognitive labs (cog labs) in May 2013 as part of tablet testing research and operational implementation. There were two main goals for the cog labs in the context of the Missouri End-of-Course (MO EOC) tablet research. The first goal was to gain insight into the cognitive processes that students are using to interact with the tablet. The second goal was to evaluate item rendering, item presentation, and usability of the tools to identify construct-irrelevant factors that may impact student responses to test questions.

Cog labs using Androids, iPads, and Chromebooks were conducted for Algebra I, Biology, and English II and included both performance events and selected-response items. Two accessories—the external keyboard and a stylus—were also experimented with.

A total of 73 students representing different achievement levels participated in the cog labs across five schools and the three types of devices. The schools were located in different parts of the state, with varying student populations. They represented a varied sample of districts. The cog lab protocol consisted of questions regarding students' use of the device and their experiences working through each item. The protocol also captured the researcher's observations of the levels of task certainty, frustration, confusion, and tool usage. Following the completion of the cog lab, students were surveyed about their experience and skills with the cog lab device, their use of the accessories, testing in the cog lab, and device preference for the MO EOC assessments.

The findings from the cog labs and surveys have implications for the continued refinement and deployment of EOC testing on new devices in Missouri, particularly for the performance events. Improving the testing interface is much needed for better tools and functions, including the graphing tools, the implementation of scrolling, and the ruler and protractor tools. Based on the findings, Questar recommends a flexible policy that allows students who take the MO EOC assessments to choose the device with which they have the most experience feel most comfortable, particularly for tests with performance events. Testing with Chromebooks presents a similar testing experience to a desktop or laptop. We recommend testing with iPads be deployed if the features that presented difficulty and challenges during the cog labs—including, but not limited to, scrolling, graphing, and ruler/protractor usage—are enhanced and demonstrate stable performance and ease of use in a field trial of the intended student population. Android devices should be introduced only when problems associated with connectivity, loading speed, compatibility, and tool functionality are more fully investigated and resolved in such a way that the experience of testing on the Android is reasonably comparable to testing on the iPad. Because many Android tablets exist, Questar recommends that their use be further screened and qualified for EOC testing.

Regarding accessories, the onscreen keyboard presented a greater source of confusion or difficulty for students than the external keyboard, which also posed a challenge. The stylus was not an accessory that most students used or were familiar with, even though a stylus could possibly provide greater precision for graphing on a tablet. Questar recommends allowing students testing on tablets to use external keyboards for performance events requiring keyboarding and to use a stylus for graphing before further research can be conducted to evaluate their effects on test performance.

INTRODUCTION

To provide student accessibility to testing using technology devices other than desktop computers and laptops, Missouri's Department of Elementary and Secondary Education (DESE) have decided to allow tablets (e.g., iPad and Androids) and Chromebooks to be used for the Missouri End-of-Course (MO EOC) assessments for selected-response (SR) items starting with the summer 2013 administration and for performance events starting with the fall 2013 administration. Questar Assessment, Inc. (Questar), in collaboration with DESE and Missouri school districts, conducted cognitive labs (cog labs) in May 2013 as part of tablet testing research and operational implementation. The focus was on the Phase I EOC tests, which include Algebra I, Biology, and English II. All of these tests consist of selected-response (SR) items and performance events (PEs).

In recent years, cog labs have increasingly been used for researching cognitive processes and finding solutions to assessment challenges. The procedure, also referred to as verbal protocol analysis or "think alouds," typically involves a two-step process (Ericsson & Simon, 1993). Initially, students are asked to think aloud while completing a task and the researcher records the information from the verbal report. At the completion of the task, the researcher may follow up with a series of questions about the task, including clarifying data from the "think alouds."

There were two main goals for the cog labs in the context of the MO EOC tablet research. The first goal was to gain insight into the cognitive processes that students are using to interact with the tablet. In the development of a user interface, assumptions are often made about the types of cognitive processes and the level of cognitive demand. The cog labs provide an opportunity to evaluate the accuracy of those assumptions before the system is deployed to a larger population. The second goal was to evaluate item rendering, item presentation, and the usability of the tools to identify construct-irrelevant factors that may impact student responses to test items. Such factors are not related to the knowledge, skills, and ability measured by the tests, but their existence may interfere with test performance.

Devices of interest included tablets (e.g., iPad and Androids) and Chromebooks. One feature that distinguishes a tablet from a desktop or laptop computer is that it has a touchscreen interface that involves pinching and swiping, with a built-in onscreen keyboard. A desktop or laptop, on the other hand, uses a physical keyboard with or without a mouse for navigation and functionalities. The Chromebook also has a thin deploying keyboard and a touch pad similar to laptop computers. However, they differ in that Chromebooks are designed to be used while connected to the Internet. They support applications that reside on the web rather than traditional applications (e.g., Microsoft Office) that reside on the machine itself, as supported by desktops and laptops. The Chromebooks were included in the cog labs as a proxy for laptop computers. In terms of testing, they present a very similar user experience to desktop or laptop computers.

These three types of devices also differ in the operating systems they use. iPads run on iOS, with Safari as the default Internet browser. Android tablets run on Android, with Google Chrome as the default Internet browser. The Chromebook uses Google Chrome OS as its operating system and Google Chrome as the default Internet browser. These devices are among the widely popular and available devices on the market; they have also been adopted by some of the Missouri high schools, based on the Missouri devices data provided by the state.

Cog labs for the three types of devices were conducted for each of the three content areas: Algebra I, Biology, and English II. Both PE and SR item types were included. In addition, students were asked to take a survey at the completion of the cog labs. Two accessories—the external keyboard and stylus—were also experimented with. A total of 73 students representing different achievement levels participated in the cog labs across five schools and the three types of tablet devices. The schools were located in different parts of the state, with varying student populations. They represented a varied sample of districts.

Data were collected in the cog lab and surveys. The cog lab protocol consisted of questions regarding students' use of the cog lab device and their experiences working through each item. The protocol also captured each interviewer's observations of participant's the levels of task certainty, frustration, confusion, and tool usage. Findings are presented by content area on the following: apparent level of certainty, apparent level of frustration, apparent sources of confusion (if any), attempts at using tools, student level of clarity with respect to using the device to respond to the item, student perceptions of the length of time it would take to respond to an item on the device versus online or in a paper-and-pencil format, and student perceptions of relative difficulty of the task in comparison to alternative formats. Following the cog lab, students were asked to take a survey about the device they used in the cog lab. Findings are presented on their general preference for technology devices, their self-assessed skills, their use of the onscreen keyboard and the stylus, their cog lab testing experience, and their device preference for taking the MO EOC test. Finally, recommendations are provided for successful implementation of the MO EOC assessments with PEs across different types of devices.

METHODOLOGY

Recruitment

Missouri school districts with any of the three types of devices – iPads, Androids, and Chromebooks – available in high schools were eligible to participate in the cog labs. A statewide e-mail preapproved by DESE was sent in early spring of 2013 to recruit school districts. The email also included a brief survey designed to qualify districts for participation. In total, nine districts responded. All of them were asked for follow-up information, including qualifications of participants. Based on schools' responses and availability, and in consultation with DESE, five schools participated in the cog labs: two using iPads, two for Androids, and one for Chromebooks. The schools were in different parts of the state, with varying student populations. They represented a varied sample of districts. The locations of the five schools, represented by A-E, are shown in Figure 1, with the black line delineating the state boundary.

Figure 1. The Five Cog Lab Sites[1].

[1] A = Android-ASUS
B = iPad
C = Chromebook
D = Android-HP
E = iPad

Participants

In each of the five participating schools, administrators nominated 15 students – five for each content area – to participate in the cog labs. Several criteria were used to select students. Eligible student participants needed to be currently enrolled in the relevant course. In addition, school administrators were asked to select students representing a range of ability in the content area(s) being assessed and to balance demographic variables—e.g., gender and race/ethnicity—in their selections.

Sample sizes of the cog lab participants are typically small because a researcher works directly with students one-on-one for some time that can vary from a few minutes to more than an hour. Nielson (1994) recommends five participants, a sample size sufficient for identifying needed information for solving most of the usability problems. To obtain valid information, Johnstone, Bottsford-Miller, and Thompson (2006) included 5–10 students per group in their evaluation of test design for students with disabilities and English language learners.

The final number of students who participated in the MO tablet cog labs was 10 for the iPad and 10 for the Android for each of the three content areas. For the Chromebook, two of the 15 students were unable to participate because of schedule conflicts, which resulted in five students for Biology, four for Algebra I, and four for English II. A total of 73 students representing different achievement levels participated in the cog labs across the five schools and three types of tablet devices. The number of students who participated in the cognitive labs is 24 for Algebra I, 25 for Biology, and 24 for English II. Table 1 presents the breakdowns by content area, tablet type, and gender.

Table 1. Cog Lab Participants by Content Area, Device Type, and Gender

Cog Lab		F	M	Total
Algebra I				
Android-ASUS		2	3	5
Android-HP		3	2	5
Chromebook		2	2	4
iPad - School 1		3	2	5
iPad - School 2		3	2	5
	Total	**13**	**11**	**24**
Biology				
Android-ASUS		2	3	5
Android-HP		3	2	5
Chromebook		2	3	5
iPad - School 1		2	3	5
iPad - School 2		2	3	5
	Total	**11**	**14**	**25**
English II				
Android-ASUS		2	3	5
Android-HP		2	3	5
Chromebook		1	3	4
iPad - School 1		3	2	5
iPad - School 2		3	2	5
	Total	**11**	**13**	**24**
Grand Total		**35**	**38**	**73**

The cog labs were scheduled to take place toward the end of the school year so that the students would have the opportunity to learn the course content and take the online test. Their recent testing experience using desktop or laptop computers would provide a basis for comparison during their evaluation of testing using the cog lab device.

Tablet Hardware Requirements

To ensure adequate support of testing using tablets and Chromebooks, minimum hardware and operating system requirements were developed and provided to the participating schools. For the iPads, iOS Version 6.0 or higher were required; for the Android tablets, Android Versions 4.0.1 (Ice Cream Sandwich), 4.1/4.2 (Jelly Bean) or higher were required. All versions of Chromebooks were allowed. In addition, screen sizes of 9.5" or larger were required for all devices. These requirements were consistent with the requirements of the two assessment consortia (PARCC, 2013; Smarter Balanced Assessment Consortium, 2013).

iOS is a proprietary mobile operating system developed by Apple Inc. to support Apple products. Android is also an operating system designed primarily for touchscreen mobile devices and is currently owned by Google. Unlike iOS, Android is open source with permissive licensing, which allows the software to be freely modified and distributed by device makers. This is why there is a large variety of Android tablets, such as ASUS, Samsung Galaxy, Google Nexus, and Microsoft Surface, to name a few popular brands. The Androids used in the cog labs also reflected this reality. Two types of Androids were used: the ASUS and the HP Touchpad. The HP Touchpad ran originally on the HP webOS and was later converted to Android. There are also different types of Chromebooks—e.g., Samsung Chromebook and Acer Chromebook. Samsung Chromebooks were used in the cog labs.

Procedure

The cog labs were conducted in computer labs or other available rooms in each of the participating schools. Three sessions (one in each of the content areas) were run concurrently. Stations were chosen such that there would be little to no interference between cog lab sessions—e.g., in separate corners of a computer lab. Each student interacted with an investigator who is also a subject matter expert (SME) of the content area in which the student was being assessed. SMEs provided each student-participant with a general introduction to the activity, demonstrated the think-aloud process with a modeling item, and worked from a script with a series of questions about each item. Students were encouraged to think aloud as they worked on each item. They were encouraged to articulate initial thoughts on viewing the item, strategies for answering the question or solving the problem, and approaches for interacting with the device to enter a response. The SMEs took notes on the students' verbalizations. An audio recording was also made of each student session. Some of the sessions were videotaped if parental permission had been granted.

Upon completion of each item, the student was asked a series of questions for information on other aspects of the device including tools, resources, and navigability of the testing interface. The SME recorded the student's responses to the questions about each item. After completion of the last item, students completed an online survey. All of this data, combined with student responses to the survey, was analyzed to make inferences regarding the cognitive processes used by students as they interact with the device.

Most of the cog lab sessions took between 45 and 60 minutes. The cog labs were conducted on one day for each school during May 6–16, 2013, for a total of five days.

In addition to the three types of devices, two accessories were also experimented with in the cog labs: the external keyboard and the stylus. Because the Chromebook already has a physical keyboard, the external keyboard was intended for the iPads and Androids to investigate the effects of its use in composing responses to the writing prompts in English II. Experimentation was conducted with the stylus to see if students thought it would assist with the graphing precision for some of the PEs in Algebra I and Biology. Every other student participating in the cog lab in these content areas on either an Android or an iPad used the stylus on items requiring precision touch. Similarly, on these devices, students alternated between virtual and external keyboarding on the items requiring key entry of information.

Cog Lab Protocol

The cog lab protocol consisted of questions regarding students' use of devices, particularly the tablets, and their experience working through the test items. The protocol also captured interviewer observations regarding task certainty, frustration, confusion, and tool usage. For each item, students were asked questions about item difficulty, tool usage, and their perception taking the item on paper-pencil or standard computer vs. on the cog lab device.

Student Survey

At the completion of each cog lab session, the student was asked to complete an online survey about their general preference for technology devices, their self-assessed skills, their use of the onscreen keyboard and the stylus, their cog lab testing experience, and their device preference for taking the MO EOC assessment.

Items

Items used in the cog labs were selected from a larger set of practice and released items across the three content areas that consist of SR items and PEs: Algebra I, Biology, and English II. The items represented different item types, formats, and features. Some PEs required a short answer in text, and others involved the use of graphing tools, the equation editor, or a data table builder. Writing prompts (WP) were included in English II. Descriptions of the items used in the cog labs are provided below by content area.

Algebra I

The Algebra I cog lab test consisted of eight items selected to address usability issues pertaining to the various item types and tools that are typical of the operational Algebra I EOC assessment. As shown in Table 2, the first six items were SRs and the last two were PEs. In general, the items required increasing levels of engagement with the tools and specialized functions that were the focus of the cog lab. The first two items were standard SRs, but the third required scrolling to view the entire item, the fourth featured a two-window display, the fifth might encourage the use of a calculator or scratch pad, and the sixth might encourage the use of a scratch pad for sketching a geometric scenario. The seventh item was the most complex: A PE consisting of four parts, it required scrolling, the filling in of tables, possible engagement with an equation editor's buttons, and the use of a graphing tool. The last item required engagement with the equation editor.

Table 2. Description of Algebra I Items

Item #	Item Type	Format	Features
1	SR	Choices arranged vertically	Horizontal stem graphic
2	SR	Choices in 2x2 block	Each choice is a graph
3	SR	Choices arranged vertically	Large scatter plot in stem, vertical scrolling required to view all choices
4	SR	Split Screen, choices arranged vertically	Large graph of linear function in left pane
5	SR	Choices arranged vertically	Five MathType expressions in stem and each choice
6	SR	Choices arranged vertically	Special symbol (%) in each choice
7	PE	Stimulus followed by four tasks	Table builder, equation editor, line graphing utility, vertical scrolling required
8	PE	Stimulus followed by 3 tasks	Equation editor with expressions and equation

Biology

The Biology cog lab test consisted of eight items selected to address usability issues pertaining to the various item types and tools that are typical of the operational Biology EOC assessment. As shown in Table 3, the first four items were SRs and the last four were PEs. In general, the items required increasing levels of engagement with the tools and specialized functions that were the focus of the cog lab. The first four items were standard SRs requiring an increasing amount of screen area to display, but no scrolling. The fifth item was a two-window PE requiring minor scrolling on the left window but none on the right (response) window. Students had to key in their answers to respond to Item 5. Items 6–8 contained the same right window display as Item 5 but required different tools and features. Item 6 required the use of the ruler, including rotation of the ruler. Item 7 required the use of several graphing and labeling tools, as well as both vertical and horizontal scrolling in order to respond completely to the item. Item 8 required the creation and labeling of a table in a response area where both vertical and horizontal scrolling was permissible.

Table 3. Description of Biology Items

Item #	Item Type	Format	Features
1	SR	Choices arranged vertically	Numerals as answer choices
2	SR	Choices arranged vertically	Special symbol ($\times$) in answer in each answer choice
3	SR	Choices arranged vertically	Bulleted list in stem, long answer choices, special symbol ($\rightarrow$) in each choice
4	SR	Choices arranged vertically	Graphic in stem
5	PE: constructed-response	Split Screen	Stimulus text and data table in left pane, large type in box with editing tools in right pane
6	PE: short answer	Split Screen	Stimulus text and data table in left pane, graphic and small type-in box with in right pane, ruler tool required
7	PE: graphing	Split Screen	Stimulus text and data table in left pane, graphing utility in right pane, vertical and horizontal scrolling required in right pane
8	PE: table builder	Split Screen	Stimulus text and data table in left pane, table builder utility in right pane, vertical and horizontal scrolling required in right pane

English II

The English II cog lab test consisted of four items selected to address usability issues pertaining to the various item types and tools that are typical of the operational English II EOC assessment. Items 1 and 2 were SRs and Items 3 and 4 were PEs. Table 4 presents the descriptions of the items. The first item was a standard SR not requiring the use of any special tools or scrolling. The second item was a two-window reading comprehension SR item requiring scrolling on the passage window. The third item was a two-window writing prompt PE requiring scrolling on the directions window and the creation of a business letter, via keyboarding, on the response window. The last item was a two-window creative writing prompt requiring the revision of a paragraph, via keyboarding, but no scrolling was required.

Table 4. Description of English II Items

Item #	Item Type	Format	Features
1	SR	Choices arranged vertically	Standalone item
2	SR	Split Screen	Passage in left pane, vertically scrolling required to view entire passage
3	WP	Split Screen	Student directions and checklist in left pane, vertically scrolling required to see entire checklist, large type-in box with editing tools in right pane
4	WP	Split Screen	Graphic and stimulus text in left pane, large type-in box with editing tools in right pane

ANALYSIS AND FINDINGS

Analysis of Cog Lab Protocol Data

Findings from the cog lab responses from both the students and the interviewers are presented by content area on the following variables: apparent level of certainty, apparent level of frustration, apparent sources of confusion (if any), attempts at using tools, student level of clarity with respect to using the device to respond to the item, student perceptions of the length of time it would take to respond to an item on the device vs. online or in a paper-and-pencil format, and student perceptions of relative difficulty of the task in comparison to alternative formats. The observation findings are presented first. Depending on the variable, results are then presented by device, and finally by item type (SR vs. PE). Special tool-related variables were captured for the PE items, and results for some of these will be highlighted.

Algebra I

Of the 24 students who took the Algebra I cog lab test, 10 used an iPad, another 10 used an Android cog lab, and four used a Chromebook. The test consisted of eight items. Because each item had 24 cases, there were a total of 192 observations on protocol questions common across items. However, for four of the students taking the cog lab on an Android tablet, Items 7 and 8 did not work on the school-supplied Android-HP devices, leaving eight fewer observations. Therefore, overall percentages reported are based on 184 observations.

Apparent level of certainty with the task. In 92% of the observations, students seemed certain of the task, while in the other 8% of observations, students either seemed uncertain or it was unclear to the interviewer. By device, the percentages are 90%, 91%, and 94% for the Android, Chromebook, and iPad, respectively. By item type, the percentages are 100% and 63% for SRs and PEs, respectively, showing a strong association between the level of certainty and task type.

Apparent level of frustration with the task. In 10% of the observations, students seemed frustrated with the task. In the other 90% of the observations, students either did not seem frustrated or it was unclear to the interviewer. By device, the percentages are 7%, 9%, and 14% for Android, Chromebook, and iPad, respectively. By item type, the percentages are <1% and 45% for SRs and PEs. Frustration appears to have been associated more with whether an item was a PE than with what type of device the student was using.

Apparent sources of confusion. In 178 observations, it could be identified whether or not there was a source of confusion for the student on the task. In 18 of those observations, there was at least one source of confusion. The most common source of confusion, mentioned in 11 observations, related to the use of the keyboard, internal or external. The internal keyboard was a greater source of confusion or difficulty for students than the external keyboard. For example, six out of seven students taking Item 6 with an internal keyboard found the keyboard a source of difficulty, whereas only two out of seven students taking it with an external keyboard found the keyboard a source of difficulty. The next common source of confusion, found in seven observations, related to the use of the equation editor. The equation editor was confusing to many students. There was mention of the equation editor insertion point not returning to a logical position and the superscript and subscript buttons not behaving as students would expect. One student could not properly enter a denominator under a numerator.

Attempts at using tools. In 181 observations, it could be ascertained whether or not an attempt or request to use a specific tool use was made. In five of those observations, a student requested a calculator. Also in five observations, a student attempted to use the notepad. In six observations, a student attempted to use the answer eliminator. In two observations, a student attempted to use the ruler.[2] In one observation, a student attempted to use the highlighter. In 30 observations, the student attempted some combination of using the equation editor, engaging with a table, or using the graphing tool. This makes sense given that the last two items required the use of some combination of these.

Student level of clarity regarding device use. On each task, students were asked if it was clear to them how to use the device to register their responses. Of the 184 observations, 158 students responded that it was "completely" clear how to use the device to mark their answer. Twenty-six reported that it was "mostly" clear or not clear. In 17%, 13%, and 13% of the Android, Chromebook, and iPad observations, respectively, students were not completely clear on how to use the device to mark their answers. Again, on this outcome measure, there was a large item type component: In only one out of 144 SR observations was a student not completely clear on how to use the device to mark their answer, in contrast with 25 out of 40 PE observations.

Student perception of relative task duration. Students were asked how long it took them to complete an item, relative to how long it would take them on computer or on paper. In 104 of the 182 responses, students said it would take "about the same" amount of time. Where there was a difference in perceived duration, it was in favor of another medium. In 47 observations students said it takes longer on the cog lab device, and in 31 observations students said it would take a shorter amount of time. The shorter vs. longer difference was associated with device type. For the 80 Android observations, students said the task would take longer on the Android in 18 cases and in seven cases they said it would take less time of the Android. For the 32 Chromebook observations, these numbers were five and seven, favoring the cog lab device. Finally, of the 80 iPad observations, 24 students said the task would take longer on the iPad and 17 said it would take a shorter time, revealing more polarization of opinion regarding relative task duration for iPad testers.

Students were more evenly split on perceived duration when it came to SR items. While in most of the 143 SR observations students did not think there would be a time difference, the remainder went approximately 50-50 longer vs. shorter. This was not the case with PE observations on this variable. Out of 39 observations, 24 indicated that the item would take longer on the device, with only five indicating shorter. Given this association of perceived relative task duration with item type, as well as the fact that eight Android observations on the PEs were lost because of technical problems, there is likely to be an even greater percentage of Android observations in which students would have perceived a task to take longer on the Android than on computer or paper.

Student perception of relative task difficulty. Students were asked to compare the ease of answering each question to what it would be like online or on paper-and-pencil. There were 184 responses to this question. Forty-seven percent of those responses indicated that the task would be equally as difficult on the device as online or on paper-and-pencil. In 28% of the observations, students felt the task would be easier, and 23% reported it would be more difficult on the device. Responses to this question were device-related. Of the 32 responses to this question for the Chromebook, 10 said that the task would be easier on that device, and only three indicated it would be more difficult. Out of the 72 responses for the Android, 14 indicated it would be easier on the device, and 16 said it would be

[2] Note: No item required the use of the ruler.

harder. And out of the 80 iPad observations on the question, 29 indicated it would be easier on the tablet and 24 indicated it would be harder. Responses were most strongly related to item type. Of the 144 SR observations on the question, 48 indicated the task would be easier on the device, with only 14 saying it would be harder. However, on the 40 PE observations, five indicated it would be easier, and 29 indicated that the PE task would be harder on the device.

Scrolling and Item 7. Item 7 was the most complex on the cog lab instrument, requiring the use of scrolling, filling in tables, engaging with the equation editor, and using a graphing tool. Most fundamentally, the item requires scrolling to read all the content and engage with all of the response areas. Of the 19 students who took this item and responded to a scrolling-related question about it, one of them said that he or she did not scroll to see the additional content on the item.

Other observation. Additional observations were recorded by the interviewer for 23 of the 24 protocol sessions. These observations list a number of issues that arose. Some are Internet access-related, others relate to problems with the device, and others are about the tools themselves. Some are interviewer notes to contextualize a particular student's responses or are suggestions for improvement of the testing experience.

Biology

Of the 25 students who took the Biology cog lab test, 10 used an iPad, another 10 used an Android, and five used a Chromebook. The test consisted of eight items. Since there were 25 cases per item, there were a total of 200 observations on protocol questions common across items. Therefore, the reported overall percentages are based on 200 observations unless otherwise noted.

Apparent level of certainty with the task. In 91% of observations, students seemed certain of the task while in the other 9% of observations, students either seemed uncertain or it was unclear to the interviewer. By device, the percentages are 91%, 88%, and 93% for Android, Chromebook, and iPad, respectively. By item type, the percentages are 92% and 90% for SRs and PEs, respectively. These results show no association between apparent level of certainty and either device or task type.

Apparent level of frustration with the task. In 6% of the observations, students seemed frustrated with the task. In the other 94% of observations, students either did not seem frustrated or it was unclear to the interviewer. By device, these percentages are 11%, 0%, and 3% for Android, Chromebook, and iPad, respectively. The apparent level of frustration was only observed in 11 % of the PE observations. Frustration appears to have been associated with whether an item was a PE and with whether a student was using something other than a Chromebook.

Apparent sources of confusion. Forty-eight observations reported at least one identifiable source of confusion. Mention of the ruler—e.g., activating it, moving it, and rotating it—occurred in 19 observations on sources of confusion. Frustration with the ruler was noted in several of these as well. The next most common source of confusion, mentioned in nine observations, related to scrolling.

Attempts at using tools. General tools for the SRs—e.g., the answer eliminator, the highlighter, and the notepad—were used by some students at one time or another on the first four items. The answer eliminator was used in 13% of the SR observations, the notepad was used in 5% of them, and the highlighter was used in 4%. The PE items called for specialized tools. Item 5 required the use of a keyboard. On this item, 10 students were asked to use the internal keyboard, while 15 used an external keyboard. The internal keyboard appeared to be a disadvantage (see Figure 2), as seven of the 10

students using it indicated that taking the item with an internal keyboard would entail a longer response time. Only two students out of the 15 taking Item 5 with an external keyboard cited the keyboard as a reason why the item would take longer than online or on paper- and-pencil. On Item 5, students participating in the cog lab with a Chromebook uniformly said that the item would be about as difficult on the device as online or in a paper-and-pencil format. Four out of the five Chromebook users also said the item would take about the same amount of time, with the other student saying it would take a shorter time on the Chromebook. No Chromebook user had difficulty using the keyboard.

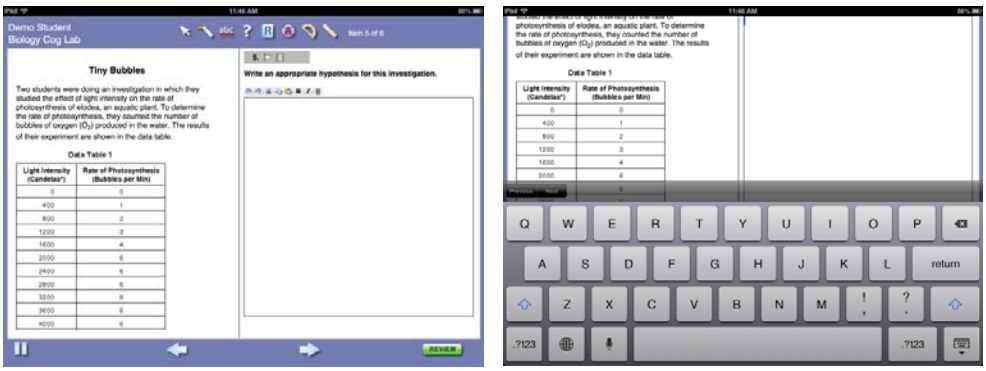

Figure 2. Biology Item 5 presented with and without the onscreen keyboard on the iPad.

Item 6 called for the use of the ruler (see Figure 3). Only two of the 25 students found the use of the ruler completely clear. Thirteen found it difficult to use, citing problems with rotating it. Chromebook users clearly benefited on this item, with all saying it was easier or of comparable difficulty to online or paper-and-pencil. Six of the 10 Android and three of the 10 iPad users said it was harder on the device. Twenty-one students responded to an ease of tool use question on Item 7, the item requiring the most tools and scrolling. Only five students found the tools “completely” easy to use. Ten students said the tools were not easy to use. Sources of tool use difficulty identified on Item 7 included the following: excessive scrolling, keyboard popping up, difficulty of drawing a straight line, and difficulty of placing labels in a precise location. “Scrolling” was the most often mentioned source of tool use difficulty on Item 7. Four of the 18 students who commented on ease of tool use for Item 8 had difficulty with the table-building tools. In three cases, these difficulties appeared to be connectivity-related or system-related: “Did not work” (two cases) and “Wouldn’t let her enter data anymore” (one case).

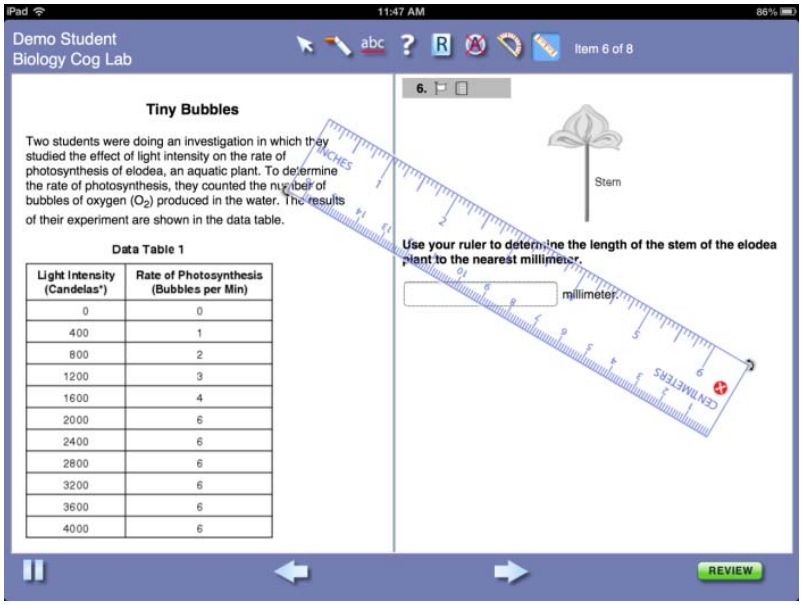

Figure 3. Biology Item 6 presented with a ruler on the iPad.

Student level of clarity regarding device use. On each task, students were asked if it was clear to them how to use the device to register their responses. Of the 195 observations on this question, 187 students responded that it was "completely" clear how to use the device to mark their answer. Only eight students reported that it was "mostly" clear or not clear. These eight responses were more or less uniformly distributed among devices, but all were associated with the PEs.

Student perception of relative task duration. Students were asked how long it took them to complete an item, relative to how long it would take them on computer or on paper. Of the 191 responses to this question, 85 of them reported it would take "about the same" amount of time. Where there was a difference in perceived duration, it was in favor of another medium: In 57 observations, students said it takes longer on the cog lab device, and in 49 observations they said it would take a shorter amount of time. The percentages of observations in which students said it would take about the same amount of time varied by device: 28% for Android, 71% for Chromebook, and 49% for iPad. SR items were perceived as the most comparable in terms of duration on a device versus online or paper-and-pencil. In 58% of SR observations, students said that working on the item would take about the same amount of time regardless of mode of administration. On the other 42% of SR observations, 90% of the students thought the item would take less time on the device. This was opposite for the PEs. In only 31% of PE observations did students report that responding to PEs would take about the same amount of time regardless of mode of administration. Seventy-seven percent of those who disagreed said that responding to the item would take longer on the cog lab device.

Student perception of relative task difficulty. Students were asked to compare the ease of answering each question to what it would be like online or on paper-and-pencil. Of the 189 responses to this question, 46% of them indicated that the task would be equally as difficult on the device as online or on paper-and-pencil. In 29% of the observations, students felt the task would be easier, and in 25% they said it would be more difficult on the device. Responses were strongly device-related. Of the 32 responses to this question for the Chromebook, 27 said the task would be of comparable difficulty, with the rest breaking about evenly between easier and more difficult. By contrast, out of the 77 responses for Android, only 22 indicated it would be of comparable difficulty. As with the Chromebook

observations on this question, the remainder broke evenly between easier and more difficult. Those taking the cog lab on the iPad provided observations on relative difficulty favoring the device: 48% indicated the task would be of comparable difficulty, but 30% indicated it would be easier on the iPad. By item type, perceptions of relative task difficulty followed the general pattern of relative task length. Fifty-four percent of the SR observations indicated comparable difficulty, while 42% indicated that the task was easier on the device. By contrast, only 38% of the PE observations for this question indicated that the task was of comparable difficulty regardless of mode of administration. The balance broke against the device, with 46% of PE observations indicating that the device made the task harder.

Other observations. Additional observations recorded by the interviewer for all 25 protocol sessions list a number of issues that arose. Some are Internet access-related, others relate to problems with the device, and others are about the tools. A number of these observations note student preferences with respect to tablet positioning (all preferring flat).

English II

Of the 24 students who took the English II cog lab test, 10 used an iPad, another 10 used an Android, and four used a Chromebook. This test had four items. Because each item had 24 cases, there are a total of 96 observations on protocol questions common across items. Unless otherwise stated, percentages are based on these 96 observations.

Apparent Level of Certainty with the Task. The apparent level of certainty with cog lab tasks was high among the students, regardless of the device they were using or the type of item with which they were engaged. In 97% of observations, students seemed certain of the task. By device, the percentages are 100%, 88%, and 98% for Android, Chromebook, and iPad, respectively. By item type, the percentages are 98% and 96% for SRs and PEs, respectively.

Apparent Level of Frustration with the Task. In 6% of the observations, students seemed frustrated with the task. In the other 94% of observations, students either did not seem frustrated or it was unclear to the interviewer. By device, the percentages are 3%, 0%, and 13% for Android, Chromebook, and iPad, respectively. By item type, the percentages are 2% and 10% for SRs and PEs, respectively. Although overall levels of frustration are low, these appear to have been weakly associated with iPad usage and PEs.

Apparent Sources of Confusion. In 39 observations, at least one source of confusion could be identified for a student on the task. Twenty-six of those observations indicated a keyboard-related source of confusion. Some students reported that the internal keyboard was slow, that it was obscuring the screen, and that it went inactive several times. The next common source of confusion, found in three observations, related to difficulties with scrolling.

Attempts at using tools. In the English II cog labs, the use of tools was largely prompted by the interviewer at Item 2. On Item 1, three of the 24 students used the answer eliminator, and nine attempted the select tool. On the PEs, each editing tool was used at one time or another by at least one of the students.

Student level of clarity regarding device use. On each task, students were asked if it was clear to them how to use the device to register their responses. In 85 of the 96 observations, students responded that it was “completely” clear on this. Nine responses reported that it was “mostly” clear, and only two reported that it was not clear. The breakdowns by device type do not reveal an association

between device type and level of clarity on device use. Among the Android, Chromebook, and iPad observations, there were five, one, and five observations, respectively, in which students were not completely clear on how to use the device to register their answer, but the total number of observations on Chromebooks was roughly 1/3 that of each of the other devices. By contrast, all the lack of clarity observations were observations on PEs, revealing that when students had difficulty understanding how to use the device when registering their answers, it was because they were working on one of the writing prompts.

Student perception of relative task duration. Students were asked how long it took them to complete an item, relative to how long it would take them on computer or on paper. Of the 95 responses to this question, 52 students said it would take "about the same" amount of time. Where there was a difference in perceived duration, it was split: In 22 observations students said it takes longer on the cog lab device, and in 21 observations they said it would take a shorter amount of time. Shorter vs. longer differences emerged for the Android and Chromebook. Slightly more than half of the 24 students taking the test on the Android said it would take the same amount of time regardless of mode of administration. For the Android, 11 of the 40 observations reported that the task would be longer on the Android than online or on paper-and-pencil, and in only five observations did students say it would take shorter. On the Chromebook, however, the split was in the other direction: one of 16 observations for longer, and eight for shorter. On the iPad, the shorter vs. longer difference was about evenly split.

Students perceived tasks on the Chromebook as requiring less time and those on the Android as requiring more time. In general, students perceived task duration on SRs to be comparable to that of online or on paper-and-pencil. However, 16 of the 48 observations perceived a difference, and it was in favor of the device: In four cases, the task was perceived as longer on the device, but in 12 it was perceived as shorter. This pattern was reversed on the PEs, where 27 of the 47 observations reported a relative task duration difference. In 18 of the those observations, the task was perceived as taking longer on the device, while in nine it was perceived as taking a shorter amount of time.

Student perception of relative task difficulty. Students were asked to compare the ease of answering each question to what it would be like online or on paper-and-pencil. This question had 93 responses. Fifty-three percent of those responses indicated that the task would be equally as difficult on the device as online or on paper-and-pencil. In 27% of the observations, students felt the task would be easier, and in 16%, they said it would be more difficult on the device. Responses were strongly device-related. Of the 15 responses to this question for the Chromebook, 11 said the task would be easier on that device. Out of the 40 responses for Android, five indicated it would be easier on the device, and eight said it would be harder. And on the 38 iPad observations on the question, nine indicated it would be easier on the tablet and seven indicated it would be harder. Responses were also strongly related to item type. Of the 47 SR observations on the question, 16 indicated the task would be easier on the device, with only two saying it would be harder. However, on the 46 PE observations, nine indicated it would be easier and 13 indicated that the PE task would be harder on the device.

Other observations. Additional observations were recorded by the interviewer for 23 of the 24 protocol sessions. These observations list a number of issues that arose. Some are Internet access-related, others relate to problems with the device, and others are about the tools. Some are interviewer notes to contextualize a particular student's responses or are suggestions for improvement of the testing experience.

Student Survey Results

At the completion of each cog lab session, the student was asked to complete an online survey about their general preference for technology devices, their self-assessed skills, their use of the onscreen keyboard and the stylus, their cog lab testing experience, and their device preference for taking the MO EOC test. Among the 73 students who participated in the cog labs, 70 took the survey. Chromebook students took only a subset of the questions relevant to their experience because the Chromebook has a built-in physical keyboard. The breakdown of survey respondents was 27 for the Android, 30 for the iPad, and 13 for the Chromebook, as shown in Table 5. The majority of the Android and iPad students (70% and 62%) came from Grade 10 while the majority of the Chromebook students (69%) were from Grade 11. Results of the Android and iPad users were combined in some of the analyses to make sense of data.

Table 5. Survey Responses by Device

Device	# of Survey Responses
Android	27
iPad	30
Chromebook	13
Total	**70**

General Preference for Technology Devices

Device preference and time of use. Three students, one from each device, said they did not like using desktop or laptop computers. When asked about if they liked using tablets, 20 Android students and 27 iPad students responding to the question said they did, while seven Android and three iPad students said they did not. On the other hand, all of the Chromebook students said they liked it. Twenty-one out of 26 Android users and 26 out of the 30 iPad users reported that they had used the device for a few months or more while the majority of the Chromebook students (9 out of 11) reported using it only for a few days at the time the cog labs took place. Twenty-three of the 27 Android users (or 85%) and 28 of the 30 iPad users (or 93%) use touchscreen devices such as smartphones or tablets on a regular basis.

Self-assessed skills. When asked to rate their skills in using the device, more of the Android and iPad users (59% and 67%) called themselves Intermediate users while more of the Chromebook students (55%) rated themselves as Beginner. The percentage of Advanced users was 11% for the Android, 17% for the iPad, and 9% for the Chromebook.

Tablet positioning. Students were asked the following question about positioning a tablet: "When you use a tablet, how do you like it positioned?" Two options—lying flat or propped up at an angle—were given. Survey results show that, across Android and iPads users, the number of students who liked each positioning was very similar (26 vs. 30). However, when self-rated onscreen typing skills were factored in, the results are different. Ten out of 13 students with Advanced typing skills kept it lying flat, while four out of the nine students with Beginner typing skills and 12 out of the 30 students with Intermediate skills preferred this position. While tablet position is in general a personal preference, it seems that the better a student's onscreen typing skills, the more likely the student kept it lying flat.

One student even preferred to position the tablet vertically so that he would be able to scroll with the onscreen keyboard up when working on the PEs.

Onscreen Keyboard and Stylus

Onscreen keyboard use and onscreen typing skills. Of the 57 iPad and Android students, 19 used the onscreen keyboard once a month and 19 used it every day. Seven students used it once a week and 12 used it two or three times a week. Among the 19 students who used the onscreen keyboard once a month, five rated their typing skills as Beginner. Among the 19 students who reported typing on it every day, 18 rated having Intermediate or Advanced onscreen typing skills. The onscreen typing skills of these 23 students seem to indicate that typing skills were associated with the frequency of the keyboard use.

Onscreen keyboard preference in light of typing skills. Table 6 presents students' like or dislike of the onscreen keyboard and their self-rated onscreen typing skills. All of the students who rated themselves as Beginner (12 out of 56) did not like typing on the onscreen keyboard. Students with Intermediate onscreen typing skills were split between "Yes" and "No." Eight of the 13 students with Advanced typing skills liked the onscreen keyboard. Students liked it because the onscreen keyboard was simple, convenient, easy to use, and fun. All of the students who liked the onscreen keyboard rated their onscreen typing skills as Intermediate or Advanced. Not surprisingly, interest in onscreen typing was strongly associated better onscreen typing skills.

Table 6. Onscreen Keyboard Typing Skills and Preferences

Onscreen Keyboard Typing Skills	**Like Onscreen Keyboard or Not?**		**Total**
	Yes	**No**	
Beginner	--	12	12
Intermediate	16	15	31
Advanced	8	5	13
Total	**24**	**32**	**56**

Stylus. Among the 55 Android and iPad students, 11 of them reported using a stylus with the tablet and three of them used it the first time during the cog lab. Another three used it all the time. One student used it mostly for playing games, and two others used it whenever it was available. Nine of the 11 students who used a stylus said they liked it, and they liked it for the reasons that it is easier to "tap," "point," "control," "touch the right spot all the time," "take a lot less time," and "(help) click on smaller icons when I had trouble trying to press it with my fingers." Some students also noted that it was more precise. Thirty-two of the 34 students who did not like using a stylus never used it before. They liked using their hands and fingers, as it was easier and more natural. Some thought it would be harder to use a stylus to tap on the screen and one student thought it was too complicated.

There were additional comments on stylus use in the cog lab experience that students shared. One student said that "using the tablet for drawing on a graph was difficult without a stylus." Another student pointed out that more practice is needed in order to continue testing on the tablet. One student who used a stylus found it was hard to use for certain graphing points, as it "kept freezing or not letting

me type." Another student noted that a stylus was needed "for the precise clicking." One student also noted that some styli may be too thick to "pinpoint what you wanted."

Testing in the Cog Lab

When asked about the ease of using the tools, buttons, and actions on the tablet, the majority of the Android and iPad students responded "easy" or "very easy," with others reporting "difficult" or "very difficult" for some of the tools. Notably, the ruler/protractor was difficult or very difficult to use for 15 out of 54 students (13 of 54 did not use it). Seventeen of the 54 students said using the Bookmark tool was difficult, and 20 out of 57 students found it hard to scroll to see a passage or question. Three of the students found it was hard to use the highlighter. Two students using the Chromebooks reported it was difficult to use highlighter, and another two students reported it was difficult to use the ruler/protractor. Other than these, students thought the tools were easy or very easy to use.

When asked "how easy was it to take a test on the tablet during the cog lab," six of the 25 Android users and seven of the 30 iPad students responded it was difficult, with 42 out of a total of 55 Android and iPad students reporting that it was easy or very easy. Students who found using the tablet easy or very easy noted that the tablet was convenient and faster. One student who participated in the iPad English II cog lab liked the iPad's interface better than a computer's interface and thought it was easy to select an answer (to SR questions) and type the responses (to WPs). The student concluded that the iPad with an external keyboard was superior to the computer. Another student noted that it was easier to click answers and draw lines on the tablet.

For those students who found tablet testing difficult, the main problems were associated with technical issues from the tools, such as graphing, scrolling, and typing, and from the sluggish responses of the tablets. Some of the students who responded "easy" also brought up the problems they encountered and pointed out improvements to be made. One prominent issue was with the graphing tool. Regardless of what a student's judgments were about taking the test on a tablet in the cog labs, 13 out of 54 students who shared their experiences reported the difficulty with the graphing item and/or expected improvements, including placing the scale, numbers, and titles in the appropriate position, lack of accuracy and precision when graphing, difficulty in plotting a line straight when using the free drawing tool, erratic responses of scrolling, and buttons that were too small. Note that there was one graphing item in Algebra I and one in Biology that required graphing the data presented in a table. One student thought the difficulty with the graphing tool was "due to general sluggishness."

> *"I've noticed that many of the tools seem to be very sluggish. For example, rotating the ruler tool was an exercise in futility, and I ended up getting it as straight as I could and estimating. The graphing tool was very difficult to use due to general sluggishness. It was impossible to do anything with accuracy, and my drawn line was not straight. The scrolling throughout the test was very erratic, and much of the interface didn't seem to really be designed with touch in mind, instead being a straight port of the computer controls. However, for the multiple choice questions, everything went very smoothly and they were very easy to complete."*

The onscreen keyboard also presented difficulties for the graphing item. Some of the PEs required typed responses. One Algebra student who used the Android-HP noted that the onscreen keyboard kept popping up in places that made it more difficult to position the text. Some of the problems noted with the tablet also existed with the desktop/laptop computers. One student wrote that the problems with the tablet were "very similar to computer, difficult to scroll sometimes."

In comparison, when asked "how easy was it to take the test on a Chromebook during the cog lab," all of the 11 Chromebook students who responded said it was "easy" or "very easy." They thought that taking a test on the Chromebook was fairly simple and straightforward like a laptop. Some students also reported problems with the tools, mostly with the graphing item, such as drawing the line and placing it along with the scales and the text in the appropriate place. One student thought that the graphs and tables would be easier on paper and another would rather take the graphing item on the computer.

Device Preferences for the MO EOC Assessments

About half of the Android and iPad students who liked the tablet (23 out of 47) also liked the onscreen keyboard. When asked about their keyboard preference in taking a test on a tablet, 16 (or 70%) out of these 23 students who liked the onscreen keyboard chose it while the other seven (or 30%) preferred the external keyboard. When asked about their device preference in taking the MO EOC assessment (see Table 7), five (or 31%) of the 16 students who chose the onscreen keyboard preferred to use the desktop/laptop. Seven of them preferred the tablet and the other four did not have any preference.

Table 7. Keyboard and Device Preferences for Taking the MO EOC Assessment for Students who Liked the Tablet and Onscreen Keyboard

Like Tablets?	Like Onscreen Keyboard?	Keyboard Preference	Device Preference when Taking the MO EOC Test			Total
			Desktop/Laptop	Tablet	No Preference	
Yes	Yes	Onscreen	5	7	4	16
		External	2	3	2	7
		Total	**7**	**10**	**6**	**23**

In total, 29 of the 57 Android and iPad users went for a desktop or laptop, 17 for a tablet, and 11 having no preference. Students liked the desktops and laptops because it was easier to use some of the test functions and, specifically, to graph with no issues with many tools. And they were familiar with laptops. One student noted the computer was easy because typing on the keyboard on the screen was "a little bit difficult." Another student who preferred the external keyboard when taking a test on the tablet and, consequently, the desktop/laptop when taking the MO EOC assessment provided the following explanation:

> *"As of now, the computer is the best performing platform. The tools all work very well, and everything goes well. If the problems with the tools on the tablet were resolved, it would be preferred. It certainly holds the most promise for raw convenience and simplicity, but, as of now, it's simply too frustrating to work through the problems."*

Apparently, these comments were associated with the student experience of using the tablet during the cognitive lab. The same student who made the above comment went on to say the following when asked to share the experience of using a tablet to take a test:

> *"I noticed that the test was being given on an HP TouchPad running CyanogenMod. I really didn't expect this to be particularly stable, and it wasn't. I'm curious if the test would run better on a native Android tablet, running an official version of Android. I also think that the tablet version of Google's Chrome browser would perform much better."*

The student realized that the performance of the testing system may depend on the devices used.

Nearly all the students who did not like the tablet (9 of 10) did not like the onscreen keyboard. Not surprisingly, these nine students preferred to use the external keyboard when taking a test on a tablet. When given the options of taking MO EOC assessments on a desktop/laptop computer or a tablet, seven of these students chose the former, with one having no preference and one choosing the latter. Students chose desktops or laptops because they were easier, faster, more responsive, and better for typing, clicking, and scrolling. And they are what the students are used to. Other reasons include "easier with a mouse, more precise," and "easier to move around on the screen without 'accidentally' hitting the wrong thing." An iPad student in the English II cog lab explained in detail why computers should be used for testing:

> *"Taking the test on the tablet was a little difficult because it took some time to get used to. I wouldn't recommend using tablets for testing because most students don't have a lot of experience with tablets and it would be a frustrating process for them. Also, most of the tools in the writing section didn't work and you couldn't use the tab key on the tablet, which you need to use often in writing. I found it difficult to type with the keyboard on the tablet because I am not used to typing on the device and the keys are very close together. Also, the auto correct kept popping up, which would need to be disabled for fair testing. Another issue is that you have to go to the other keyboard to get common punctuation marks. Reading on the tablet was okay but not ideal. I like reading sections on the computer better because you can see more text at one time and it is easier to scroll. It was kind of tricky to scroll and some students might not bother to read the rest of the passage if they cannot scroll. Ultimately, I think that testing should just be done on computers because tablets aren't as easy to navigate for most people."*

On the other hand, the one student who chose the tablet for EOC testing provided the following explanation:

> *"I would rather take it on a tablet because it is easier for me to stay in the test and it would show a better understanding on the graphs and stuff like that. And it would help kids in the future because everything is going to technology like this."*

In comparison, when asked about their device preference for taking the MO EOC assessment, six of the 11 Chromebook students chose the Chromebook while the other five did not have any preference. In the mind of these students, the Chromebook was either a better testing device or it can be used interchangeably with a desktop or laptop. Students preferred it for reasons that they "wouldn't have to go all the way to the library to take our tests on regular computers", "... this will prevent the students from having to leave their classes…less time wasted, and less noise in the hallway," and a Chromebook was "smaller, more personal." In spite of these, one student pointed out without hesitation that the Chromebook "could use some improvement" while another "would like the graphing part on another computer though."

SUMMARY AND RECOMMENDATIONS

Summary

Overall, students were relatively certain about the tasks in the cog lab tests. There was an association between task certainty and item type for Algebra I, but not in the other content areas. In no content area was certainty with the task associated with the type of device on which students took the task. Across content areas, there was only a small amount of frustration with the cog lab tasks. Higher levels of frustration were in all cases associated with item type. Students also tended to be more frustrated with the tablet devices (Android and iPad). Use of the internal keyboard and scrolling was a source of confusion across all content areas. The internal keyboard was a greater source of confusion or difficulty for students than the external keyboard. The use of the onscreen ruler was a major source of confusion for Algebra I students, and the equation editor and graphing tools were uniformly confusing.

Reported level of clarity with using the device to mark answers was always associated with task type, not with the device. However, perceptions of relative task duration and relative task difficulty were generally related to both device and task type. Performance events were generally perceived to take longer and to be more difficult on a cog lab device. However, when students were taking an item on a tablet (Android or iPad) and they felt a task would not be of equal duration or difficulty on the device in comparison to online or on paper, they tended to perceive that the task took longer on the tablet and/or was more difficult. This finding did not emerge for the Chromebook, where students either felt that the task was of comparable duration and difficulty, and when they disagreed, their responses tended to favor the Chromebook.

Student survey results indicate that most of the Android and iPad students liked the device while all of the Chromebook users liked Chromebooks. The majority of the Android and iPad students reported using the device for a few months or more, and the majority of the Chromebook students reported using it only for a few days when cog labs were conducted. Students vary considerably in terms of their experience with the device and accessories, their self-assessed skills, and their preferences. Students' like or dislike of the onscreen keyboard was associated with the level of tablet skills they classified themselves into. None of the students who rated themselves as "Beginner" liked typing on the screen while more of the "Advanced" users liked the onscreen keyboard. Onscreen typing skills also seem to be related to students' attitudes toward the onscreen keyboard, as all of the students who liked the onscreen keyboard had "Intermediate" or "Advanced" typing skills. The stylus is an accessory that most of the students in the cog lab did not use before, and, for that reason, nearly all of these students did not like it. On the other hand, the few who used it actually liked it. One student also noted that the stylus would provide more precision with graphing an item.

In using a tablet, slightly more than half of the Android and iPad students preferred keeping it lying flat to propping it up at an angle. However, most of the students with "Advanced" typing skills preferred to keep it lying flat. While tablet position is in general a personal preference, the better a student's onscreen typing skills, the more likely the student kept it lying flat.

Taking a test on the tablet during the cog lab was reported to be easy by most of the Android or iPad students and by all of the Chromebook students. Regardless of the device, it did present difficulties, particularly for the Android and the iPad. Most of the reported difficulties of taking the cog lab test on the tablet were technical problems associated with the tools needed for the PEs and the related tools

and functions, such as scrolling, graphing, and use of the ruler. When asked about their device preference in taking the MO EOC assessment, half of the Android and iPad users went for a desktop or laptop, with a portion (or 20%) having no preference. Students liked the desktops and laptops because it was easier to use some of the test functions and, specifically, to graph. In comparison, when asked about their device preference for taking the MO EOC assessment, half of the Chromebook students chose the Chromebook while the other half did not have any preference. In the mind of these students, the Chromebook was either a better testing device or could be used at least interchangeably with a desktop or laptop. Even for some of the Chromebook students, graphing was not the type of item one would like to work on using a Chromebook.

The difficulties students experienced in the cog labs – be it frustration with the tasks or challenges in using the tools – pose threats to validity of test scores. They should be addressed before operational testing with PE items can be implemented with these devices.

Recommendations

1. *Improvement of the tools* – The findings of the cog lab research indicate the continued need to improve the testing interface for better tools and functions, including the graphing functionality, the implementation of scrolling, and the functionality of the ruler and protractor tools. Graphing tools should be improved for adding text and drawing lines. A scroll bar or some other indicator of hidden content should be introduced. The way in which scrolling is implemented should be revised so that scrolling activation is more in line with scrolling conventions on tablets. Small buttons on the editing toolbars, such as bold, italic, and underline buttons, should be eliminated if they are not necessary.

   As an update, since the completion of the cog lab and the field trial of the SR items, Questar has been studying student feedback and designing technology solutions to improve the performance of the tools. A more intuitive solution has been found and implemented which minimizes student confusion and frustration with these tools. A scrolling solution is being investigated that will provide ease with scrolling yet maintain content presentation as similar as possible to the online version of the test. Once these tools are revised to address students' concerns, a field trial will be needed to target the PEs and testing of the tools before these devices are adopted for testing.

2. *Keyboard* – Students vary considerably in terms of their preference, experience, and skills using Androids and iPads and the accessories. The onscreen keyboard is a greater source of confusion or difficulty for students than the external keyboard, which also presented challenges. During the cog lab setup, problems arose with connecting the external keyboard. An external keyboard connects with the device via Bluetooth, which allows for data exchange wirelessly over short distances. Because of this wireless feature, the external keyboard does not need to be attached to the device and can be placed away from it. Once the Bluetooth is turned on, it starts searching for the keyboard to connect with the device. As multiple setups were going on at the same time, one for each device, Bluetooth had trouble locating the right keyboard to connect with the right device. Based on the findings and experience from the cog lab about the onscreen keyboard and the external keyboard, Questar recommends allowing students taking tests on tablets to use external keyboards for PEs requiring keyboarding. Questar also recommends advising test- takers that in taking tests with PEs, they should use an external keyboard unless they are fully fluent and comfortable with internal keyboards and know how to bring them up and retract them. School IT

professionals should have sufficient time to set up the device with the external keyboard appropriately before testing begins.

3. *Stylus* – Most of the students had never used a stylus before. While a stylus provides precision with clicking and graphing of Algebra I and Biology PEs, the cog lab did not find, nor was it designed to find, to what extent the use of a stylus was related to test performance. And no studies have been published that evaluate its effects either. Questar recommends allowing students to use a stylus in place of a finger if they wish to and if they think it helps them to demonstrate the knowledge and skills they already possess. These students should be encouraged to do practice tests with a tablet and a stylus. Should such a stylus be used, it should be one designed for precision purposes.

4. *Tablet positioning* – How to position an Android or an iPad for test taking is essentially a personal choice. Students can practice with the two different types of positioning—i.e., lying it flat or propping it up at an angle—and find the way that is conducive for them to take the test.

5. *Devices for MO EOC testing* – Questar recommends a flexible policy that allows students to be able to choose the device that they have most experience with and feel comfortable with when taking the MO EOC assessments, particularly for tests with PEs. Testing with Chromebooks presents a similar testing experience to a desktop or laptop. We recommend testing with iPads be deployed if the tools which presented difficulty and challenges during the cog lab, including but not limited to scrolling, graphing, and ruler/protractor, be enhanced and demonstrate stable performance and ease of use in a field trial of the intended student population and only for students who use these devices regularly and consistently for a variety of school-related purposes. Android devices should be introduced only when problems associated with connectivity, loading speed, compatibility, and tool functionality are more fully investigated and resolved in such a way that the experience of testing on the Android is reasonably comparable to testing on the iPad. Because many Android tablets exist, Questar recommends that their use be further screened and qualified for the EOC testing.

REFERENCES

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. (rev. ed.). Cambridge, MA: The MIT Press.

Johnstone, C. J., Bottsford-Miller, N. A., & Thompson, S. J. (2006). *Using the think aloud method (cognitive labs) to evaulate test design for students with disabilities and English language learners (Technical Report 44*). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved February 11, 2013, from the World Wide Web: http://education.umn.edu/NCEO/OnlinePubs/Tech44/

Nielson, J. (1994). Estimating the number of subjects needed for a thinking aloud Test. *International Journal of Human-Computer Studies, 41*(3), 385–397.

PARCC (2013). *Technology guidelines for PARCC assessments (Version 2.1)*. Retrieved March 25, 2013 from http://www.parcconline.org/sites/parcc/files/PARCCTechnologyGuidelines2dot1_Feb2013Update.pdf

Smarter Balanced Assessment Consortium (2013). *The Smarter Balanced technology strategy framework and system requirements specifications (executive summary)*. Retrieved March 25, 2013 from http://www.smarterbalanced.org/wordpress/wp-content/uploads/2011/12/Technology-Strategy-Framework-Executive-Summary_2-6-13.pdf