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

White Paper

Ellen Strain-Seymour, Ph.D.
Jason Craft, Ph.D.
Laurie Laughlin Davis, Ph.D.
Jonathan Elbom

July 2013

ALWAYS LEARNING

PEARSON

About Pearson
Everything we do at Pearson grows out of a clear mission: to help people make progress in their lives through personalized and connected learning solutions that are accessible, affordable, and that achieve results, focusing on college-and-career readiness, digital learning, educator effectiveness, and research for innovation and efficacy. Through our experience and expertise, investment in pioneering technologies, and promotion of collaboration throughout the education landscape, we continue to set the standard for leadership in education.For more information about Pearson, visit http://www.pearson.com/.

About Pearson's Research & Innovation Network
Our network **mission** is to spark innovation and create, connect, and communicate research and development that drives more effective learning. Our **vision** is students and educators learning in new ways so they can progress faster in a digital world. Pearson's research papers share our experts' perspectives with educators, researchers, policy makers and other stakeholders. Pearson's research publications may be obtained at: http://researchnetwork.pearson.com/.

Abstract

Tablets' affordability and the intuitiveness of manipulating on-screen objects directly via touch-screen make a compelling case for the use of iPads and Android tablets in the classroom. However, in the same way that comparability studies are used to investigate fairness when a high stakes test is delivered both online and in print, cross-device comparability studies should be used to inform state policies around the acceptable range of devices used for high-stakes testing. As a precursor to further research to inform policy, a study was conducted in Spring 2012 to observe primary and secondary school students' interaction with assessment materials on touch-screen tablets, including essay-writing tasks accessed with and without an external keyboard. The goal was to understand how the tablet might provide intuitive access to test materials and ease of use for capturing student responses as well as any challenges presented by the devices. Of interest was students' interaction with the physical or hardware aspects of the tablets as well as with the user interface provided by the computer-based testing software.

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

An analysis of any performance-impacting differences between test-takers' interaction with touch-screen tablets and with computers will be informed by two areas of study: comparability studies and human-computer interaction (HCI). While comparability studies address issues of fairness and validity in the interpretation of assessment results across diverse testing conditions, HCI addresses assessment contexts less specifically but contributes analytical methods for understanding how individuals interact with input devices in order to perform a range of tasks related to on-screen content. Comparability studies use pre-existing assessment instruments and rigorous data analysis methods to compare student outcomes across different test administration conditions (Bugbee, 1996). The detection of mode effects – test-taker performance differences across administration conditions – suggest the presence of confounding variables that interfere with accurately correlating test scores with knowledge of the constructs being assessed. HCI, on the other hand, uses a variety of qualitative and quantitative methods – observation, surveys, interviews, eye-tracking, screen/mouse action logging, measures of speed/accuracy, and post-task assessment – to understand some of those confounding variables when they stem from the ways humans interact with an electronic device (Hartson, 1998).

Cross-device Comparability

The need for comparability when an assessment is delivered via both paper and computer is addressed by a number of professional bodies such as the American Psychological Association and is mandated by the U.S. Department of Education as a component of the No Child Left Behind peer review process (APA, 1986; AERA, APA, NCME, 1999, Standard 4.10). Comparability research over the last quarter century has largely focused on differences between print and computer-delivered assessment, while less research has mined the wider implications of Randy Bennett's definition of comparability as

"the commonality of score meaning across testing conditions including delivery modes, computer platforms, and scoring presentation" (Bennett, 2003). However, with a shift towards online testing coinciding with the proliferation of devices appearing in the classroom, the sub-genre of cross-device comparability studies can be expected to flourish over the next few years.

Although no rigorous comparability research to date has focused on touch-screen usage for test-taking, foundations have been laid for cross-device comparability studies in areas such as screen size and keyboard differences – factors likely to vary across devices. One of the most wide-ranging studies done to date was conducted by Bridgeman, Lennon, and Jackenthal (2001) looking at screen size, resolution, internet connection speed, operating system settings, and browser settings for the SAT®. While the survey results showed that some test-takers noted frustration around small screen sizes and latency or wait times caused by slow internet speeds, the most critical factor was the amount of information available on screen without scrolling. While math scores appeared to be unaffected, lower scores were observed in verbal skills when smaller screen resolutions led to a lower percentage of the reading materials being visible at one time. A 2010 study by Keng, Kong, and Bleil (2011) kept the amount of information shown on screen, the screen resolution, and the amount of scrolling constant across test conditions but varied screen sizes. The results showed no difference in student performance across test-takers using netbooks with screen sizes of either 10.1 or 11.6 inches and students using the 14- to 21-inch screens common on desktop and laptop computers. These two studies suggest that the amount of information available on screen at one time seems to have more potential for negative impact than small shifts in the size of information on screen, assuming that content is presented at a large enough size for basic legibility.

Comparability studies comparing writing on laptops and desktops have also approached this idea of the impact of a smaller screen size, although these studies also probe another aspect of device difference: the keyboard. A study of Graduate Record Exam

results for test-takers using laptops and desktops was not specifically focused on writing, but found that essay writing was the only area with performance differences between the two conditions (Powers & Potenza, 1996). Since all test-takers used both devices for some portion of the test, it was possible to collect survey feedback comparing the two experiences: 48% of test-takers said that it was easier to take the test on a desktop computer, while 15% felt that it was easier on a laptop (36% responded that the two experiences were roughly equivalent). Survey respondents reported issues regarding screen size, keyboard size, the position of keys on the keyboard, and the feel of the keyboard. This study was conducted in the mid-1990s when experience with laptops was less pervasive; 94% of test-takers reported routine use of desktop computers while only 21% reported routine use of laptops.

Device/model familiarity may have played a role in a similar study of the National Assessment of Educational Progress (NAEP) assessment in which eighth-grade students used either desktop computers owned by their schools or laptops brought to schools by NAEP test administrators (Horkay et al., 2006). For one of two essays, test-takers using laptops scored significantly lower than students using school computers. When this study was widened to a larger sample size, results differed. No differences were apparent within the aggregated scores, but female students performed significantly lower on the NAEP laptops on both essays. The design of the study does not allow the performance differences to be attributed to aspects of the laptop design versus to student familiarity with particular computer models used regularly within school computer labs, but it does point to the need for greater understanding of device differences.

HCI: Screen Size Effects

Since differing amounts of on-screen information due to screen size seem to be a critical factor within these comparability studies, it is worth turning to the field of HCI for further examination of this issue as it relates to reading and writing tasks. A study by De

Bruijn, De Mul, and Van Oostendorp (1992) showed that subjects using a 15-inch screen to consume an extended textual description needed less learning time than subjects using a 12-inch screen when the amount of information on a single screen was less for the smaller screen size. However, both groups of subjects scored similarly on a follow-up comprehension assessment. The findings of Dillon, Richardson, and McKnight (1990) were not definitive but did include data trends to suggest higher levels of comprehension when college students saw more lines of journal article text at once using larger screens. Study results around reading comprehension seem to be more conclusive with more radical differences in screen size; a study of 50 participants reading Web site privacy policies demonstrated that the comprehension levels of desktop users, as evinced by scores on a Cloze test, were more than double the comprehension scores of readers accessing the privacy policy via an iPhone-sized mobile device (Singh, Sumeeth, & Miller, 2011). Other studies have measured productivity for scanning and searching large quantities of text rather than reading comprehension and noted improvements with larger monitors (Simmons & Manahan, 1999; Simmons, 2001; Kingery & Furuta, 1997).

The field of HCI's investigation into the effect of screen size on writing tasks has tended to focus on college-level or adult subjects engaged in academic writing or job-related text editing. A 1991 study by Van Waes (translated into English in 2003) observed that academic writers using a larger screen that made more text visible at one time were observed making revisions at a greater distance from the last point of insertion. In other words, extending the amount of written text in view directly affected the range of the writer's revisions, as the writing process typically involves scanning and re-reading for revision and further planning (Van Waes & Shellens, 2003; Flower et al, 1986; Van Oostendorp & De Mul, 1996). A more recent study conducted at University of Utah but commissioned by monitor manufacturer NEC involved 96 participants given text-editing tasks and randomly assigned a range of dual and single monitor set-ups using 18", 20", 24" and 26" screens. Time and editing performance measurements showed monitor size was a

more significant determinant of speed and accuracy than single versus double monitor

configurations (University of Utah, 2008).

Device Differences

As we move from a general look at different devices and varying screen sizes to the

specifics of a tablet, it is worth pausing to provide a definition of tablets and describe some

of their unique attributes. As a mobile computer, a tablet is usually larger than a mobile

phone but differs from a laptop or desktop computer in that it has a flat touch-screen and is

operable without peripherals like a mouse or keyboard. The boundaries between tablets and

computers sometimes blur with quick-start computers and with hybrids and convertible

computers, which combine touch-screens with keyboards that can be removed, swiveled, or

tucked away. Tablets were once best differentiated from computers not by size or screen

but by their mobile operating systems: iOS, Android, BlackBerry Tablet OS, webOS, etc.

However, software manufacturers are now presenting operating systems, such as Windows

8, as tablet- and computer-ready. In another instance of blurring boundaries, some e-

readers are becoming increasingly indifferentiable from tablets in their use of mobile

operating systems, similar size and shape, color touch-screen, long battery life, wi-fi

connectivity, and support of downloadable "apps." However, e-reader screens, unlike tablet

LCD screens, are optimized for reading even under bright light conditions, while tablets tend

to be designed with more memory and storage space for supporting multiple media forms

and a wider range of applications.

The key differences between tablets and computers that are deserving of further

analysis within cross-device comparability studies include physical size, ergonomics, screen

size, touch-screen input, and keyboard functioning. In regards to size, most tablets weigh 1

to 2 pounds and are designed to be handheld, used in a flat position, placed in a docking

station, or held upright by a foldable case. A tablet has no singular correct position, which is

reinforced by re-orientation of the on-screen image to portrait or landscape based on the

position of the device. Although the most typical tablet screen size of 10 inches was popularized by Apple's iPad, screen sizes of 5 and 7 inches are not unheard of. For instance, the 7-inch Samsung Galaxy Tab, with about half of the surface area of a 10-inch device, resembles the size of a paperback book. The 5-inch Dell Streak can fit in a pocket and resembles a large smart phone.

These relatively diminutive sizes and weights, the variability in mounting strategies, and most tablets' support for different orientations lead to more fluid interactions between a user's posture, lines of sight, and the device's angle and distance from the user. The popular media abound with references to these very different and more variable ergonomics, when compared to a desktop computer, and the possible negative impacts including "gorilla arm" caused by prolonged use of a touch-screen in a vertical position (Korkki, 2011; Davis, 2010; Carmody, 2010). A survey of students with ubiquitous classroom and home access to tablets noted a variety of positive and negative observations including "prevalent visual and musculoskeletal discomfort" (Sommerich et al., 2007). A more targeted study focused on tablet ergonomics confirmed that tablet users take advantage of their devices' many potential display positions, changing device position and their bodily position based on the task. However, high head and neck flexion postures, rather than low-strain neutral positions, were associated with some of these viewing postures (Young et al., 2012).

Within studies of input devices such as touch-screens, comparisons are made between the benefits of the immediacy of direct input, where moving on-screen objects resembles moving objects in the physical world, and those of mechanical intermediaries, such as the indirect input of a mouse. While speed, intuitiveness, and appropriateness for novices are benefits of direct input, mechanical intermediaries often extend human capability in some way (Hinckley & Wigdor, 2011). For instance, a comparison could be made between two types of buttons used to ring a buzzer or bell. In the first case, a game-show buzzer benefits from the direct and immediate input of a fist or palm that slams down

on the button to indicate that a contestant has the answer before his or her competitors. In the second case, a strong-man carnival game involves slamming a mallet down on the base hard enough to project upwards a counter weight to ring the bell at the top. While no Jeopardy contestant would want to lose time to picking up a mallet, the arc of the arm lengthened by the mallet in the carnival game extends and concentrates human strength even as it takes more time to use it and more skill to maneuver it. Similarly, tablets' touch input is immediate and direct, while mouse input aids accuracy and allows one small movement to equate to movement of the cursor across a much larger screen distance. However, the acquisition time – the time required to move one's hand to the input device and use it to point – and learning time associated with a mouse usage is much greater than with touch input.

As pointing devices, mice and touch return coordinates as inputs to a system. Touch inputs are associated with high speed but reduced precision; they are typically faster than mouse inputs for targets that are larger than 3.2 mm, but the minimum target sizes for touch accuracy are between 10.5 and 26 mm, much larger than mouse targets, which tend to be more limited by human sight than by cursor accuracy (Albert 1982; Vogel, Baudisch, & Shift, 2007; Hall et al., 1988; Sears & Shneiderman, 1991; Meyer, Cohen & Nilsen, 1994; Forlines et al., 2007). Touch-screen input accuracy may suffer from spurious touches from holding the device and from occlusion when the finger blocks some part of the graphical interface (Holz & Baudisch, 2010). Mouse input is associated with a single coordinate, whereas touch inputs for multi-touch screens can include multiple coordinates at once, such as an intentional two-fingered gesture or a large thumb touch registering as two simultaneous coordinates.

Despite their similarity in returning coordinates based on mouse or finger position, these two input modalities are associated with different types of events. A touch-screen can detect that the pointing device is out of range when no finger or stylus is touching. The mouse-driven cursor, on the hand, is constrained to the screen and never out of range – a

coordinate is always being communicated to the system. The cursor's ability to stay in place where it was left reduces reacquisition time – resuming from a prior position – in comparison to touch input, where muscle control must be used to keep a finger in a similar position but without touching the screen. A mouse-controlled cursor can be moved without triggering an active selection state; cursor movement is differentiable from dragging. The cursor shows the user the precise location of the contact location before the user commits to an action via a mouse click (Buxton 1990; Sutherland 1964). A touch-screen, on the other hand, does not have these two distinct motion-sensing states; pointing and selecting, moving and dragging, are merged. No "hover" or "roll-over" states as distinct from selection states can exist on a touch-screen, which removes a commonly used avenue of user feedback within graphic user interfaces. Similarly, without a cursor, touch-screen interfaces cannot have cursor icons, which can be used to indicate state or how an object can be acted upon (Tilbrook 1976). For instance, if an erasure state is on, an eraser icon can indicate that clicking or mouse-down movement will erase marks or remove objects. A cursor icon change from an arrow to a pointing finger can indicate to a user that an on-screen object can be selected, clicked, or otherwise acted upon. A change to a flashing vertical bar can indicate that text can be inserted in a given area. Cursor responsiveness can indicate that the system is active and functioning, just as a clock or hour glass or the failure of the cursor to move can indicate that the system is temporarily unresponsive. Interfaces designed for multi-touch screens compensate for this narrower range of events by responding to touch durations and gestures, which, like double-click and right-click events on a mouse, tend to be learned rather than immediately known elements of an interface design.
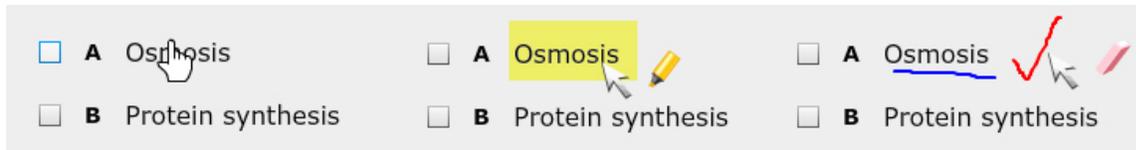
The pointing finger icon and the blue-outlined box on hover indicate that the answer choice can be clicked anywhere to select it. The highlighter cursor icon shows that you have the highlighter tool on, so answer choices can not be selected while in this mode. The eraser icon and the line to be erased turning from blue to red on hover are examples of user feedback unavailable on a tablet.

While tablets can be supplemented with external keyboards, they typically include on-screen touch keyboards used for typing. Performance indicators for human interaction with computers sometimes draw on speed and accuracy measurements, which are particularly well-known measurements for typing. While 40 words per minute (wpm) is considered an average typing speed for computers, reports for average typing speeds with on-screen touch keyboards range between 15 and 30 wpm (Sax, Lau, & Lawrence, 2011). A number of reasons are cited for these lower performance measures. Pressure on a key, the edges of a key, the distance traveled to press that key, and the sound of key pressing provide tactile, kinesthetic, and aural feedback that are missing on a touch-screen keyboard, although some on-screen keyboards use sound or haptic feedback such as vibrations to increase feedback. Without sufficient feedback, users must partially rely on vision for knowledge of a key's position and those of surrounding keys. This visual attention to the keys increases eye movements, or saccades, between the keys and the textual display. These factors, along with accidental inputs and missed inputs that can occur with fingernails that do not register on a capacitive touch-screen, lead to reduced speed and decreased accuracy, with possible differential impact on female users. Touch-screen keyboards have one fewer input state than physical keyboards, since a finger can be off a key or pressing/touching a key on a touch-screen but fingers can not rest on the keyboard without activating those keys. Thus, the traditional typing technique of keeping fingers resting on the ASDL and JKL; keys on a QWERTY keyboard can not be utilized. Keeping fingers pulled back to avoid unintended key taps can lead to fatigue. While typing speed and

accuracy under these conditions are easily measured, it is more difficult to measure other effects of fatigue and of removing reliance on the procedural memory associated with keyboarding skills, which diverts cognitive energy from writing composition to on-screen verification of typed characters (Hinckley & Wigdor, 2011; Ryall et al., 2006; Benko et al., 2009; Hinrichs et al., 2007; Barrett, 1994; Sax, Lau, & Lawrence, 2011; Findlater & Wobbrock, 2012).

Methods

The methodology chosen for the study consisted of a format used on prior occasions by this research team to understand interactions between computer-based user interface and the cognitive processes used for exhibiting knowledge and skills in an assessment context. This approach is best described as a hybrid between a usability study and a cognitive laboratory, making use of the observation and cognitive psychology's think-aloud protocol, also known as "concurrent verbalization" (Ericsson & Simon, 1993).

COMPARISON OF STUDY TYPES	
Usability Study	
Involved Specialists	Usability engineer, user experience designer
Method	One-on-one session, observation, subject is asked to "think aloud"
Key Question	Does the application interface enable a user to accomplish tasks successfully (e.g., efficiently, accurately, in an engaged way)?

Cognitive Laboratory for Assessment	
Involved Specialists	Cognitive psychologist, educator/content expert
Method	One-on-one session, observation, subject is asked to "think aloud"
Key Question	Does a test-takers response to an assessment task appear to produce evidence of the test-taker's knowledge or skill in the targeted area?

Hybrid Usability Study / Cognitive Laboratory	
Involved Specialists	Superset of the above
Method	One-on-one session, observation, subject is asked to "think aloud"
Key Question	Do the tools and interactivity provided within an item allow the test-taker to exhibit their knowledge/skill level without introducing construct-irrelevant variance stemming from usability issues?

While a usability study can be used to discover usability issues, and a cognitive laboratory can reveal potential validity issues – i.e., weaknesses in the construction of the assessment task that suggest that a test-taker's response can not always be taken as indicative of test-taker knowledge or ability – the issues surfaced by a hybrid study could be in either of these two above categories, or the issues could be at the intersection of the two. In other words, a hybrid study can be used to explore whether the interface enables or impedes a student's problem-solving methods or response creation in a way that would help validate or question the use of the student response as evidence of construct mastery.

Sample Issues Revealed by a Hybrid Study
Using the interface produced frustration or increased the load on working memory in a way that could negatively impact student performance and compromise validity.
The interface supported the processes and steps used in the "expert" way of solving the problem (i.e., beginning with the recognition that graphing this type of function would involve a parabola), but the not the novice method of arriving at the correct answer in a longer period of time through trial and error and plotting many possible points.
The interface limited the degree to which the task could be considering a discerning item in that it constrained the possible ways of constructing a response, thereby preventing students from proceeding with a common misconception that would have led to an incorrect response.

Instrument

The study was conducted using a small range of grade-level appropriate assessment items from a Virginia Standards of Learning (SOL) field test. The primary area of interest

was writing, but functionality was included to investigate a range of interactive features used in test items and within the overall online testing environment:

- Multiple-choice answer selections

- "Hot spot" items involving selecting one or more elements or areas of an image

- Drag-and-drop items

- Passages displayed through a paging interface

- Tools such as highlighter, pencil tool, and answer eliminator

- Navigational controls

- An essay-writing interface

A number of decisions were made in how these test items and the online testing interface were moved over to the tablet and in how the tablet was configured for the study. Rather than first making tablet-specific design changes to an existing interface, the researchers decided to use this study to isolate potential problems requiring tablet-specific interface elements. Thus, the existing, browser-based interface was ported from Adobe Flash to Adobe Integrated Runtime in order to run on iPads and Android-based tablets. For the study, 10" Samsung Galaxy Tab tablets were chosen, along with Bluetooth® external keyboards. User interface elements without touch-screen equivalents – in this case, iconic cursors and mouse-rollover "hover" effects – were removed. The testing interface intentionally avoids any use of double-click or right-click (mouse interactions that younger children or computer novices might not be aware of), so all mouse-click based interactions were able to be translated to tap-triggered actions on the touch-screen. Based on existing research described above, the decision was made to keep the same amount of information on screen as would appear on a laptop or desktop. (In the interest of fairness, the testing software is designed to keep this aspect of the presentation consistent across different screen resolutions and screen sizes.) For logistical reasons, it was easiest to bring the tablets preloaded with the abbreviated tests in the form of a native app rather than rely on getting the devices on to school networks for retrieving the content via the internet. Thus,

tablet connectivity and performance issues related to web-accessed content were not included as part of the study.

The Samsung on-screen keyboard packaged with the device was used, but with auto-complete, auto-correct, and auto-capitalize turned off. Although such features are intended to compensate for the decreased accuracy of touch keyboards, options were chosen for this study to more closely mimic computer keyboards and avoid any potential distraction, lack of familiarity, or frustration that might accompany imperfect auto-completion/correction. Sound and vibration upon key press were turned on in response to research citing 20% decrease in input errors, 20% increase in input speed, and lowered cognitive load when using haptic or tactile feedback with touch-screen keyboards, in comparison to non-haptic touch-screens (Brewster, Chohan & Brown, 2007). An external keyboard choice was made based on wireless capability via Bluetooth® and user reviews around ease of use for typing.

Participants & Procedure

Twenty-four students from two Virginia school districts – one that used iPads or iPods in instruction and one that did not – were chosen by school personnel to participate in the hybrid study. Students were selected to represent three grade ranges: grade 4, grade 8, and high school. Since the Standards of Learning (SOL) test is administered almost entirely online at multiple grade levels in Virginia, all students were familiar with the online testing environment and with the format of SOL tests. Students were instructed to use the device to answer several test items and to read the essay prompt before composing an essay as they would for the SOL test. Initially, students used the on-screen virtual keyboard native to the device to construct their essays. Once students had created significant portions of their essays with the on-screen keyboards, the study proctors presented them with compact external wireless keyboards to use in completing their essays. Students were asked to think aloud throughout the study, although during the essay-writing portion proctors granted the

study subjects more quiet concentration time and fewer prompts and reminders regarding thinking aloud. Following this process, students were invited to compare and contrast the two keyboards and were asked to fill out a brief survey before receiving a gift certificate to thank them for their participation.

Results

A number of notable observations were made during the study.

Familiarity with the interface facilitated ease of use. All students participating in the study had used this particular online testing interface for multiple tests in the past. Students reported immediately recognizing the interface despite its translation to a touch-screen, and many were able to describe how they tend to use the tools during a test, using the version on the tablet to illustrate. In fact, one feature used in the Virginia writing field test was not included due to a delay in translating the functionality to the Android. A number of students noticed and even lamented its absence.

Overall, **students found the device compelling**. Excitement and positive comments within a usability study can often be the result of students' eagerness to please adult visitors to their campus and sometimes pleasure in having been selected to participate, thereby missing class. While the effects of this tendency can not be discounted completely, the students in this study immediately started using the device, often used adjectives such as "cool," and voiced how easy it was to use, even on a few occasions when the student was simultaneously repeating an action multiple times and failing to get the desired effect on the first try. A positive affect seemed to be attached to the device.

Students were able to read, **select**, **and navigate with ease**. In the cases where controls and manipulatable objects were larger than finger size, students did not have trouble navigating, reading, page turning, selecting answers, or dragging objects such as marking tools or draggable images in a drag-and-drop interaction.

The most frequently occurring usability issue that cut across different types of system functionality involved the object, button, or control being smaller than or close in size to the area of a student's fingertip. This is a common usability problem found in applications designed originally for use with a mouse but then accessed on a tablet or mobile device, due to the differences in ideal target size for a mouse versus a finger as described above. This small target size problem was found in two contexts: (1) user interface controls such as buttons that are not unique to an item and (2) item-specific content such as images that can be dragged in a drag-and-drop interaction. In regards to the latter, draggers and "hot spots" of different sizes were included to garner a sense of what sizes performed successfully and which did not. One extreme example involved dragging commas over to a sentence. The dragger itself was completely obscured by the student's finger, and the user feedback that tells a student when a dragger is over an area where it can be dropped was also obscured. A number of students gave up after numerous attempts to make the comma "stick."

Tool use was more intuitive but less precise. Students either used tools like the highlighter and pencil as they answered items, or they were prompted with questions like "Do you ever use these tools in this top bar, and if so, how do you typically use them?" and "Do you think they would work similarly on a tablet like this?" Students used the tools with ease, and a few indicated that it was easier because it was more "direct" or because you did not need to use the mouse. However, when students tried to highlight single words, circle a significant word or number, or underline a phrase, several students tried a couple of times to position the mark correctly with a couple acknowledging that the mark was not made exactly where intended but "close enough."

Sometimes students had to touch a button more than once before seeing an effect. The Next button, tools, and the paging turning control sometimes required a student to tap more than once. A number of students experienced this but did not verbally note it, and no student seemed frustrated by it. Further analysis will be required to understand

whether it was device sensitivity, the online testing software itself, or another effect of the control being small with the finger tap being positionally imprecise.

The lack of roll-over or "hover" effects and cursors, as is standard for touch-screens, led to less user feedback for troubleshooting purposes when the application did not respond as they expected. In prior usability testing of this system, most students either used the ability to click anywhere on the answer choice to select it or answered correctly when asked, "Can you click anywhere on the answer choice to select it?" The application ported over to the tablet provided less opportunity for discovering this feature through a roll-over effect (e.g., the cursor changing to a pointing finger when anywhere over the answer choice thereby indicating clickability). When answering multiple choice questions on the tablet, several students continued to click the radio button, sometimes requiring more than one attempt due to the small size of the radio button. One place where student frustration was encountered due to lack of user feedback involved several confounding variables. Within this online testing application, a student can not select an answer when using a tool like the highlighter. This is to avoid accidentally selecting an answer choice when using the highlighter or pencil to mark up some part of the answer choice. In prior usability studies, a few students have been observed forgetting this fact but immediately "course-corrected" by turning off the tool before trying again to select an answer choice. Within the tablet study, some students experienced frustration in this same situation and did not recover or "course correct" as quickly. When queried by the proctor, they indicated that they were aware that answers could not be selected with a tool on. Nonetheless, students in this situation often tried several times before recovering, not knowing whether it was an issue of the device not responding. Unlike the computer-based version of the application, there was no cursor icon to remind students that they had the highlighter, underline, or pencil tool on. There was no roll-over effect on the answer choice to advertise a selectable or non-selectable state. (On a computer, in order to remind the student that answer selections can not be made with a tool, the application replaces the

standard cursor with a tool icon and does not show the pointing finger when over an answer choice if a tool is turned on.) Lastly, since students had experienced other controls where the first tap was not successful, they were more likely to try several taps before switching strategies.

Satisfaction with the on-screen keyboard varied by age and keyboarding ability. With the group of students included in this study, younger students were unsurprisingly observed to have less proficient keyboarding skills, while older students tended to have more proficient keyboarding skills, although a range of ability was self-reported among the high school students. Most of the younger students "hunted and pecked" using either one index finger or two to type on both the on-screen keyboard and the external keyboard. They were either equal in speed on both keyboards or slightly faster on the on-screen keyboard. These students reported liking the visual and slight vibration feedback that they got with the on-screen keyboard. Another reason cited for liking the on-screen keyboard was the simpler keyboard (since not all characters are represented at once), although some students from all age groups took a bit of time to find numbers and other symbols when queried. Some of the younger students could not find the number symbols and asked the proctor to show them. Although no student commented on this, the younger students were observed to have fewer problems with capital letters on the on-screen keyboard than on the external keyboard. Instead of requiring users to hold down the shift key at the same time as another key, the on-screen keyboard involved subsequent key presses (capitalization key followed by letter key) in addition to changing the display of all keys to the capital letter while capitalization was in effect. Older students were either frustrated with the on-screen keyboard or commented that it would "take some getting used to." Their keyboarding input was markedly slower on the on-screen keyboard, with the differential between the two speeds being greater the better a student's typing skills were. One student commented with surprise that the technique he had learned regarding resting

his fingers on the ASDF and JKL; keys did not work, since finger resting led to letters being typed.

The on-screen keyboard covered a portion of the essay being typed. While most students did not complain about a portion of the screen covered up by the on-screen keyboard, some students did close the on-screen keyboard one or more times as they reviewed their essays. Other students did not collapse the keyboard, but not all students were aware of how to remove the on-screen keyboard from view with the designated key (the design of which may vary by device or by keyboard in the case of third party on-screen keyboards). Review and contemplation of the essay topic appeared to require some work on the part of students, since the essay prompt was on the prior screen, but the Previous button was obscured by the on-screen keyboard. Thus, the on-screen keyboard appeared to compound an existing issue: the lack of screen space to show as much of the essay as possible without scrolling and the essay prompt.

While proficient typists preferred the external keyboard over the on-screen keyboard, the external keyboard introduced some new challenges for most students. Some of these issues were related to the particular external keyboard used (or to incomplete integration or impartial compatibility of device and external keyboard), while others might be factors with a wide range of external keyboards. Among the issues in the former category were: presence of some inactive keys ; a tendency for the on-screen keyboard to still open and block part of the screen, even while the external keyboard was connected via Bluetooth; and, occasional key responsiveness issues. Most notable in the latter category was some difficulty with the physical configuration. While students appeared to be comfortable working with the tablet flat on the table when using the on-screen keyboard, the addition of the external keyboard caused observable awkwardness, generating several student comments. Once the external keyboard was added, some students lifted the tablet at an angle, and then set it back down, as if looking for a way to prop it up. With the external keyboard, greater distances within head movements related to

saccades were observed, as most students looked at their fingers during typing and then up at the essay text box, scanning back and forth as they worked on their essays. One student characterized this drawback as "everything not being in one place." Some students appeared to find it awkward to switch between using the external keyboard to type and using one's finger to select text and place the cursor. Switching back and forth between use of the keyboard and on-screen text-based actions did not always occur deftly. It is not clear whether this was related to the physical set-up of two separate devices lying flat, which students appeared to be uncomfortable with, or to a mental shift between "now I'm using this device to work on text, but now I need to switch to this other interface to do other text-related things," as some text functions moved to the external keyboard, but others did not. The small size of the external keyboard felt "cramped" to some of the older students, even as most commented that the external keyboard was still preferable to the on-screen keyboard. Some of the more proficient typists expressed that the small size made this keyboard still not "like a real keyboard," which interfered with how quickly and accurately they could type. Younger students mentioned the small size when describing their preference for the on-screen keyboard with its larger keys. Lastly, in the area of logistical issues, proctors charged all devices prior to sessions, but since the Bluetooth keyboards did not automatically turn off after a certain period of time, one battery depletion issue was experienced during the study. This situation was resolved with additional batteries on hand.

Students had difficulty selecting text and repositioning the text cursor on the tablet. Students found that the imprecision of the finger to indicate location interfered with their ability to select text when trying to replace words or to fix spelling, punctuation, or capitalization. Some students were observed deleting most of a sentence or multiple words just to fix an error, rather than using their fingers to reposition the text cursor in an earlier part of the sentence. In the case where older students declared that they needed a keyboard if this were a "real test," it was most frequently at the moment when students felt like they were making more typing errors, since their keyboarding skills were not easily

transferred to the on-screen keyboard, and tried to correct them but found it difficult to select text to make the appropriate changes. A couple students who were familiar with the iPad commented that selecting and manipulating text was even harder on this device than on the iPad. (Devices with the iOS operating system have a magnifying glass feature that aids text selection.) It was rare that students noticed that they could use the arrow keys on the external keyboard to help overcome this problem of imprecise cursor positioning using the touch-screen. One student, upon discovering that the facilitator had a external keyboard hidden out of sight, asked whether there was also a mouse in the facilitator's bag to help resolve the problems he was experiencing with text editing.

While a couple students tried to use an unsupported gesture, generally students did not assume that the interface drew on gestural conventions or would change appearance when re-oriented. A couple students familiar with the iPad tried to use the pinch-zoom gesture to magnify content, although they did not explicitly complain about content legibility on the 10" screen. No students tried to slide the screen to navigate to the next item, and no one tried to use a swipe to turn a page within the passage, which uses a paging interface. Similarly, no one was observed turning the tablet 90 degrees in an attempt to switch from a landscape to a portrait view.

No substantive differences were observed between the school district that uses iPods and iPads in the classroom and the one that did not. One possible reason for this may be the limited extent of classroom tablet usage in one school district, which tended to use iPods more extensively than iPads and tended toward a greater concentration of iPad usage in kindergarten and first grade than in the older grades. Another reason may stem from the fact that the ported application did not rely on any gestures; thus, if swipe and pinch are more known as conventions to some students, this knowledge would not translate to more adept usage of this application.

Without desktop lockdown, certain actions would move the student to an environment outside of the test, with some effort required by the student or assistance on the part of the proctor to return the student to the test. The particular device used in this study provided a number of ways to navigate outside of the test, many of which were triggered accidentally. Persistent on-screen controls to navigate to a "home" screen were sometimes accidentally activated. Various messages were visible regarding connectivity, which some students inquired about and others accidentally tapped. Several configurations related to the on-screen keyboard were also available via pop-up windows. Some students accidentally opened these while looking for particular keys on the keyboard (e.g., a way to get to numbers and symbols or to close the on-screen keyboard).

Summary and Discussion

Observations from this study in many cases underscored existing research:

- Touch-screen interfaces allowed for direct and immediate input but sometimes involved less precision than mouse-based input, particularly when targets were insufficient in size and finger occlusion was involved.

- Fewer avenues of user feedback were available on the touch-screen version of the testing interface, which occasionally led to usability challenges in the areas where the original design (intended for computer not tablet delivery) of the online testing interface relied heavily on roll-over effects and iconic cursors.

- Reading and navigation were easily achieved by students using the existing online testing interface (which has similar page-turning conventions to an e-reader) on the tablet.

- Use of the touch-screen keyboard required visual attention by all students, even those with keyboarding skills.

- Text-editing was difficult due to the small target size involved with placing a cursor or selecting a range of characters.

The least expected results related to the positive reaction of younger students and novice typists to the on-screen keyboard and the difficulties experienced by most students in regards to the external keyboard. Students without keyboarding skills found the on-screen keyboard easier to use than the external keyboard. While students with more advanced keyboarding skills preferred the external keyboard, they experienced some of the same frustrations with the external keyboard as other students. The physical set-up of a non-upright screen and a keyboard was awkward; the small size of the keyboard was noted as difficult or frustrating; alternating between touch and keyboard use did not seem natural to students; and, the compatibility of the keyboard with the tablet fell short of 100% in that a few keys did not work and the on-screen keyboard occasionally appeared, even with the external keyboard connected.

The issues encountered can generally be assigned to one of three categories:

- A problem with the online testing interface's translation to the tablet that may be resolvable through user experience changes;

- An aspect of the particular physical configuration and hardware elements used, which may be addressed through different but currently available choices;

- Issues related to some essential aspects of the tablet that may be more difficult to resolve.

In regards to the translation of the online testing interface to the tablet, the most successful solution may be a compromise between capitalizing on student familiarity with the current interface and adjusting the software's user experience to draw on some tablet conventions and alternate modes of user feedback. Classroom testing may involve students experiencing assessment materials on desktop and laptop computers as well as tablets, moving interchangeably between devices. Students should not be required to learn an entirely different interface when alternating between devices and should be able to transfer their knowledge and techniques from one device to another. At the same time, anywhere

that inadequate user feedback is available due to the absence of roll-over effects and iconic cursors, alternate means that are suitable for the tablet need to be sought.

The use of an external keyboard to leverage students' keyboarding skills holds some promise, as evinced by older students' greater comfort with the external keyboard than with the on-screen keyboard. However, attention will need to be made to the seamlessness of the external keyboard's compatibility with the tablet and the set-up of the tablet for an optimal viewing position when used for a writing task. For instance, students may benefit from using the tablet in different positions during different portions of the test: handheld for extended reading, flat or slightly tilted for answer selection and drag-and-drop tasks, and upright when used with the keyboard. Increased tablet usage in the classroom may lead to greater student familiarity with different use models such that students will become adept at adjusting tablet position and their own position to suit the task. Additional research will need to be undertaken around use of the stylus for text selection, since adults experience similar levels of precision with a stylus as with a mouse but some studies of children's use of styluses have shown mixed results for use beyond drawing (Mack & Lang, 1989; MacKenzie, Sellen, & Buxton, 1991; Couse & Chen, 2010).

On-going research in this area will take place against a background that could be considered a moving target in at least three ways. First, students' experience with tablets will grow as the use of such devices in the classroom increases, thereby lessening the potential contribution of unfamiliarity to construct-irrelevant variance. Secondly, the types of tasks that are a part of computer-based assessments may grow in complexity with Common Core State Standards consortia pursuing performance-based assessments. In light of this movement, subsequent research will need to take into account not just the basic interfaces for navigation and digital tool use, but also the demands of individual items. It is worth noting the three situations where individuals report preferring to use a desktop or laptop over a tablet (Browne, 2011):

- Tasks that require extensive text entry

- Tasks requiring the interrelated use of multiple tools, windows, or tabs

- Tasks requiring complicated applications or detailed manipulations

Thus, in addition to further study of essays, cross-device comparability studies should include items with high complexity levels, such as those that require accessing a variety of materials and tools – tables, calculator, formula charts – or that require detailed constructions such as graphic organizers or the plotting of two lines and a shaded solution set.

Lastly, the design of tablets will continue to evolve, including innovative ways to provide access to a physical keyboard, such as Microsoft Surface's integrated keyboard built in to the tablet's 5mm-thick case. Advances are also underway with technologies that can detect finger proximity above the touch-screen and respond by expanding the target area as the finger approaches (Yang et al., 2011). Ways to interact with touch-screens may be expanding through the use of sound detection to differentiate between finger tip, pad, nail, and knuckle use (Harrison, Schwarz, & Hudson, 2011). The range of haptic feedback may increase using piezoelectric actuators and voltage controlled protuberances, which can generate pulses and vibrations resembling resistance and other types of tactile feedback (Fruhlinger, 2008; Kaaresoja, Brown, & Linjama, 2006; Laitinen & Mäenpää, 2006; Leung et al., 2007). Tactile screen technologies are being pursued to allow applications to present transparent press-able buttons on demand, rising from a deformable touch-screen surface and then receding back to a smooth surface (Westerman, 2009; Strange, 2012).

A number of other typing systems have been introduced to supplement or replace QWERTY keyboards with systems that are more usable on small touch-screens. Adaptive keyboards begin with traditional key positions but then adjust those positions based on the user's finger position. Swype and ShapeWriter draw on the QWERTY key layout but allow for dragging between letters, lifting only between words. ThickButtons enlarges the keys that are most likely to follow the prior character based on typical key combinations. Dasher, on the other hand, departs from QWERTY altogether to use a zooming model, where typing one

letter leads to the display of additional letters using a predictive model to anticipate likely next characters. SnapKeys is an invisible keyboard designed for thumbs, dividing characters into four categories based on their shape and providing access to these characters through four buttons. 8pen makes use of circular movements to navigate through four quadrants choosing letters. Most of these systems also take advantage of probabilistic predictive models, such as TouchType, which involves more time choosing words from a suggested word list generated by its predictive engine than typing letters. These various advances will ideally bring a range of engaging and affordable devices into the classroom and provide fodder for on-going cross-device comparability research.

References

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

American Psychological Association. Committee on Professional Standards, American Psychological Association. Board of Scientific Affairs. Committee on Psychological Tests, & Assessment. (1986). *Guidelines for computer-based tests and interpretations*. The Association.

Arthur, K. (2009). TouchType fast text entry with word prediction. *Touch Usability.* Retrieved from: http://www.touchusability.com/blog/2009/5/11/touchtype-fast-text-entry-with-word-prediction.html

Arthur, K. (2009). Crocodile touchscreen keyboard with triangular buttons. *Touch Usability*. Retrieved from: http://www.touchusability.com/blog/2009/6/3/crocodile-touchscreen-keyboard-with-triangular-buttons.html

Ball, R., & North, C. (2005). "An analysis of user behavior on high-resolution tiled displays," in *Tenth IFIP International Conference on Human-Computer Interaction*, 350-364.

Ball, R., Varghese, M., Sabri, A., Cox, E., Fierer, C., Peterson, M., Carstensen, B., & North, C., (2005). "Evaluating the benefits of tiled displays for navigating maps," in *IASTED International Conference on Human-Computer Interaction*, 66-71.

Barrett, J. (1994). Performance effects of reduced proprioceptive feedback on touch typists and casual users in a typing task. *Behaviour & Information Technology,*13(6), 373-381.

Benko, H., Morris, M. R., Brush, A.J.B., & Wilson, A.D. (2009). Insights on interactive tabletops: A survey of researchers and developers. *Microsoft Research Technical Report MSR-TR-2009-22*.

Bennett, R.E. (2003). *Online assessment and the comparability of score meaning*. Princeton, NJ: Educational Testing Service.

Breland, H., Lee, Y. W., & Muraki, E. (2005). Comparability of TOEFL CBT essay prompts: Response-mode analyses. *Educational and Psychological Measurement*, *65* (4), 577-595.

Brewster, S., Chohan, F., & Brown, L. (2007, April). Tactile feedback for mobile interactions. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 159-162).

Bridgeman, B., & Cooper, P. (1998, April). Comparability of scores on word-processed and handwritten essays on the graduate management admissions test. Paper presented at the American Educational Research Association, San Diego, CA.

Bridgeman, B., Lennon, M. L., & Jackenthal, A. (2001). *Effects of screen size, screen resolution, and display rate on computer-based test performance* (RR-01-23). Princeton, NJ: Educational Testing Service.

Brown, Q., & Anthony, L., (2012, May). Toward comparing the touchscreen: Interaction patterns of kids and adults. Paper presented at CHI 2012, Austin, TX.

Browne, B., (2011, August 5). Five lessons from a year of tablet UX research. UX Magazine. Retrieved from: http://uxmag.com/articles/five-lessons-from-a-year-of-tablet-ux-research

Bugbee, A.C., (1996). The equivalence of paper-and-pencil and computer-based testing. *Journal of Research on Computing in Education,* 28, 282–99.

Carmody, T. (2010). Why 'Gorilla Arm Syndrome' rules out multitouch notebook displays. Retrieved from: http://www.wired.com/gadgetlab/2010/10/gorilla-arm-multitouch/

Cooper, M. (2009, May 15). Poll results: Touchscreen & QWERTY combo voted your perfect set-up. Conversations by Nokia. Retrieved from: http://conversations.nokia.com/2009/05/15/poll-results-nokia-touchscreen-qwerty-combo-voted-your-perfect-mobile-set-up/

Couse, L. J. & Chen, D. W. (2010). A tablet computer for young children? Exploring its viability for early childhood education. *Journal of Research on Technology in Education*, 43, 75-98.

Czerwinski, M., Smith, G., Regan, T., Meyers, B., Robertson, G., & Starkweather, G. (2003). Toward characterizing the productivity benefits of very large displays. In *Proc. Interact* (Vol. 3, pp. 9-16).

Davis, C. (2010, January 27). The Apple iPad; this Apple has a few worms. *The Ergolab*. Retrieved from: http://ergonomicedge.wordpress.com/2010/01/27/the-apple-ipad-this-apple-has-a-few-worms

De Bruijn, D., De Mul, S. & Van Oostendorp, H. (1992). The influence of screen size and text layout on the study of text. *Behaviour and Information Technology*, 11(2), 71-78.

Dillon, A. (1992) Reading from paper versus screens: A critical review of the empirical literature. *Ergonomics*, 35(10), 1297-1326.

Dillon, A., McKnight, C., & Richardson, J. (1988) Reading from paper versus reading from screens. *The Computer Journal,* 31(5), 457-464.

Dillon, A., Richardson, J. & McKnight, C. (1990). The effects of display size and text splitting on reading lengthy text from screen. *Behaviour and Information Technology*, 9(3), 215-227.

Duchnicky, R.L. & Kolers P.A. (1983). Readability of text scrolled on a visual display terminal as a function of window size. *Human Factors,* 25(6), 683-692.

Duncan, G. (2012, January 26). Are tablet ergonomics a pain in the neck? Digital Trends. Retrieved from: http://www.digitaltrends.com/mobile/are-tablet-ergonomics-a-pain-in-the-neck/

Eklundh, K. S., & Sjöholm, C., (1991, November). Writing with a computer: A longitudinal study of writers of technical documents. *Carin International Journal of Man-Machine Studies*, 35(5), 723-749

Eklundh, K. S., Romberger, S. & Englund, P. (1992) Writing on sheets of paper: a spatial metaphor for computer-based text handling. In Proceedings of the International Conference on Electronic Publishing and Document Manipulation, Lausanne, Switzerland.

Ericsson, K., & Simon, H. (1993). *Protocol Analysis: Verbal Reports as Data* (2nd ed.). Boston, MA: MIT Press.

Findlater, L. & Wobbrock, J. O. (2012). Plastic to Pixels: In search of touch-typing touchscreen keyboards. *Interactions*, 19(3), 44-49.

Flower, L., Hayes, J. R., Carey, L., Schriver, K., & Stratman, J. (1986). Detection, diagnosis, and the strategies of revision. *College Composition and Communication*, 37(1), 16–55.

Forlines, C., Wigdor, D., Shen, C., & Balakrishnan, R. (2007, May). Direct-touch vs. mouse input for tabletop displays. Paper presented at CHI, San Jose, CA.

Fruhlinger, J., (2008, July 8). Nokia's Haptikos tactile feedback tech revealed in patent application. Engadget. http://www.engadget.com/2008/07/08/nokias-haptikos-tactile-feedback-tech-revealed-in-patent-applic/

Gibson, J. J. (1977). The theory of affordances. In Robert Shaw & John Bransford (Eds.), *Perceiving, Acting, and Knowing* (pp. 67-82). Hillsdale, NJ: Lawrence Erlbaum.

Gliksman, S. (2011, April). Assessing the impact of iPads on education one year later. Tablet Computers in Education. Retrieved from: https://edutechdebate.org/tablet-computers-in-education/assessing-the-impact-of-ipads-on-education-one-year-later/

Hall, A.D., Cunningham, J.B., Roache, R.P., & Cox, J.W. (1988). Factors affecting performance using touch-entry systems: Tactual recognition fields and system accuracy. *Journal of Applied Psychology*, 4, 711-720.

Hansen, W. J., & Haas, C., (1998, September) Reading and writing with computers: a framework for explaining differences in performance. *Communications of the ACM*, 31(9), 1080-1089.

Harrison, C., Schwarz, J., & Hudson, S., (2011, October). TapSense: Enhancing finger interaction on touch surfaces. Paper presented at the *UIST* conference, Santa Barbara, CA. Retrieved from: http://www.chrisharrison.net/projects/tapsense/tapsense.pdf

Hartson, H. R., (1998). Human-computer interaction: Interdisciplinary roots and trends. *Journal of Systems and Software*, 43(2), 103–118.

Heffernan, V. (2010, August 13). The promise and peril of 'smart' keyboards. *The New York Times*. Retrieved from: http://www.nytimes.com/2010/08/15/magazine/15FOB-medium-t.html?_r=1

Hinckley, K. & Wigdor, D. (2011). Input Technologies and Techniques. In Andrew Sears and Julie A. Jacko (eds), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications* (pp. 161-176). CRC Press.

Hinrichs, U., Hancock, M., Collins, C., & Carpendale, S. (2007, October). Examination of text-entry methods for tabletop displays. In *Horizontal Interactive Human-Computer Systems, 2007. TABLETOP'07. Second Annual IEEE International Workshop on* (pp. 105-112). IEEE.

Holz, C., & Baudisch, P. (2010, April). The generalized perceived input point model and how to double touch accuracy by extracting fingerprints. In*Proceedings of the 28th international conference on Human factors in computing systems* (pp. 581-590). ACM.

Horkay, N., Bennett, R. E., Allen, N., Kaplan, B., & Yan, F. (2006). Does it matter if I take my writing test on computer? An empirical study of mode effects in NAEP. *Journal of Technology, Learning, and Assessment, 5*(2).

Hourcade, J. P., Bederson, B., Druin, A., & Guimbretière, F. (2004.) Differences in Pointing Task Performance Between Preschool Children and Adults Using Mice*. ACM Transactions on Computer-Human Interaction,* 11(4), 357–386.

Kaaresoja, T., Brown, L. M., & Linjama, J. (2006, July). Snap-Crackle-Pop: Tactile feedback for mobile touch screens. In *Proceedings of Eurohaptics* (Vol. 2006, pp. 565-566).

Keng, L., Kong, X. J., Bleil, B. (2011, April). Does Size Matter? A Study on the Use of Netbooks in K-12 Assessments. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA. Retrieved from: http://www.pearsonassessments.com/hai/images/PDF/AERA-Netbooks_%20K-12_Assessments.pdf

Kingery D. & Furuta R. (1997). Skimming electronic newspaper headlines: A study of typeface, point size, screen resolution and monitor size. *Information Processing and Management*, 33 (5), 685-696.

Korkki, P. (2011, September 10). So Many Gadgets, So Many Aches. *New York Times*. Retrieved from http://www.nytimes.com/2011/09/11/jobs/11work.html?_r=0

Laitinen, P., & Mawnpaa, J. (2006). Enabling mobile haptic design: Piezoelectric actuator technology properties in hand held devices. In *Haptic Audio Visual Environments and their Applications, 2006. IEEE International Workshop on* (pp. 40-43).

Leung, R., MacLean, K., Bertelsen, M. B., & Saubhasik, M. (2007, November). Evaluation of haptically augmented touchscreen gui elements under cognitive load. In *Proceedings of the 9th international conference on Multimodal interfaces* (pp. 374-381). ACM.

Lovelace, E.A. & Southall, S.D. (1983). Memory for words in prose and their locations on the page. *Memory and Cognition*, 1, 429-434.

MacCann, R., Eastment, B., & Pickering, S. (2002). Responding to free response examination questions: Computer versus pen and paper. *British Journal of Educational Technology, 33*(2), 173-188.

Mack, E. (2012, January 12). Snapkeys' quest to assassinate QWERTY. *CNET*. Retrieved from http://ces.cnet.com/8301-33377_1-57358223/snapkeys-quest-to-assassinate-qwerty/

MacKenzie, I. S., Sellen, A., & Buxton, W. A. S. (1991). A comparison of input devices in element pointing and dragging tasks. In proceedings of the Conference on Human Factors in Computing Systems (pp. 161 – 166), New Orleans, LA. ACM.

MacKey, D., (2011). Dasher. Retrieved from: http://www.inference.phy.cam.ac.uk/dasher/DasherSummary2.html

Manalo, J. R., & Wolfe, E. W. (2000). A comparison of word-processed and handwritten essays written for the Test of English as a Foreign Language. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (pp. 13-103). New York, NY: Macmillan Publishing,.

Meyer, S., Cohen, O., & Nilsen, E. (1994, April). Device comparisons for goal-directed drawing tasks. In *Conference companion on Human factors in computing systems* (pp. 251-252). ACM.

Mills, C., & Weldon, L. (1987) Reading text from computer screens. *ACM Computing Surveys,* 19 (4), 329-358.

Morris, M. R., Lombardo, J., & Wigdor, D. (2010, February). WeSearch: supporting collaborative search and sensemaking on a tabletop display. In*Proceedings of the 2010 ACM conference on Computer supported cooperative work* (pp. 401-410). ACM.

Neal, A. & Darnell, M. (1984) Text editing performance with partial line, partial page and full page displays. *Human Factors,* 26(4), 431-441.

Nielsen, J. (2011, September 26). Mobile usability update. Retrieved from: http://www.useit.com/alertbox/mobile-usability.html

Norman, D. A., (1999). Affordances, conventions and design. *Interactions,* 6(3), 38-43.

Norman, D. A., (1990) *The design of everyday things*. New York: Doubleday.

Paek, P. (2005). Recent trends in comparability studies. (Pearson Educational Measurement Research Report 05-05). Retrieved from: http://www.pearsonedmeasurement.com/research/research.htm

Powers, D. E., Fowles, M. E., Farnum, M., & Ramsey, P. (1992). Will they think less of my handwritten essay if others word process theirs? Effects on essay scores of intermingling handwritten and word-processed essays. (RR-92-45). Princeton, NJ: Educational Testing Service.

Powers, D. & Potenza, M. T. (1996). Comparability of testing using laptop and desktop computers (RR–96–15). Princeton, NJ: Educational Testing Service.

Richardson, J., Dillon, A., & McKnight, C. (1989). The effect of window size on reading and manipulating electronic text. In E. Megaw (ed.) *Contemporary Ergonomics* (474-479). London: Taylor and Francis.

Russell, M. (1999). Testing on computers: A follow-up study comparing performance on computer and on paper. *Education Policy Analysis Archives, 7*(20).

Russell, M. & Haney, W. (1997). Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via paper-and-pencil. *Education Policy Analysis Archives, 5*(3).

Ryall, K., Forlines, C., Shen, C., Ringel Morris, M., & Everitt, K. (2006). Experiences with and Observations of Direct-Touch Tabletops. *Proc. Tabletop 2006*, 89-96.

Sax, C., Lau, H., & Lawrence, E. (2011, February). LiquidKeyboard: An ergonomic, adaptive QWERTY keyboard for touchscreens and surfaces. In *ICDS 2011, The Fifth International Conference on Digital Society* (pp. 117-122).

Sears, A., & Shneiderman, B. (1991). High precision touchscreens: design strategies and comparisons with a mouse. *International Journal of Man-Machine Studies*, *34*(4), 593-613.

Siegenthaler, E., Bochud, Y., Wurtz, P., Schmid, L., & Bergamin, P., (2012). The effects of touch screen technology on the usability of e-reading devices. *Journal of Usability Studies*, 7(3) 94-104.

Simmons, T., & Manahan, M. (1999, September). The effects of monitor size on user performance and preference. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 43, No. 24, pp. 1393-1393). SAGE Publications.

Simmons, T. (2001). What's the optimum computer display size? *Ergonomics in Design*, 9(4), 19-25.

Singh, R.I., Sumeeth, M., & Miller, J., (2011). Evaluating the readability of privacy policies in mobile environments. *International Journal of Mobile Human Computer Interaction*, 3(1), 55–78.

Siozos, P., Palaigeorgiou, G., Triantafyllakos, G., Despotakis, T., (2009, May). Computer based testing using "digital ink": Participatory design of a Tablet PC based assessment application for secondary education. *Computers & Education*, 52(4), 811–819.

Sommerich, C. M., Ward, R., Sikdar, K., Payne, J., & Herman, L., (2007). A survey of high school students with ubiquitous access to tablet PCs. *Ergonomics, 50*(5), 706-727.

Strange, A., (2012, June 8). Tactus Unveils Physical Keyboard for Touch-screen Displays. *PC Magazine*. Retrieved from: http://www.pcmag.com/article2/0,2817,2405504,00.asp

Sweedler-Brown, C. (1991). Computers and assessment: The effect of typing versus handwriting on the holistic scoring of essays. *Research & Teaching in Developmental Education*, *8*(1), 5–14.

"Monitor size and aspect ratio productivity research," commissioned by NEC and conducted by the University of Utah (2008). Retrieved from: http://www.scribd.com/doc/34875662/NEC-Productivity-Study-0208

Van Oostendorp, H., & De Mul, S. (1996). *Cognitive aspects of electronic text processing* (Vol. 58). Ablex Publishing Corporation.

Van Waes, L., & Schellens, P. J., (2003). *Writing profiles: the effect of the writing mode on pausing and revision patterns of experienced writers. Journal of Pragmatics*, 35 (6). pp. 829-853

Vogel, D., & Baudisch, P. (2007, April). Shift: a technique for operating pen-based interfaces using touch. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 657-666). ACM.

Wang, T., & Kolen, M. J. (2001). Evaluating comparability in computerized adaptive testing: Issues, criteria and an example. *Journal of Educational Measurement,* 38, 19-49.

Wang, S., Jiao, H., Young, M. J., Brooks, T., & Olsen, J. (2008). Comparability of computer-based and paper-and-pencil testing in K–12 reading assessments. *Educational and Psychological Measurement*, 68(1), 5-24.

Way, W. D., Davis, L. L., & Strain-Seymour, E. (2008). The Validity Case for Assessing Direct Writing by Computer. *A Pearson Assessments & Information White Paper*. Available from http://www.pearsonedmeasurement.com/news/whitepapers.htm

Way, W. D., Lin, C., & Kong, J. (2008, March). Maintaining score equivalence as tests transition online: Issues, approaches, and trends. Paper presented at the annual meeting of the National Council on Measurement in Education, New York, NY.

Westerman, W. C. (2011). *U.S. Patent No. 7,920,131*. Washington, DC: U.S. Patent and Trademark Office. Retrieved from http://www.faqs.org/patents/app/20090315830

Wise, L. & Plake, B. S., (1989). Research on the effects of administering tests via computers. *Educational Measurement: Issues and Practice,* 8(3), 5-10.

Wolfe, E. W., Bolton, S., Feltovich, B., & Bangert, A. W. (1996). A study of word processing experience and its effects on student essay writing. *Journal of Educational Computing Research*, 14(3), 269-284.

Wolfe, E. W., Bolton, S., Feltovich, B., & Niday, D. M. (1996). The influence of student experience with word processors on the quality of essays written for a direct writing assessment. *Assessing Writing*, 3(2), 123-147.

Wolfe, E. W., & Manalo, J. R. (2004). Composition medium comparability in a direct writing assessment of non-native English speakers. *Language Learning & Technology*, 8(1), 53-65. Retrieved from: http://llt.msu.edu/vol8num1/wolfe/default.html

Yang, X. D., Grossman, T., Irani, P., & Fitzmaurice, G. (2011, May). TouchCuts and TouchZoom: enhanced target selection for touch displays using finger proximity

sensing. In *Proceedings of the 2011 annual conference on Human factors in computing systems* (pp. 2585-2594). ACM.

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

Yu, Livingston, S. A., Larkin, K. C., & Bonett, J. (2004). Investigating differences in examinee performance between computer-based and handwritten essays (RR-04-18). Princeton, NJ: Educational Testing Service.