SMARTPHONE INPUT METHOD PERFORMANCE, SATISFACTION, WORKLOAD, AND PREFERENCE WITH YOUNGER AND OLDER NOVICE ADULTS

A Dissertation by

Amanda Lynn Smith

Master of Arts, Wichita State University, 2009

Bachelor of Arts, University of Kansas, 2003

Bachelor of Fine Arts, University of Kansas, 2003

Submitted to the Department of Psychology
and the faculty of the Graduate School of
Wichita State University
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

May 2013

SMARTPHONE INPUT MEHOD PERFORMANCE, SATISFACTION, WORKLOAD, AND PREFERENCE WITH YOUNGER AND OLDER NOVICE ADULTS

The following faculty members have examined the final copy of this dissertation for form and content, and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Psychology.

____

Barbara S. Chaparro, Committee Chair

____

Alex Chaparro, Committee Member

____

Evan Palmer, Committee Member

____

Darwin Dorr, Committee Member

____

Anthony DiLollo, Committee Member

Accepted for the College of Liberal Arts and Sciences

____

Ronald Matson, Dean

Accepted for the Graduate School

____

Abu S. Masud, Interim Dean

DEDICATION

To the one who gives me eskimo kisses

ACKNOWLEDGEMENTS

I am sincerely grateful to my advisor, Dr. Barbara Chaparro. I would not have been able to reach this milestone without her unwavering patience, encouragement, and guidance throughout the dissertation process; likewise, through all of my years in graduate school. Thank you for always being so understanding and generous with your time. You are truly appreciated, and I am excited to learn more from you in the coming years!

I extend my gratitude to my defense committee members, Dr. Alex Chaparro, Dr. Evan Palmer, Dr. Darwin Dorr, and Dr. Anthony DiLollo, for their time, effort, advice, and helpful criticism.

I want to thank the Chaparros, Rondell Burge, and Curtis Lauterbach for allowing me to borrow key pieces of equipment during data collection. I also wish to thank Tyler Crocco for her help with data entry.

Lastly, and most importantly, I must acknowledge my family. My husband Daniel has been absolutely instrumental in this process. He has shared in the sacrifices that accompany graduate studentship, and has supported me through all of it. He did not even complain (excessively)! As far as I am concerned, this is your degree, too. I also mention Emma, our daughter, who has simultaneously been the greatest impediment in finishing this project and the strongest reason to finish. I am extremely thankful for my parents, who have always encouraged me and devoted so much of their time and energy so that I could be successful. I hope that I have made both of you proud. I love you all very much.

ABSTRACT

Smartphones have dominated the American mobile phone market since mid-2012 (Nielsen, 2012). Forty-five percent of American adults own smartphones (Smith, 2012), and report frequently using their devices for the same tasks: text messaging, emailing, and social networking (comScore, 2013), all of which are dependent on text entry. Hardware and/or software text input options are available, but recently vendors have begun to abandon physical keyboards in their device portfolios.

Empirical research does not indicate, comparatively, which smartphone text entry methods are the fastest, most accurate, and most preferred by consumers. Furthermore, potential relationships that users' anthropometry, voice qualities and age-related limitations may have with the accuracy and satisfaction of these input methods have not been addressed.

Two experiments explored the impact that five frequently used smartphone input methods (physical and onscreen Qwerty keyboards, tracing, handwriting, and voice recognition) had on novice user performance, perceived workload, satisfaction, and preference. Relationships between anthropometry, speech and voice qualities with the input methods were also examined. Results from Study 1 demonstrate that younger adults were fastest with voice recognition, but committed fewer errors, reported lower workload, higher satisfaction, and preference with the physical keyboard. Results from Study 2 revealed that older adults were fastest, most accurate, reported lower workload, higher satisfaction and preference for voice recognition. A comparison between age groups suggested that older adults were generally slower and committed more errors. However, performance differences were not found between age groups for voice recognition entry rates and word error rates, as well as for physical Qwerty uncorrected error rates. Additionally, both groups had similar satisfaction and preference ratings for most methods.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

Chapter Page

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF TABLES (continued)

Table Page

LIST OF TABLES (continued)

Table Page

LIST OF FIGURES

Figure Page

LIST OF FIGURES (continued)

Figure Page

LIST OF FIGURES (continued)

LIST OF FIGURES (continued)

Figure Page

CHAPTER 1

INTRODUCTION

Six years after the first iPhone launch, the communications landscape has changed. As of late 2012, forty-five percent of American adults are smartphone owners (Smith, 2012). In mid-2012, smartphone users dominated as the American mobile majority for the first time (Nielsen, 2012a), and it was reported that two-thirds of new mobile device purchases were smartphones (Nielsen, 2012b). Globally, it is anticipated that vendors will ship more smartphones than feature phones for the first time this year (International Data Corporation, 2013).

Why have these phones become a consumer necessity? It seems that their utility is driving their prolificity. A smartphone is more than a simple communications device. It offers the two-in-one flexibility of being essentially a miniature computer with call capability. The small size and light weight typical of these devices allows them to be more portable and convenient for many tasks (e.g., text messaging, email, and social networking) compared to larger form factor devices, such as tablets (comScore, 2013; Edison Research, 2012). Smartphone users are more likely to keep their phones near them at all times, compared to non-smartphone owners (Edison Research, 2012), and twenty-five percent of Americans prefer to use their smartphone over their computer to browse online (Smith, 2011a).

The rapid adoption of smartphones has created a fiercely competitive market. Vendors aim to please with a variety sizes and features, such as high-resolution screens and intelligent personal assistants (e.g., Apple iPhone’s Siri). Some phones are thin and lightweight, designating the touch screen for all forms of input, while others are bulkier, offering a physical keyboard for navigation and typing. Touch screen input methods come standard on both devices, and can also be upgraded by downloading software applications directly onto the phone. Some examples of touch screen input methods include the onscreen keyboard, tracing (e.g., ShapeWriter), handwriting, and voice recognition. With so many options, how does one choose a smartphone? It appears that consumers' top three purchase considerations hinge on the service network, the operating system and the cost of the device (comScore, 2013). In comparison, the user-friendliness of the phone's input method may seem trivial to the consumer's decision, which influences vendors' choices of the input methods offered in their device portfolios. The recent trend is to limit physical keyboards on smartphones. In fact, this is HTC's direction, a major smartphone manufacturer, which announced they are suspending production of physical keyboards in favor of pursuing touch screen haptic technology (Oryl, 2012). However, these decisions are driven by opinion, as empirical research does not indicate which input method is "best" for entering text on a smartphone. Judgments about input methods should not be taken lightly, as many essential smartphone activities require text entry. Almost all users send texts (90.5%), emails (77.8%), and access social networking (65.3%) on their smartphones (comScore, 2013). The consequences of poor input method usability have recently been a source of entertainment (and embarrassment), as illustrated in *Damn You, Autocorrect!* (Madison, 2011) and on websites such as damnyouautocorrect.com and autocorrectfail.org (see Figure 1).

Figure 1. Screenshot demonstrating an experience from a first-time onscreen keyboard user. (K. Lenz, personal communication, April 19, 2013)

A special concern is the usability of smartphone text input methods for the older population. Baby boomers started reaching retirement age in 2011, and it is projected that the older population will account for 20% of the population by 2030 (Administration on Aging, 2010). The smartphone market is entering the "late majority" stage of the adoption curve (comScore, 2013), where marketing strategy will shift to win over those who have not yet bought in to smartphone technology. Many older adults fit into the "late adopter" group. Though they currently are not a large smartphone user group, 31% of American adults aged 55 – 64 and 13% aged 65+ own smartphones (Smith, 2012). Smartphone users in this age group perform the same essential tasks on their devices as the younger users, however, cognitive and physical limitations become more pronounced with aging (Carmeli et al., 2002; Holzinger et al., 2007; Torre & Barlow, 2009). In order to convince this population that smartphone technology is beneficial, the ease with which they can use the device is paramount. Thus, investigating usability of text input methods with this population is important.

Given the ubiquity of smartphone devices and the consideration that tasks dependent on text entry are completed by users at all age levels, it is necessary to explore how different text input methods impact performance, satisfaction, workload and preference. Data was obtained from both younger and older adults to address the strengths and weaknesses of each of five current text entry methods, regarding performance and subjective opinion. This research identified the more successful and gratifying input methods for each age group, and explored comparisons between age groups to suggest that similar performance can be achieved with some input methods.

CHAPTER 2

LITERATURE REVIEW

This chapter includes a brief review of technologies, whose features have been integrated into current smartphone input methods. A classification of current smartphone text input methods is provided. The commonly used text production methods, correction protocols, and metrics for evaluating text input methods are presented. Literature regarding comparative studies between input methods is reviewed. Finally, age-related limitations, which may impact text entry, are identified.

Brief Smartphone History

Forty-five percent of all American adults own smartphones (Smith, 2012), and by late 2012, smartphone users dominated as the American mobile majority (Nielsen, 2012a). Understanding the smartphone’s influence as a consumer necessity requires appreciation for the device beyond its dedication as a communication medium. It combines the call and text capabilities of a mobile phone with the social media, games, and email of a computer, adding the organization of a personal digital assistant, and further offering a camera capable of photos and video. It is an all-in-one device that has the convenience of being mobile.

This array of functionality appeared over decades of iteration, beginning with the Simon personal communicator, introduced in 1993 by IBM and BellSouth. Simon was advertised as a cellular phone with pager, fax, and email capabilities, and equipped with additional features including an appointment scheduler, address book, calculator, Scramble game, and pen-based sketchpad (“Bellsouth”, 1993). More features, such as music, a camera, and maps, could be accessed by inserting a memory card into the phone. It weighed 18 ounces and had a monochrome resistive touch screen and a stylus. Users could either touch letters on a digitized keyboard, or write notes using the stylus (see Figure 2.1).

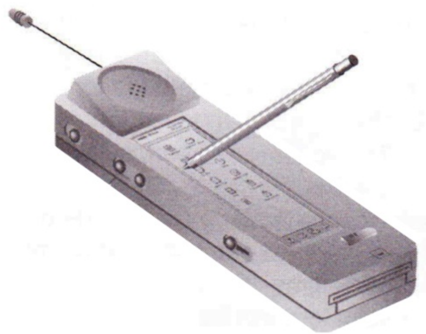


Figure 2.1. The Simon personal communicator (IBM Corporation, 1994).

The Simon was not a commercial success, and was off the market by early 1995. The Nokia Oy Communicator 9000, appearing in 1996, was the market's attempt at a "personal digital assistant phone" and included pocket organizer, fax, email, messaging, and web browsing capabilities. This device allowed the owner to customize ringtones, came equipped with a physical keyboard, featuring the familiar "Qwerty" layout, and ran the GEOS V3 operating platform (Nokia, 1996; Computer Business Review, 1996). A year later, Stockholm Smartphone, affiliated with Ericsson, initiated the "smart phone" product category with the unveiling of the GS88 "Penelope" concept phone.

Figure 2.2. The Nokia Oy Communicator 9000i (n.d.). Retrieved March 25,1013, from http://conversations.nokia.com/wp-content/uploads/2012/09/nokia-9000i.jpg

A few years later, Ericsson began to release smartphones running on the Symbian operating platform, including the R380 World, which included Communication Intelligence

Corporation's Jot® handwriting recognition software in addition to a touch screen Qwerty keyboard. Handwriting recognition software was also a feature for the Palm Treo 180 smartphones (CNET, 2002), which were either sold with a physical Qwerty keyboard or the Graffiti handwriting recognition software. Research in Motion also released their first smartphone, the BlackBerry 5810 model (see Figure 2.3), in 2002. The BlackBerry physical Qwerty, optimized for thumb input, was so successful that it is now considered synonymous with the brand.

Figure 2.3. Research in Motion's BlackBerry 5810 (n. d.). Retrieved March 25, 2013, from http://dienthoaiblackberry.vn/uploads/images/sanpham/3.jpg.

Apple's release of the iPhone in 2007 was revolutionary. The device had a capacitive multi-touch screen, and the iOS interface was designed with the finger as the input device. The full touch Qwerty keyboard included predictive software to boost typing accuracy (Apple, 2007).

Figure 2.4. Example of the iPhone onscreen keyboard (n. d.). Retrieved April 30, 2013, from http://c185824.r24.cf1.rackcdn.com/3835484-3076032-thumbnail.jpg.

Two years after the iPhone arrived on the market, Motorola released the Droid (2009), which ran on Google’s Android 2.0 operating platform. The Droid came equipped with a physical slide-out Qwerty keyboard, as well as a soft keyboard accessible on a multi-touch screen (see Figure 2.4).

Figure 2.5. Motorola Droid (Motorola, 2009).

Android’s voice search was a desirable feature, and was implemented on other smartphones using its operating system. Interest in voice input technology led to Apple’s acquisition of SRI International, a nonprofit research organization, and the release of the Siri voice assistant function on the iPhone 4S in late 2011 (Milian, 2011).

Input Method Classification

Iterations of the smartphone hardware and interface over the past twenty years have afforded many opportunities to test new ground in the realm of input technology. This section includes a brief review of current dominant smartphone text entry methods, grouped into the physical mini-Qwerty, on-screen methods, and voice recognition categories. Strengths, weaknesses, and some of the variability within each type of input are briefly discussed.

Physical Mini-Qwerty.

The “Qwerty” name is derived from the first six keys on the top row of a standard (Silfverberg, 2007) keyboard (Yamada, 1980). On a full-sized keyboard, the layout has been demonstrated to be suboptimal, and can result in excessive wrist extension and ulnar deviation, which enables poor posture and muscle pain (Duncan & Ferguson, 1974). The physical mini-Qwerty keyboard is a hardware feature on many current smartphones. It is designed for thumb input, and is much smaller than its desktop counterpart. It makes use of the familiar Qwerty layout, and the physical buttons have tactile and audio “click” responsiveness. The keyboard is displayed either on an open-face form factor, typically referred to as the “candybar”, or it may be stored underneath a multi-touch screen, accessible as a slide-out or flip-out extension of the hardware. Physical keyboards intended to be operated in portrait mode typically are open-faced, and in landscape mode most keyboards slide out from underneath the device. Though the concept of this input method is straightforward, there are many variations of keyboard attributes which increase its complexity.

Characteristics of the keys include their size (length, width and height), shape, and profile. Backlighting, the color of the keys, and the contrast of the printed digits and symbols on the keys are highly variable. The materials from which the keyboard is manufactured, including the texture and glossiness of the components, and the responsiveness of the keys are also diverse. The overall layout and spacing between keys is inconsistent between devices. For example, many smartphone keyboards sport four rows of keys, but there are others which have three or five. Also, most keyboards keep the traditional Qwerty layout of the keys, while others split the keyboard in half. Some keyboards stagger or curve the key layout, while others organize the keys as a grid.

Clarkson and colleagues (2005) conducted a longitudinal investigation of mini-Qwerty thumb typing rates. Fourteen novice participants (aged 18-34) completed twenty 20-minute sessions with either a Targus or Dell mini-Qwerty. Participants were fastest on the Targus, reaching a mean speed of 34.33 Words per Minute (WPM) after the first 20 minutes and a mean speed of 61.44 WPM during the last session. In addition, they reported the first session average total error rate across the two keyboards as 6.12%. This study illustrates that fast typing speeds are attainable on a small physical Qwerty keyboard. However, this study does not address the perceived satisfaction or workload associated with the participants' experience.

On-Screen Methods.

Touch screen technology has been implemented into smartphones since IBM's Simon in 1993. Since then, with the invention of multi-touch software, it seems that the possibilities for interaction are limitless. On-screen text input allows for a more efficient use of the device's real estate. The type of input and orientation of the input interface can easily be changed to suit the user's preference. Additionally, on-screen methods typically incorporate a greater amount of visual feedback compared to physical methods, including changes in key appearance corresponding to its current state, using pop-ups as a selection notification, and the appearance of a visual "trail", following the user's actions on the interface.

There are some disadvantages to on-screen methods as well. Vibratory feedback and audio cues may both be enabled in the device's settings, but may not be acceptable substitutes for physical buttons. Chapuis & Dragicevic (2011) suggest other disadvantages include the hand's occlusion of small targets on the screen, landing and take-off imprecision, and the "fat finger" issue of selecting small hot spots with larger fingers. Also, "parallax", the perceived shift in object position, can be caused by the user's viewing angle and discrepancy between the displayed and actual location of a control within the touch screen matrix. This issue may decrease accuracy. Fortunately, capacitive touch screens have minimized the issue of parallax.

On-Screen Qwerty. An on-screen Qwerty keyboard is simply a functioning graphical representation of a Qwerty keyboard rendered on a touch sensitive screen. Smartphones are released with their own default on-screen Qwerty, usually dependent upon the operating system on which the device runs. Currently, Android is the most widely-used operating system, with 53% of the American smartphone market share, followed by Apple's iOS (36%), BlackBerry (6%) and Microsoft's Windows Mobile OS (3%) (comScore, 2013). In addition to the default, smartphone owners can download software for other keyboards for free or for a minimal fee from markets like Google Play or Apple's App Store. An example is the SwiftKey keyboard shown in Figure 2.5.

Figure 2.6. The SwiftKey split keyboard layout.

The keyboard only appears when a text field is selected. When the keyboard first appears, all letters are in uppercase. After the user selects the first key the Graphical User Interface shifts the appearance of the keys to lowercase for the remainder of the sentence. In addition, when a key is selected on some keyboards (e.g., default Android and iOS) the user receives feedback in the form of a larger, pop-up graphic superimposed on the rest of the keyboard, if this setting is enabled. Other keyboards display an outline around the selected key (e.g., FlexT9) or a color change on the selected key (e.g., Android keyboard). To select numbers or special characters, users can switch modes by selecting symbol keys, using the "alt" mode key, or by holding a finger on a key for an extended time to bring up additional options.

A problem with the on-screen Qwerty is that users receive little, if any, physical feedback for key presses and subsequently do not have any indication of whether their fingers are selecting the center of an intended key or hitting between two keys. This lack of sensation contributes to mistakes. However, users familiar with the physical Qwerty layout on mobile devices have achieved entry speeds of 22.5 WPM and an 11.8% total error rate on a touchscreen device (Castellucci and MacKenzie, 2011). This is a faster entry rate than reported by Arif and colleagues (2010), which demonstrated a 15.92 WPM for the iPhone keyboard, however they reported a similar total error rate of 10.38%.

Tracing. Tracing is also known as "shape writing", and was first introduced by Zhai and Kristensson (2003). This is a gestural method of text input, where users slide their finger across the keyboard from letter to letter (typically in the Qwerty layout) to form a word. If the word contains "double" letters, the user makes a loop gesture over the key. For example, in the word "hello" the user would make a loop over the "l" key (see Figure 2.6).

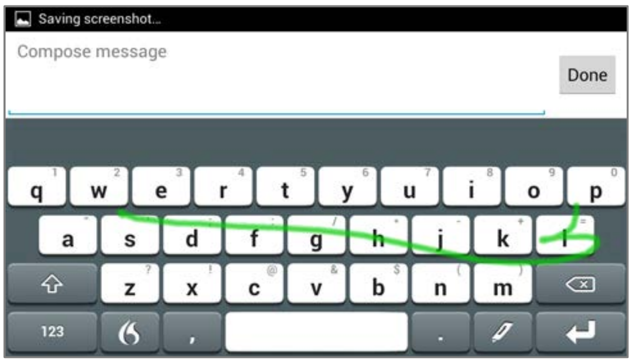


Figure 2.7. Demonstration of the tracing input method.
The loop above the "l" to create the double-l in "hello".

In the most often used programs (e.g., Swype, FlexT9, SwiftKey), a line "follows" the path traced by the user. The look and responsiveness of this line varies across software, and can also be customized. Once the trace is complete and finger is lifted off of the screen, the software compares the user's trace pattern with geometric templates located in the program's lexicon. The best match is then retrieved and displayed as the output. The precursor to most tracing software, ShapeWriter, was reported to enable novice entry rates of 25 WPM (Kristensson, 2007). A couple of problems with this method are that the user's hand typically obscures the screen while tracing. This can be more problematic for those who are less familiar with the Qwerty layout. Another issue occurs when the finger is lifted off of the screen before the user is done tracing the word. The software returns the word closest to the partially traced word, which leads to corrections.

Handwriting. Traditionally, computer-based handwriting systems utilized a stylus. This method was frequently used to interact with personal digital assistants (PDAs) from the early 1990s, and it has been an option on some smartphone devices since the Ericsson R380 in 2000. Handwriting recognition software has been used to analyze different types of user input, from cursive writing to block letters to unistrokes (MacKenzie & Zhang, 1997). Unistroke methods introduced the concept of writing each letter of the alphabet with a single stroke. An advantage to the unistroke method is that the user can write without having to worry about stroke order. A disadvantage is that the alphabet takes a while to learn as some letters do not closely resemble their Roman counterparts. The Graffiti alphabet, a variation of the unistroke method, is more intuitive and uses more than single-stroke characters. Other software (e.g., FlexT9) allows the user to print in a manner that resembles their normal writing style. For example, some software allows cursive handwriting or will accommodate languages other than English (e.g., DioPen).

On today's smartphone devices, users trace letters with a finger. When the user composes text, they can trace letterforms on top of each other, or from left to right. Typically, software will provide a colored trail following the finger's creation of each letterform (see Figure 2.7). In some cases, numbers and letters are created in the same space (e.g., FlexT9), while other software will provide specific areas in which to compose letters or numbers (e.g., Graffiti). To space between words, the user can hit the space bar, or use a gesture such as a left-to-right trace. Likewise, a right-to-left gesture in some software (e.g., FlexT9) will delete a letter.

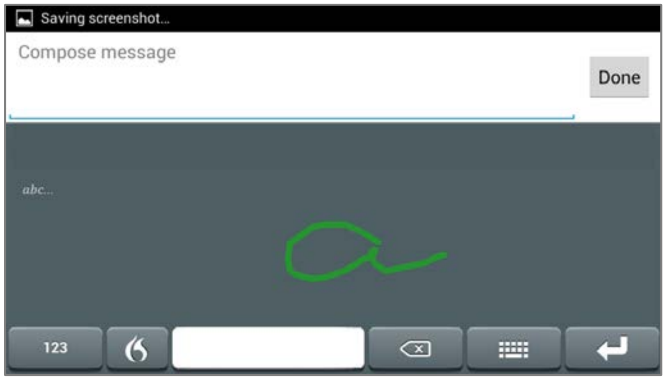


Figure 2.8. The FlexT9 handwriting interface.

Though handwriting is a familiar way to compose text, the entry rate is slow compared to other input methods. Bailey (1996) demonstrated that unconstrained handwriting is roughly 20 WPM. Given familiarity with everyday handwriting, users may expect to immediately achieve the same entry rates on mobile devices (MacKenzie & Soukoreff, 2002), though most handwriting recognition software is often slower. However, the familiarity of handwriting can make it advantageous over other methods of entry for novice users. In a study of novice users, handwriting was initially faster (21.5 wpm) than keyboarding (19.6 wpm). Even after 250 minutes of practice, performance between the two methods revealed similar entry and error rates (Kristensson & Denby, 2009).

Voice Recognition.

Speech is a natural way to communicate, and it has the potential to be much faster than any manual form of input. For instance, expert keyboardists type at a rate of 80 WPM, whereas speech rate can reach a speed of 200 WPM (Cox et al., 2008). Until recently, voice recognition on mobile phones was not technically feasible. The software used to process the speech signals took more processing power than was available. This technology has been in development since the middle of the last century, and yet some of the most accurate software still looms at the 90% level of accuracy. There are many variables that can affect the accuracy of speech recognition software, including user characteristics (i.e., age, gender, native language, individual speech qualities), user behaviors (i.e., long-term singing or smoking), and characteristics of the environment (i.e., type of background environment and noise level) (Benzeghiba et al., 2007). In spite of the myriad of conditions that affect performance, speech recognition is a favorable option for those in eyes-busy situations or for people with visual impairments. It may also be useful for those with decreased manual dexterity.

Voice interfaces have a wider range of functionality, including text-to-speech communications, such as composing an email or text message, to commands, like searching the internet. Virtual assistants are more popular after Apple's release of Siri on the iPhone 4S in 2011. A virtual assistant is a voice interface that acts as a personal assistant or secretary. Stoyanchev (2012) described two classifications of assistants, "mute" and "speaking". An example of a "mute" assistant is Google Voice Search, which uses visual and audio prompts to communicate. An example of a "speaking" assistant is Apple's Siri, which addresses the user, displays emotion, and apologizes.

Typically, smartphone voice recognition software is accessed through the phone's graphical user interface displayed on the touch screen. However, there are instances in which it may also be accessed through hardware (e.g., the HTC myTouch "Genius" button). In many cases, the user will receive a visual prompt (e.g., VLingo, Google Voice) or audio prompt (e.g., "Genius" button) to begin speaking. Once the user is finished speaking, they can either tap the touch screen or pause to start the processing phase. Processing usually contains some type of visual progress bar or other progress graphic to provide visual feedback. After the software processes the input, the output is displayed. Some software provides a vibration once the software is finished processing (e.g., Google Voice).

Input Method Evaluation

The process of evaluating an input method involves decisions about the methods of text production, handling corrections, and which metrics best explain the participants' experiences. This section describes current practices in evaluation.

Text Production.

There are two common methods used for text production, text copying and text creation. The first method, text copying, is a method that requires participants to copy a presented body of text. The researcher decides which text to transcribe, creating greater experimental control. However, for long bodies of text, the participant must divide their attention between the presented text and monitoring their transcription progress. The attentional demands of the task, rather than the input method itself, may result in an inaccurate depiction of performance for the method under evaluation (Soukoreff, 2002).

The second method, text creation, instructs the participant to either memorize or spontaneously generate their own body of text. While this may seem more natural to the participant, this production method has many drawbacks. Text created on impulse is not likely to represent the frequency distribution of the language's word usage. Without reference text, the accuracy of the output is more likely to incorporate error due to confounds related to limited resources available in working memory and spelling capability. Entry rates may also be confounded by the time taken to decide what to transcribe. Furthermore, allowing participants to spontaneously generate text makes the comparison of participants' intent and the transcribed text impossible (Soukoreff, 2002). One way to regain experimental control, and maintain a balance between internal and external validity, is to either dictate the text to participants (Ward et al., 2000) or present them with brief phrases that are simple to memorize. A phrase set, developed by MacKenzie and Soukoreff (2003), is comprised of phrases which are 4 – 8 words in length, easy to remember and representative of the English language. This phrase set is commonly used in text input research.

Correction Protocol.

There are essentially three different error correction protocols for text entry experiments: none, recommended, or forced correction (Arif & Stuerzlinger, 2009). In the “none” protocol, participants are not allowed to correct errors. They are told to enter the text as quickly and accurately as possible, and to also continue transcribing the phrase even if they discover they had made an error. In the “recommended” condition, participants are asked to correct any errors that they notice as they are transcribing the text. They may be asked to pretend they are writing the text to a friend, which may help reduce the number of errors left in the final transcription (Castellucci & MacKenzie, 2011). In the “forced correction” condition, participants are told they must correct all transcription errors before moving onto the next trial. Arif and Stuerzlinger (2009) investigated the impact of all three correction protocols on text entry performance metrics with expert desktop Qwerty typists. They found that there were no significant differences between the three correction conditions for entry rate. The number of edit keystrokes between the “recommended” and “forced correction” conditions also did not significantly differ.

Metrics.

Text entry performance metrics are inconsistently reported in the literature, which makes comparing studies difficult. The type of metric used is dependent on the study’s established correction protocol and the level at which the data is observed and recorded. Within this section it is helpful to mention some useful terminology (Arif & Stuerzlinger, 2009; Wobbrock, 2007):

- *Presented text* is the stimulus string (*P*) presented to the participant. |*P*| is the length of *P*, which includes all characters, spaces and punctuation.
- *Transcribed text* is the final text string (*T*) entered by the participant. |*T*| is the length of *T*, which includes all characters, spaces and punctuation.
- *Input Stream (IS*) refers to all keystrokes made during the entry process, including backspaces, and |*IS*| refers to the length of *IS*.

Entry rate. Entry rate is a measure that describes transcription speed with the input method under investigation. The calculation of entry rate is dependent upon the unit at which the transcribed string is produced. In other words, text input methods that produce the transcribed string at the word level would empirically report entry rates using the Words per Minute metric. Input methods that produce the transcribed string at the character level may utilize any of the following metrics, each of which addresses slightly different text entry behavior (Wobbrock, 2007).

Words per Minute and Characters per Second. Words per Minute (WPM) is the most frequently reported measure of text entry performance. WPM has been used as an empirical measure of entry rate with input methods that produce text at the word-level and at the character-level. However, the definition of what constitutes a "word" unit varies depending on the input method. For instance, a "word" unit in speech-based text entry is defined as a spoken word. However, it is common practice to define a "word" produced by other input methods as a 5-character unit, including spaces and punctuation (Yamada, 1980). This does not take into account the total amount of keystrokes in the input stream, rather, only the length of the final transcription and the time to create it. In other words, backspaced characters and shifting are not reflected in the measure. WPM is calculated as

$$WPM = \frac{|T| - 1}{S} \; x \; 60 \; x \; \frac{1}{5} \tag{2.1}$$

T is the length of the final transcription. S is the time taken to enter the transcribed string in seconds, starting from the entry of the first character and ending with the last character, including backspaces. This is the reason "1" is subtracted from the transcription length in the numerator. The quotient is multiplied by 60, the number of seconds in a minute. Then the product is divided by 5, which is the average word length (Wobbrock, 2007). As an example:

```
we are subjects and must obey
|                           |
```

t = 0 sec   t = 10 sec

There are 29 characters (including spaces), WPM could be calculated as 28/10, then multiplied by 12. WPM for this phrase is calculated to be 33.6.

A character-level measure, Characters per Second (CPS), is less often reported by researchers, but can be derived by simply calculating the first ratio in equation (2.1).

$$CPS = \frac{|T| - 1}{S} \tag{2.2}$$

Using the same example, the calculated CPS would be 2.8.

Adjusted Words per Minute. There is a trade-off between speed and accuracy when entering text. Typically, entry rates decrease to avoid making or to correct errors. However, if errors remain in the final transcription, the calculated WPM or CPS entry rate is likely inflated. The Adjusted Words per Minute (AdjWPM) metric was created for this reason. AdjWPM, defined by equation (2.3), reflects a more accurate estimation of speed by adding a penalty exponent for the proportion of uncorrected errors left in the transcribed phrase (Matias et al., 1996; Wobbrock et al., 2006).

$$AdjWPM = WPM\ x\ (1 - U)^{a} \tag{2.3}$$

U is the uncorrected error rate (ranging from 0 – 1). The penalty exponent, *a*, is usually set to 1. This value can be increased to more severely penalize the raw WPM, however, this is somewhat arbitrary. Because of this, some researchers recommend use of the “forced correction” protocol and to discard trials containing errors (Lewis, 1999).

Keystrokes per Second. WPM and AdjWPM do not consider all of the participant’s behaviors during the creation of the transcribed string (i.e., backspaces and character reentry). Keystrokes per Second (KSPS) is a useful tool for this. It is calculated as:

$$KSPS = \frac{|IS| - 1}{S} \tag{2.4}$$

IS refers to the input stream, containing all characters, spaces, and backspaces utilized to create the final transcribed string in *S* seconds. For example:

we arw<**e subjec**ra<<**ts and** n<**must obey**

| |

t = 0 sec t = 10 sec

In this example, backspaces are indicated by “<” and the final transcription is indicated by the characters in bold face. Even though the final transcription is 29 characters, there were actually 37 characters in the input stream, including 4 erroneous keystrokes and 4 backspaces used to correct them. In this case, KSPS is calculated as 36/10 or 3.6. Even though it is “keystrokes” per second, it can also be applied to any non-typing method that produces a measurable input stream (Wobbrock, 2007).

Accuracy. There is little consensus in the literature regarding error rate metrics. The choice of which error metric to use depends upon the correction protocol used in the study, the method used to gather the performance data, and the goals of the researcher. Error rates are typically classified into “corrected errors” made during the production of the input stream, and “uncorrected errors”, which are errors left in the final transcription.

Keystrokes per Character. Keystrokes per Character (KSPC) is a frequently used accuracy metric. This measure is applicable to error quantification during the text entry process. It’s the ratio of the total number of characters and backspaces entered in the input stream (*IS*) to the total number of characters in the transcription, shown in equation (2.5).

$$KSPC = \frac{|IS|}{|T|} \tag{2.5}$$

Taking the previous example:

```
we arw<e subjecra<<ts and n<must obey
|                                   |
t = 0 sec                           t = 10 sec
```

There were 37 keystrokes, but only 29 characters in the transcribed string. This gives us a KSPC ratio of 37/29 = 1.28. A KSPC value of 1 indicates perfect entry (Wobbrock, 2007).

A disadvantage to this method is that, though some characters may be backspaced and reentered, there is no way to indicate whether some of these characters were initially correct (Soukoreff & MacKenzie, 2004). To do this, a character-level analysis of the errors should be conducted.

Minimum String Distance. Minimum string distance (MSD) is a sequence comparison algorithm that calculates the minimum number of operations (e.g., substitutions, insertions and deletions) required to convert the presented string (P) into the transcribed string (T). It is based on Levenshtein's theory of self-correcting binary codes, but was applied to text entry research by Soukoreff and MacKenzie (2001). For instance, consider the following strings:

Presented: `we are subjects and must obey`

Transcribed: `we aresubjects and musr obeey`

It can be seen that the errors in the transcribed string include an omission of a space, a substitution of "r" for "t", and an insertion of an additional "e". For this pair, MSD is calculated as 3. However, these calculations are not typically straightforward for two reasons. One concerns the length of the transcribed string, and the other concerns the categorization of differences between the two strings. An example given by Wobbrock (2007) addresses the deviation between strings:

Presented: `quickly`

Transcribed: `qucehkly`

In this case, the MSD error rate (Soukoreff & MacKenzie, 2001) is calculated as

$$MSD\ error\ rate = \frac{MSD(P,T)}{MAX(|P|,|T|)}\ x\ 100\% \tag{2.6}$$

Equation (2.6) uses the longer of the two strings as the value in the dominator, which ensures that the quotient never exceeds 1, as well as applies a penalty for transcribing too few or too many characters. In this example, MSD error rate is (3/8) x 100% = 37.5%.

However, even after discovering the error rate, which describes the number of errors, it is often very difficult to classify the errors occurring during the input stream. In the above example, there are 3 errors, but how can these errors be categorized? The actual intentions of the person transcribing the text are unknown. Thus, algorithms exist that can evaluate all of the optimal alignments between the two strings. Plausible alignments between the presented and transcribed strings are given in the following example (Wobbrock, 2007):

Presented$_1$: `qu-ickly`

Transcribed$_1$: `qucehkly`

Presented$_2$: `qui-ckly`

Transcribed$_2$: `qucehkly`

Presented$_3$: `quic-kly`

Transcribed$_3$: `qucehkly`

Presented$_4$: `quic--kly`

Transcribed$_4$: `qu-cehkly`

The first pair shows an insertion of “c”, and substitutions of “e” and “h”. The second pair demonstrates the substitutions of “c” and “h”, an insertion of “e”. The third pair shows substitutions of “c” and “e” and the insertion of “h”. The last pair demonstrates the omission of “i” and insertions of “e” and “h”. Frequencies of optimal alignments for each type of error will allow the researcher to calculate the weights, and probabilities, associated with each error type for each character. For example, of the four plausible situations, the chance that “c” was substituted for “i” is 50%, since this was demonstrated in half of the above scenarios. Likewise, “i” was omitted 25% of the time.

Unified Error Metric. Typically, the Minimum String Distance method has been used only to compare the presented and transcribed strings, thus, another measure, such as Keystrokes per Character, was needed to describe the input stream. However, these metrics cannot be combined to express total error rate. The Unified Error Metric (Soukoreff & MacKenzie, 2003) was developed to address this concern. All characters are assigned to one of four categories: correct (*C*), incorrect but fixed (*IF*), a fix (*F*), and incorrect but not fixed (*INF*). *C* refers to all characters in the transcribed string that match the presented string. *IF* applies to all incorrect characters deleted in the input stream. *F* takes into consideration all backspaces, both for correct and incorrect characters, in the input stream. *INF* refers to any incorrect characters left in the transcribed string. For example:

$$\underbrace{\texttt{we ar}}_{C}\underbrace{\texttt{ }}_{INF}\underbrace{\texttt{subject}}_{C}\underbrace{\texttt{d}}_{IF}\underbrace{\texttt{<}}_{F}\underbrace{\texttt{s and must obey}}_{C}$$

This allows the researcher to devise the following error rates

$$Corrected\ Error\ Rate = \frac{IF}{C+INF+IF} \tag{2.7}$$

$$Uncorrected\ Error\ Rate = \frac{INF}{C+INF+IF} \tag{2.8}$$

$$Total\ Error\ Rate = \frac{IF+INF}{C+INF+IF} \tag{2.9}$$

One advantage to the Unified Error Metric is that backspaced characters, which were initially correct, can be counted as such. Typically, corrected error rate is larger than uncorrected error rate. Entry rate is also slower for higher corrected error rates. An input method that has a high entry rate and a low uncorrected error rate, even in the face of numerous corrected errors, could still be considered a success (Wobbrock, 2007).

Comparative Research

This section presents comparative studies between input methods that have applied some of the mentioned evaluation principles. Arif and Steurzlinger (2009) examined data presented in several journal articles and inferred that of a full-length standard and projection Qwerty, thumb Qwerty and on-screen stylus Qwerty, Twiddler, 12-key multitap and Graffiti, the standard full-size Qwerty was the fastest (*M* = 75.84 WPM), followed by thumb mini-Qwerty (*M* = 50.86 WPM). The on-screen Qwerty and Graffiti methods were slower (*M* = 24.83 and 11.62, respectively). However, this was deduced from several different sources, each of which used different hardware, data collection protocols, performance metrics, numbers of participants, participant expertise, and stimuli.

Költringer and Grechenig (2004) compared handwriting (Graffiti 2) to an onscreen Qwerty keyboard on a Palm Pilot emulator (graphic tablet with pressure sensitive pen and a 19 inch display). Participants consisted of 12 young adults (*M* = 27.17 years old) with some basic experience with the type of input methods. They input three each of three different types of phrases for each device. One phrase type utilized the MacKenzie phrases (MacKenzie & Soukoreff, 2003), others were commands (e.g., e-mail addresses), and others were numbers. They found that the virtual keyboard was fastest for each of the phrase types (ranging from 10.25 WPM for commands to 14.55 WPM for numbers) and that handwriting was slower (ranging 6.64 WPM for commands to 9.24 WPM for phrases). Total error rate was also much higher for handwriting (19.35% during phrase entry) than the on-screen keyboard (4.11% for phrase entry). There no significant differences in reported workload or satisfaction between the two input methods. However, participants generally preferred Graffiti 2 and reported it was more intuitive than the onscreen Qwerty.

Hoggan and colleagues (2008) tested a physical Qwerty, onscreen Qwerty and an onscreen Qwerty with vibration feedback with 12 university students, who had experience sending text messages with a physical Qwerty keyboard. Participants tested these input methods in a lab and also on a subway train. Their results indicated that the physical keyboard was the fastest and most accurate in both environmental settings, and that the standard touchscreen was least accurate. There were no significant differences between the physical keyboard and the tactile touchscreen regarding accuracy. However, the physical keyboard was significantly faster than both the tactile and standard touchscreen keyboards. While the lab and subway conditions did not differ for accuracy, participants did enter text significantly more slowly in the subway condition. Additionally, participants reported that the standard touchscreen keyboard produced higher perceived workload than for the tactile touchscreen and physical keyboards.

Castellucci and MacKenzie (2011) tested an on-screen Qwerty, handwriting (DioPen) and shape writing (SWYPE) with six younger adult users, who had some level of exposure with these methods, on an Android device. Participants entered ten phrases with each method in portrait mode. Results indicated that SWYPE had the lowest total error rate (6.2%), followed by Qwerty (10.4%) and DioPen (25%). Corrected error rates were highest for DioPen. They were also high for Qwerty because participants were deleting entire words to insert a space where one was missed or an erroneous character was typed. In terms of entry speed, Qwerty was the fastest at 21.4 WPM, SWYPE was 17.4 WPM and DioPen was the slowest at 6.1 WPM.

Though speech-to-text is mostly studied as it relates to computer dictation software, there are some studies comparing voice recognition to manual forms of input. Karat and colleagues (1999) investigated speech input against the keyboard and mouse. Though speech was faster for entering text (107 uncorrected WPM), it took nearly 3 times as long for users to detect and correct errors with speech input than the time taken to enter the text. Cox and colleagues (2008) investigated speech and both multitap and predictive 12-key text input for text messages that they generated from actual undergraduate text content. They measured speed in WPM, word error rate, and workload. Speech was shown to be the fastest, with the highest satisfaction and lowest workload compared to the other techniques and combinations thereof. Predictive text was shown to have the lowest word error rate, followed by speech and multitap.

For novice older adults, a study by Rau and Hsu (2005) compared several input methods for completing tasks on the web. They hypothesized that direct manipulation interfaces (handwriting, touch screen and voice) would outperform keyboard and mouse. They found that handwriting and touch screen use allowed faster time to task completion and had lower error rates in search tasks compared to voice and keyboard/mouse input.

Read and colleagues (2001) evaluated mouse, keyboard, speech, and handwriting input with children for computer copying and composing tasks. The metrics that they used varied somewhat across input types, for example, efficiency was measured in Characters per Second (CPS) for all input methods besides speech, which was measured in Words per Second and then multiplied by 0.33 to approximate CPS. This was an exploratory study, and they did not definitively report that any one method was better than another. However, their reported error rates seem to show that speech input was the least accurate and keyboard and mouse input were most accurate.

Physical Limitations and Older Adults

Older adults, those aged 60 and above, have not typically been viewed as “smartphone users”. Indeed, most American smartphone “early adopters” are in the 18 – 34 year age range (Smith, 2012). Some older adults are included in this group, in fact, 31% of adults aged 55-64 and 13% of adults above the age of 65 own smartphones (Smith, 2012), which reflects a more rapid adoption rate than expected (Stroud, 2012). Younger and older adults use their smartphones to engage in the same tasks, such as sending/receiving text messages and email, and social networking (Smith, 2011b). However, older adults are a more heterogeneous group regarding sensory impairment, communicative ability, and familiarity with technology (Newell & Gregor, 2004), and age-related cognitive and physical declines can create substantially greater difficulty in accomplishing these tasks. This section reviews some of the age-related limitations relevant to text entry.

Psychomotor impairments.

Carmeli and colleagues (2002) suggested that age-related changes in hand functioning often result from musculoskeletal, vascular, and nervous system declines. Structural deterioration occurs in joints, muscles, tendons, bones, nerves, blood supply, and skin. In this age group, changes in tendons can result in decreased range of motion for joints and decreased flexion power. Even though the hand does not lose as much muscle mass as in other areas of the body, there is still a 20-25% decline in hand-grip strength after the age of 60 (Carmeli et al., 2002). Thumb abduction, pinch and grip strengths decrease (Boatright et al., 1997), and reduced index finger abduction strength and greater force variability occurs (Galganski et al., 1993). The age-related decline in motor performance is attributed to the atrophy of 25% of hand muscle motor axons (Carmeli et al., 2002). Tactile and vibratory sensitivity is reduced (Kenshalo, 1979) from sensory mechanoreceptor atrophy (Carmeli et al., 2002). This decrease in sensitivity may be responsible for reduced ability to grip and lift objects (Cole et al., 1998). Combined, these declines make it more difficult for aging adults to complete tasks that involve precision dexterity, however, changes in performance (i.e., performance time, range of motion) are more apparent after the age of 75 (Carmeli et al., 2002). Additionally, it is common for these functional declines to be compounded by pathological conditions, such as arthritis, osteoporosis, and Parkinson's disease.

Changes in performance time on task can also result from the more conservative strategies deployed by older adults. For example, a study by Walker and colleagues (1997) demonstrated older adults are more error averse, meaning that they usually more cautious in movement towards a target even when told that there is no error penalty. Spatial finger and wrist coordination declines (Contreras-Vidal et al., 1998) and movements take longer to execute

(Becker & Webbe, 2006). Furthermore, older participants take more time to verify accuracy of their movement and are slower after making an error than their younger counterparts (Vercruyssen, 1996). The cues older adults use are also relative to the type of task. For example, in settings where time pressure is not a factor, older adults utilize proprioceptive cues, or information regarding joint position and muscle force, and sensory cues similar to younger adults (Chaput & Proteau, 1996a), however, when time pressure is introduced older adults rely more on proprioception (Chaput & Proteau, 1996b). However, proprioceptive acuity is reduced in older adults and Adamo and colleagues (2007) suggested that functional movement errors in complex sensorimotor tasks may translate to an average of 4.4 cm displacement in finger-tip position. Holzinger and colleagues (2007) state that five main factors that vary with age are time to learn a task, speed of performance, error rate, retention over time and subjective satisfaction. Additionally, older adults tend to report greater difficulty, fatigue (Czaja & Sharit, 1993), as and workload (Fox et al., 1995) than younger adults on the same tasks.

These declines can create difficulty when interacting with various modes of input. A study by Kobayashi and colleagues (2011) investigated older Japanese adult (aged 60-70) performance with four methods of touch screen interaction on an iPod over a week's time. They found a 39% error rate with a tapping method when selecting a 30-pixel target, however, they also showed that dragging and pinching gestures improved over time. In fact, participants reported that pinching and dragging gestures were more comfortable and easier than tapping. They suggested that parallax and large finger contact area contributed to missed targets, and recommended that targets on smartphone devices should be larger than 8 mm. A study by Jin et al. (2007) demonstrated that target size for adjacent buttons should be no less than 16.51 mm and separated by no less than 3.17 mm for older adults. They also revealed that users with a greater degree of manual dexterity preferred and were faster and more accurate with smaller target sizes and decreased spacing between targets. Research by Wacharamanotham and colleagues (2011) has shown that older adults may fare worse when using a tapping mode of interaction with touch screens, due to increased finger oscillation. However, a swabbing interaction (i.e., touch, slide, lift) resulted in higher accuracy and satisfaction rates.

Speech.

In addition to manual function declines, variability in speech characteristics increases over time. Physical changes affect the elasticity of the respiratory system, which leads to decreased lung pressure. Calcification, muscle atrophy and changes in mucous secretions of laryngeal tissues can result in less stable vibrations of the vocal folds (Linville, 2000; Gorham-Rowan & Laures-Gore, 2006). Tongue strength decreases, teeth may be lost, and the volume of both the oral cavity and vocal tract increase (Linville, 2000; Xue & Hao, 2003). Other changes occur from reduced sensory feedback (presbycusis), diminished motor speed and control, and lesser cognitive-linguistic function (Torre & Barlow, 2009).

Age-related acoustic changes have been demonstrated with fundamental frequency, increased jitter, shimmer, and breathiness, as well as a slower speech rate (Vipperla et al., 2008). Fundamental frequency (F_0) is how high or low a person's voice sounds; the perceptual correlate of F_0 is pitch. F_0 decreases with age (60-89 years old) for women, but increases with age for men, though men still have significantly lower F_0 than women. Variability increases for both sexes (Torre & Barlow, 2009). Jitter, or frequency perturbation, is the cycle-to-cycle variation of pitch caused by instability in vibrations of vocal folds. Shimmer, or amplitude perturbation, is the variability of peak-to-peak amplitude, and is strongly correlated with age. Both jitter and shimmer are associated with the hoarseness or roughness quality of a voice. Voice turbulence, breathiness, is likely caused by incomplete glottal closure, and typically displays a higher than normal frequency signal (Vipperla et al., 2010).

Though the literature demonstrates differences between voices of younger and older adults, its impact on speech recognition systems is scarcely documented. Typically, “corpora”, structured bodies of text, used for research and model estimation with Automatic Speech Recognition (ASR) systems are not representative of older adult speech, which can impact the system’s accuracy for older users (Benzeghiba et al., 2007). One ASR system used with older adult voices (aged 60+) demonstrated a word error rate increase of 50%, when compared to younger adults (aged 18-59), even after training and modifying the system (Wilpon & Jacobsen,1996). Another study showed that word errors increased more rapidly for those above 65 years of age (Vipperla et al., 2008). More recently, older adults were shown to have a 10% increase in word error rates compared to younger adults, and error rates were higher for females (13.7%) than for males (8.7%) (Vipperla et al., 2010). Additionally, older people interact with spoken dialog systems (SDS) differently, where they tend to use more words when interacting with the system than younger adults, and will talk to the SDS as if it were human (Möller et al., 2008).

Vision. Usually, it is pertinent to view an interface to interact with a device. Changes in vision can severely impair older adults, creating difficulty in perceiving an interface and, thus, a device’s intended use. Fozard and colleagues (1977) suggest that there are two stages of noticeable decline. The first occurs between the ages of 35 to 45, where there is less light reaching the retina due to smaller pupil size and the eye’s ability to accommodate to focus at short distances decreases. The eye is also slower to adjust focus when shifting between focal depths (Kuwabara, 1975). The second period of noticeable decline typically occurs between the ages of 55 and 65. During this time the previously mentioned problems worsen and are accompanied by changes to the retina and from changes in neural pathways and cortical function (Salthouse et al.,1996). Additionally, the peripheral area of the visual field is reduced and problems with dim lighting are more apparent. The lens continues to thicken, becoming more opaque and yellowed (Charness & Holley, 2004).

Several abilities commonly decline with these age-related differences. The thickening of the lens results in an increase of light scattered in the eye, creating a blurring effect, which has a direct impact on reducing visual acuity, contrast sensitivity and increasing susceptibility to glare (Charness & Holley, 2004). Visual acuity is the ability to detect fine detail, which is measured by letter recognition under fixed conditions. Contrast sensitivity is an individual's ability to detect differences in illumination within gratings, composed of lines varying in their degree of fineness. Contrast sensitivity has been shown to significantly decline between 20 and 50 years of age, and again at 80 years of age, which demonstrates that older adults have difficulty seeing middle and high spatial frequency displays (Owsley et al., 1983). Color recognition, especially for the blue-green wavelengths, begins to suffer at age 20 and becomes much more noticeable by the age of 70 (Domey et al., 1960).

Older Adults and Smartphone Input Methods.

Smartphone design solutions and marketing strategies have overwhelmingly targeted the younger, more dexterous users who are comfortable incorporating technology into their everyday lives (Pedlow et al., 2010). As of late 2012, the landscape of mobile communications changed. Last year, smartphone owners became the mobile majority (Nielsen, 2012a), which shifted the market into the "late majority" stage of the technology adoption lifecycle (comScore, 2013). Late adopters are older, more frugal, and skeptical of technology (comScore, 2013). The number of people in this demographic is growing at an accelerated pace, with older adults projected to account for 20% of the American population by 2030 (Administration on Aging, 2010). The potential exists to equip this audience with smartphones, which could ameliorate social isolation, improve quality of life, and increase security and empowerment (Jerram et al., 2010). Though older adults are more apprehensive about learning novel tasks that involve new technology (Tomporowski, 2003), they have also been shown to favor and adopt technology when the benefits of the technology are perceived to meet their needs for social interaction and autonomy (Brown et al., 1990). In fact, a study by Barrett (2011) reported that 53% of people aged 50 and older are interested in using a mobile device to track their health, and other agencies are investigating the use of smartphones for medication reminders and safety alerts (CTIA, 2011).

However, given pronounced age-related physical and cognitive limitations, this audience will need more than different marketing strategies to win them over. Since older adults are poised to form a larger percentage of our population, it is imperative to supply them with usable devices (Stone, 2008). However, smartphones are largely inaccessible by this group. The small smartphone form factor limits the size of the hardware and touch screen, which in turn limits the size and spacing of physical and software controls, icons, and the displayed font size. Small keys and touch targets can be especially problematic for older adults, resulting in decreased accuracy and slower selection speeds. This problem could be compounded, depending upon the materials used to manufacture the keyboard, the contrast and size of the printed font on the keys, and the crowning (or height) of the keys, all of which can affect findability of individual keys and the ease with which they can be selected. The small touch screen used to display the interface accommodates smaller fonts, which is problematic for individuals with reduced visual acuity.

The screen itself emits light and is a reflective surface, potentially increasing blur for those who have thicker lenses.

Summary.

This section described the backgrounds and research investigating the five most common modes of text input (i.e., physical and onscreen Qwertys, tracing, handwriting, and voice) used on smartphone devices today, as well as the different approaches in methodology and metrics used in these evaluations. It has been demonstrated that the results across input methods are not only quite variable in metrics and methodology, but also in the participants' demographics and levels of expertise with the input methods and the software and/or hardware tested. Reported entry and error rates are inconsistent and incomparable across studies. This uncertainty regarding which input methods are more favorable is further complicated when trying to generalize their use by older adults. The reliance on proprioceptive cues by older adults may result in an advantage for physical keyboards, which provide rich kinesthetic information, over touchscreen methods. In addition, the changes in the physiology of voice structures may lead to higher levels of inaccuracy with voice recognition systems. Many of the limiting factors, cognitively and physically, may create greater difficulty in learning how to operate the input methods, as well as slower entry rates, higher error rates, and lower levels of perceived satisfaction.

CHAPTER 3

PURPOSE AND RESEARCH QUESTIONS

While many studies have been completed to investigate various performance metrics using different input methods, these studies are not directly comparable. Within this body of literature, the speed, accuracy, and efficiency metrics used to evaluate input methods are inconsistent. The examined devices and input methods run the gamut in size and mechanical properties. Variability exists in the selected test stimuli and protocols for gathering data, including methods for generating the transcribed string, the number of transcriptions used in data analyses, and the length of time participants spend completing their tasks. User samples are frequently small and are typically conducted between-groups, introducing additional variability in motivation, ability, attitudes, fatigue, and experience. Studies habitually ignore novices and older users. Most studies compare only a couple of input techniques, and many of the evaluated technologies are no longer current. Voice recognition as an alternative to the dominant manual entry techniques for smartphone devices has not been investigated. Aside from performance measures, very few studies report user satisfaction, workload, or preference related to initial exposure with these input methods.

Two studies were conducted to investigate the relationship of five current text entry methods with novice user performance, satisfaction, workload, and preference on the same smartphone device. The purpose of the first study was to investigate whether significant relationships exist for these measures within a sample of younger users (aged 18 – 35). In addition, individual characteristics which could potentially influence the collected data (e.g., baseline keyboarding speed, speech rate and voice characteristics, hand anthropometry and function) were also assessed. The second study was identical to the first study, except a sample of older users (aged 60 – 82) was evaluated.

CHAPTER 4

COMMON METHODOLOGY

Though Study 1 and Study 2 examined two different demographics, they were nearly identical in the methodological components of the materials, dependent measures, and procedure. The same protocol was also used during the data screening and cleaning stage, for identification of outliers and for handling non-normal data distributions. Additionally, the same alpha correction procedure was used. This chapter provides the methodology common to both studies. Any differences will be noted in subsequent chapters for each respective study.

Method

Materials.

Questionnaires. An online background survey, the Attitudes Towards Computers Questionnaire (ATCQ) (Jay, 1989), and a modified System Usability Scale (Brooke, 1996) were administered using Google Forms. Workload was assessed using an online version of the NASA Task-Load Index (NASA-TLX) (Hart & Staveland, 1988). These questionnaires are listed in Appendix A. Preference for each input method was assessed using a printed 50-point scale and printed cards displaying text descriptors of the input methods pre-test and screenshots of the input methods post-test (see Appendices B through D). Comments regarding “likes”, “dislikes”, and “suggested improvements” were typed by the facilitator.

Screening Tools. Binocular near visual acuity was assessed using a Good-Lite® Sloan letter eye chart with an affixed string, held at a testing distance of 16 inches. Digit sensitivity for the thumb, index and pinky fingertips was assessed using Semmes-Weinstein monofilaments. The hand set included the 2.83, 3.61, 4.56, 4.31 and 6.65 level monofilaments, corresponding to .07, .4, 2, 4, and 300 grams of force, respectively. A folded hand towel was used to stabilize the participant’s hand during the test, and a printed diagram and colored pencils, corresponding to different levels of force, were used to record sensitivity at each location on a printed form (see Appendix E).

Baseline Evaluations. Prior to testing, several baseline measures were recorded. The intent was to assess whether certain characteristics might be related to data obtained on the dependent measures. Areas of interest included perceptual and cognitive abilities, handwriting, hand function, dexterity and anthropometric measurements, keyboarding and motor speeds, speaking rate and voice characteristics. More detailed information on these measures is given in Appendices E through G.

Visual scanning, visual memory and visual-motor processing were measured using the Digit Symbol Substitution (SDS) test, printed on a sheet of paper and completed with a pencil. Time was kept using a stopwatch (see Appendix H).

Baseline handwriting, hand function, and dexterity were assessed with several measures. To gain insight into participants’ natural letterform style, they wrote each letter of the alphabet in upper and lowercase with a pencil on a blank sheet of paper. Hand function was assessed using Jebsen-Taylor sentences (Jebsen et al., 1969), printed on a notecard, copied by the participants in cursive onto a sheet of paper with a pencil. The Grooved Pegboard test (Trites, 1989) was used to examine dexterity for both hands. Time for the Jebsen-Taylor and the Grooved Pegboard tasks was kept using a stopwatch. A Functional Range of Motion (FRoM) test (Gilbert et al., 1988) assessed the range of motion and digit reach of the left and right thumbs along a 180 degrees arc. The arc was printed on a board, which was the same depth as the smartphone device. Thumb location was marked on the board using a dry erase marker, and thumb positions were recorded using a ruler and protractor. Anthropometric measurements were gathered using digital calipers and a flexible measuring tape. The anthropometric measurements used in this study were the same as a set of measurements reported in a study by García-Cáceres and colleagues (2012).

Baseline keyboarding speed (WPM) was assessed by gathering the average time taken to type three pangrams in lowercase (e.g., “the quick brown fox jumped over a lazy dog”) into a Microsoft Word document using a standard computer keyboard. Motor speed was assessed using a modified Finger Tapping Test (FTT) on a laptop keyboard. Participants tapped the space bar with their left and right thumbs, independently, for 10-second intervals. Time was kept for each of these tasks using a stopwatch.

Participants read Grandfather Passage (VanRiper, 1963) printed on a sheet of paper. Speech rate was calculated by time to read the passage aloud. Voice characteristics were captured by the Multi-Dimensional Voice Program (MDVP) Model 5105 (KayPENTAX, 2008) running on a Dell Vostro 1400 laptop, using a SHURE PG81 microphone, and M-Audio MobilePre USB 200F amplifier.

Smartphone Device and Input. Participants used a Motorola Droid 4, running Android OS 2.3.5 (also known as Gingerbread). The device’s form factor is 69.9 x 126.7 x 12.9 mm, with a 4-inch TFT LCD display (540 x 960 resolution). Screen protectors were not used during the study, however, the screen was wiped with a soft cloth between input methods. The smartphone’s physical Qwerty keyboard and the FlexT9 Text Input (Nuance Communications, 2011) on-screen Qwerty (version XT9 V08.00.00.07), tracing (version XT9 V08.00.00.07) , handwriting (version T9WRITE V05.02.01) and voice input (Dragon Version 2) modes were evaluated. All physical input methods were evaluated in landscape mode. The presentation order for the input methods was partially counterbalanced. Settings for each input method are given in Appendix I.

Phrases. The performance test stimuli were phrases taken from the list developed by MacKenzie and Soukoreff (2003), see Appendix J. These phrases were chosen because they are comprised of words most frequently used in English, and the phrases are commonly used as stimuli in text entry research. The Text Entry Metrics on Android (TEMA) software application (Castellucci and MacKenzie, 2011) was used to randomly present phrases and collect performance data for all of the input methods. In order to keep the phrase available while the participant entered text, the appropriate phrase was retrieved from a dropdown box in Microsoft Visio and displayed on a 19” monitor in front of the participant.

Recording Software. Back-up recordings of phrase entry and participant comments were captured using a Microsoft LifeCam Cinema webcam and Morae Recorder® Version 3.2.1 (TechSmith, 2011), as well as a Sony Handycam Hybrid HDR-SR11.

Sound Level. Background noise level was recorded at the beginning of each set of voice input method trials using a Digital Instruments SL-814 sound level meter. The level of background noise remained constant during all testing sessions.

Procedure.

Participants completed the background questionnaire prior to participation in the study. Before each session, the facilitator recorded environmental noise levels using the sound level meter. Background noise was shown to fluctuate slightly, from 49 to 53 dB, across participants. Additionally, user data stored on the phone was deleted between participants.

Upon arrival, participants signed a consent form (see Appendix K) and were read an introduction to the study, informing them about the baseline measures that would be collected and that they would evaluate five text input methods in the study. They were also informed that they would be receiving breaks periodically, and that they could request more if they desired.

First, the screening measures were administered. Participants' binocular near visual acuity was recorded, and then their thumb and fingertip sensitivity was evaluated, using the monofilaments.

Next, the Digit Symbol Substitution test was administered, allowing participants to code as many symbols as they could within 2 minutes. Then, handwriting and hand function were evaluated. Participants were asked to write the alphabet in uppercase, then in lowercase, writing letterforms as they normally would. They were asked to write one Jebsen-Taylor sentence as quickly and clearly as they could, in cursive, with their non-dominant hand, and repeating the task with a different sentence using their dominant hand. The Grooved Pegboard dexterity task was administered, first using the participants' dominant hand and then using the non-dominant hand. Participants were handed a board with a printed 180 degrees arc, and told to hold the board with their thumbs placed in a comfortable position. Then, they were asked to pivot their thumbs to the 0˚, 45˚ and 90˚ (if possible) marks, and then as low as they could comfortably manage. Functional range of motion and digit reach was recorded for both hands at each position. Anthropometric measurements (hand breadth, hand length, thumb and index width, length and circumference) were taken and recorded for each hand. Manual motor speed was captured by asking participants to hit the space bar on a laptop keyboard with their index finger as quickly as they could for 10 seconds, without moving the rest of their hand. They completed 5 trials for each hand, resting 30 seconds in between each trial. After every third trial, they received 60 seconds of rest.

Voice characteristics were captured in MDVP by asking participants to hold the microphone 6 inches away from their mouth and to voice and sustain a flat tone utterance for 4 seconds. They repeated this three times. Keyboarding speed was assessed by asking participants to key three pangrams, time was recorded for each independently. Next, they were asked to read the Grandfather Passage at their normal rate of speech. Time to read the passage was recorded. Finally, participants completed the Attitudes Toward Computers Questionnaire. This portion of the study lasted about 45 minutes.

Following a break, participants were read instructions regarding the performance portion of the study. Participants were given brief descriptions of each of the five input methods in the order that they were to be presented. They were then asked to rate each input method, using cards with one-word descriptors of each input method, on a 0 – 50 point scale (0 = least preferred and 50 = most preferred) based on their initial impressions of how desirable or valuable they perceived each method to be for entering text on a smartphone. Comments were recorded by the facilitator.

During the text entry portion of the study, participants were told that they would enter 20 phrases with each input method, and the first five phrases would be practice. They were told to input the text as quickly and accurately as they could, without capitalization, punctuation or shorthand. They were allowed to make mistakes, however, they were asked to keep the accuracy of their messages at a level they would feel comfortable sending to a friend. Participants were allowed to hold the device where it was most comfortable for them, as long as it remained in the landscape position for all of the physical methods. For the voice input method, participants were instructed to hold the phone at a comfortable distance in portrait mode and to speak as they normally would. Prior to the block of trials for each method, the facilitator gave more detailed use descriptions and demonstrated each input method by entering the phrase “hello world”. Detailed descriptions for each input method are given in Appendix L. Following the block of trials for each input method, participants completed the satisfaction and workload questionnaires, and were probed to discuss their likes, dislikes and suggested improvements for the method.

After all trials for each input method were complete, participants again rated their preference of the input methods using screen shots along the same 0 – 50 point scale, and again, comments were recorded by the facilitator. This portion of the study lasted between 1 ½ to 2 hours, including time taken for breaks. The facilitator's guide for both studies is given in Appendix L.

Dependent Measures.

Adjusted Words per Minute**.** Adjusted Words per Minute were calculated across all of the input method conditions (Physical Qwerty, Onscreen Qwerty, Tracing, Handwriting, and Voice). Time on task was captured using the TEMA software and video. TEMA logs for each of the input methods for each participant were reviewed. Time recorded for text entry of each phrase was examined by watching video and using a stopwatch. Text entry duration for each trial began when the participant's finger touched the screen (or hit a key, as with the Physical Qwerty keyboard and Voice Recognition launch icon), and ended when the final word or character appeared on the screen. Any duration of time in which participants were talking to the experimenter, fidgeting or looking at the phrase on the monitor was trimmed from time on task. Time was also trimmed if the participant accidentally navigated out of the TEMA software, shifted into a different text entry mode (e.g., from Onscreen Qwerty to Handwriting), or had server connection problems (as with the Voice condition).

Adjusted Words per Minute was calculated using equation (2.3), which takes into account the number of characters in the phrase, the time required to transcribe it, and corrects for inflated entry rates, due to remaining transcription errors. All of the manual input method Words per Minute were figured using equation (2.1). Since Voice produces words as a unit, rather than by character, Adjusted WPM was calculated from figuring WPM by number of words in the presented phrase, rather than characters, and using Word Error Rate to adjust entry speed. It should be noted, however, that there were no significant differences between the calculations of Voice WPM completed at both level. From this calculation, the same adjustment was made as with the manual input methods for figuring Adjusted WPM.

As the first five trials of each input method were considered “practice”, these trials were not included in the analysis. Additionally, due to the nature of the TEMA software, it was possible for participants to enter a phrase before they had made desired corrections by hitting the “Enter” or “Done” buttons. Often, participants would verbally express that they had accidentally entered the phrase. These trials, making up .91% of all of the non-practice trials, were discarded.

***Error Rates*.** Since only the manual input methods produce words at the character-level, the Physical Qwerty, Onscreen Qwerty, Tracing and Handwriting methods were compared with the character-level error rates using the Unified Error Metric (Wobbrock, 2007) for Corrected Error Rate, Uncorrected Error Rate, and Total Error Rate. Formulas for these methods are given in equations (2.7 – 2.9). “Errors” were constituted as any incorrect character, or omission of an intended character, occurring in the duration of composing the transcribed phrase. The TEMA logs and videos for each participant were examined, and each error was recorded, categorized and tallied. Occasionally, people would misspell a word. If this was noticed by the experimenter, the transcribed characters were not counted as errors. It should be noted that the autocorrect feature was enabled, and set to “Low”, during the study. If a word derived from the autocorrect feature did not share 50% of the letters with the intended word, all of the letters in the autocorrected word were counted as errors (the number of characters that should have been in the transcribed word added to the number of extra characters appearing in the word, if any). This calculation is derived from typical error categorization schemes (substitutions, omissions and insertions). If most of the letters in the transcribed word were shared with the intended word, only the erroneous characters were counted.

The Voice input method produces text at the word-level, and so rather than counting errors in the input string they were instead counted in the transcribed phrase. In order to compare the number of errors remaining in the transcribed string across input methods, errors were counted at the word level. Errors were defined as any character in the transcribed phrase that was different from characters in the intended phrase, and there need only be one erroneous character within a word to result in a word error. Extra spaces were not counted as errors, however, an omission of a space was counted as one word error. Occasionally, in the Voice condition, numeric output would be substituted for a word (e.g., “1” instead of “one”). This was not counted as an error. As Word Error Rate is a reflection of the number of errors left in the final transcribed phrase, it should be noted that it does not capture the number of errors created in the input string.

Satisfaction. Satisfaction was evaluated across all of the input method conditions, using a modified version of the System Usability Scale (SUS). The SUS (Brooke, 1996) is a 10-item survey, with each item measured on a 1 – 5 Likert-type scale (1 = strongly disagree, 5 = strongly agree). Once the data is obtained, 1 is subtracted from the odd-numbered data, and 5 is subtracted from the even numbered scores, which reverse-scores these items. The recoded data is summed, and then multiplied by a constant of 2.5. The range of satisfaction scores ranges from 0 – 100, and the higher score indicates higher levels of perceived satisfaction with the system being tested. This instrument is typically used to measure satisfaction in usability studies, and it has been found to be reliable (Tullis & Stetson, 2004). It has been implemented in hundreds of studies, and has been shown to have an average score of 68 (Sauro, 2011). A grading system has been suggested, which assigned grades "A" through "F" to percentiles of the overall distribution of SUS scores from the collection of studies using the measure. It was suggested that SUS scores above 80.3 were considered as achieving an "A" and scores below 51 were considered an "F" (Sauro, 2011).

Workload. Workload for all text input methods was obtained using the NASA Task Load Index (NASA-TLX), which captures six dimensions of workload on a 1 – 21 Likert-type scale (Hart & Staveland, 1988). The dimensions of workload are Mental, Physical, Temporal, Performance, Effort, and Frustration. This instrument has been found to have excellent reliability in its traditional form, which includes participants' paired comparisons of "importance" of the dimensions to assign weights to the dimension scores for an overall measure of workload, and in its modified form without the paired comparisons. This study used the modified administration of the NASA-TLX items. Descriptive statistics are presented for each of the individual scales. A reliability analysis was completed with the six dimensions for each of the input methods. Reliability was acceptable for the items for each of the input methods, thus, a composite score was created by summing the workload ratings for each item for each input method. Cronbach's alphas (unstandardized) for the workload dimensions for each input method with each group are given in Table 4.

TABLE 4

RELIABILITY OF WORKLOAD COMPOSITE SCALES

	Physical Qwerty	Onscreen Qwerty	Tracing	Handwriting	Voice
Younger Adult Group	.851	.771	.891	.831	.713
Older Adult Group	.812	.837	.889	.827	.741

Note: Cronbach's alphas are for unstandardized scores.

Preference. Preference was obtained by asking participants how desirable or valuable each input method was perceived to be on two occasions, pre- and post-test. Participants were given a 0 – 50 point scale (0 = least preferred, 50 = most preferred) and asked to place notecards with a printed one-word descriptor of each input method (i.e., "Tracing") along the scale pre-test and notecards with images of each input method along the scale post-test.

Results

Data Cleaning.

Before running analyses, data distributions for each condition for each dependent measure were examined visually, with the use of histograms and probability plots, and numerically, using measures of central tendency, variance, skewness and kurtosis generated by SPSS. An outlying score whose absolute standard normal value exceeded 2.58 (a value recommended for a sample size smaller than 100) was recoded to 1 unit larger or smaller than the next most extreme score (Fidell & Tabachnick, 2003; Tabachnick & Fidell, 2007). A list of outliers for each study is provided in Appendix M. It should be noted that data were analyzed both before and after outlier removal, and though the results were the same in both instances, outlier removal facilitated fulfillment of the assumptions for selected statistical tests.

Statistical significance of skewness and kurtosis values were also examined using α = .001, recommended for a small to moderate sample size (Tabachnick & Fidell, 2007). Homogeneity of variance was evaluated using Hartley's F_{max} test, which considers the ratio of the largest to smallest cell variance. Ratios not exceeding 10 were considered acceptable, given the equality of n in each cell (Tabachnick & Fidell, 2007). If significant skewness, kurtosis, or F_{max} was greater than 10, data transformations were explored. Distributions of proportional data, all proportions generally below 40%, were transformed with an arcsine square root. In every case that a transformation was applied, the descriptive statistics are reported after back-transformation to the original scale of measurement. Likewise, graphs depicting means and standard error are showing back-transformed values. After executing the parametric model, standardized residuals were examined. Guidelines adopted from Field (2009) suggest a model has an acceptable fit to the data if 5% (or less) of the standardized residuals exceed an absolute value of 1.96, if 1% (or less) exceed 2.58, and that none exceed 3.29. Model fit was acceptable for all dependent measures.

Partial eta squared was the calculated measure of effect size for all of the Analysis of Variance tests within studies 1 and 2. As the effect size given by partial eta squared is equal to that of eta squared for a single factor (Pierce et al., 2004), the magnitude of these effects can be interpreted as .01 for a “small” effect, .06 as a “medium” effect, and .14 as “large” (Cohen, 1988). The effect size magnitude of Pearson’s correlation coefficients were interpreted using the guidelines set forth by Cohen (1988), where .10 is interpreted as “small”, .30 is interpreted as “medium”, and .50 is interpreted as “large.

Alpha Adjustment.

Sixty-one comparisons were warranted by the stated hypotheses for the younger adults, seventy-one for the older adults, therefore, the Holm’s sequential Bonferroni procedure (Holm, 1979) was applied to control for family-wise Type I error across multiple comparisons. Given the exploratory nature of investigating correlations between voice characteristics and Voice input method accuracy, and anthropometric measures with Satisfaction and Total Error Rate for the manual modes of input, the significance level was left unadjusted at $\alpha = .05$, two-tailed, for these tests. Additionally, evaluations made between age groups in Chapter 7 are also exploratory, thus, the significance level for these comparisons is also left at $\alpha = .05$, two-tailed. The hope is that the results from these investigations will aid hypothesis generation for future studies, rather than to make definitive claims regarding the correlations in these studies, as well as the comparisons between studies.

Error Bars.

Both studies had a repeated measures design. Thus the error bars depicted in the bar graphs for each dependent measure represent 95% confidence intervals, calculated to remove within-subject variance. The method for creating the error bars is taken from Cousineau (2005), with the correction proposed by Morey (2008).

Relationship between Measures.

Person's correlation coefficients were calculated to describe the relationships between assessment measures (e.g., cognitive, anthropometric, hand function, etc.) within each gender for each age group. These coefficients are provided in Appendix P.

CHAPTER 5

STUDY 1: SMARTPHONE INPUT METHOD PERFORMANCE, SATISFACTION, WORKLOAD, AND PREFERENCE WITH NOVICE ADULTS AGED 18 - 35

Purpose

Previous research has demonstrated inconsistent results on performance measures between different types of input methods (e.g., Arif & Stuerzlinger, 2009; Clarkson et al., 2005; Kristensson & Denby, 2009; Költringer & Grechenig, 2004; Read et al., 2001; Castellucci & MacKenzie, 2011), and few studies have reported findings specifically regarding smartphone input method performance (e.g., Hoggan et al., 2008; Castellucci & MacKenzie, 2011). Furthermore, the smartphone input method studies used a small number of young adults who had experience with a smartphone physical Qwerty. No studies have directly compared all five of the most common smartphone input methods. None of the studies compared subjective satisfaction or preference ratings for these input methods. Given the variability of the evaluated input methods, participant expertise, and the unknown impact that these input methods may have on the user experience, it was difficult to formulate directional hypotheses, thus, this study was exploratory in nature. This study investigated possible differences between five current smartphone text entry methods on novice adult performance, perceived satisfaction, workload and preference. Additionally, it was of interest to explore whether certain assessments of individual characteristics (e.g., attitudes, cognitive and motor abilities, speech characteristics, etc.) may have been correlated with relevant dependent measures. This was a 5-level, within-subjects design. It was hypothesized that:

H_1: Significant main effects for text input method with performance measures (Adjusted Words per Minute, Corrected Error Rate, Uncorrected Error Rate, Total Error Rate, and Word Error Rate) would be found.

H_2: Significant main effects for text input method with subjective measures (reported Satisfaction, Workload and Post-Test Preference) would be found.

H_3: Significant relationships between individual assessments and the dependent measures would be discovered, where:

- H_{3a}: Baseline hand function (Jebsen-Taylor sentences) would be significantly correlated with Handwriting Adjusted Words per Minute.
- H_{3b}: Baseline keyboarding speed would be significantly correlated with Physical and On-Screen Qwerty Adjusted Words per Minute.
- H_{3c}: Larger hand and digit dimensions would be significantly correlated with Total Error Rate, for physical modes of input.
- H_{3d}: Larger hand and digit dimensions would be significantly correlated with satisfaction, for physical modes of input.

H_{3e}: Speech characteristics (e.g., speech rate, Jitter, Shimmer, etc.) would be significantly correlated with Voice Adjusted Words per Minute and Voice Word Error Rate.

Method

Participants.

Twenty-seven participants were recruited through the Wichita State University SONA system. Flyers were posted at Wichita State University and Barton County Community College. Additionally, snowball sampling occurred through Facebook, email, and word of mouth. Participants were compensated $30 cash. Three participants' data was discarded, as one participant was unable to adequately interact with the smartphone device, one participant was unable to adequately read the words in the task's presented phrases, and one participant had misrepresented his level of expertise with many of the evaluated input methods.

Twenty-five participants (17 female, 8 male) were retained for analysis. Participants had an age range of 18-35 years (*M* = 24.4 , *SD* = 5.6). Nearly all participants reported they had attended college (*n* = 23), with many reporting themselves as current students (*n* = 16). Most were right-handed (*n* = 23), and did not report any major current finger, hand or wrist problems. None of the participants had current speech disorders, and four participants reported that they currently smoke tobacco on a weekly or monthly basis. All participants were novices to each of the smartphone input methods investigated in this study (for summary of relevant demographics, see Appendix N). All participants had acceptable binocular near visual acuity and fingertip sensitivity.

Materials.

All materials used in this study are described in Chapter 4.

Procedure.

The procedure is explained in detail in Chapter 4.

Results

All data screening, cleaning, and alpha adjustment procedures, as well as guidelines for effect size interpretations, are described in Chapter 4. A list of outliers for each dependent measure is given in Appendix M.

Adjusted Words per Minute.

After applying a square root transformation to the data, a one-way repeated measures analysis of variance was conducted. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(9) = 32.75$, p <.001, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .62$). The results showed a large significant main effect for input method, $F(2.49, 59.85) = 192.03$, $p < .001$, $\eta_p^2 = .89$.

Follow-up paired-samples *t* tests indicated that Voice (*M* = 44.62, *SD* = .58) had a significantly higher entry rate than Physical Qwerty (*M* = 28.84, *SD* = .32), $t(24) = 8.20$, Onscreen Qwerty (*M* = 16.59, *SD* = .63), $t(24) = 12.88$, Tracing (*M* = 20.24, *SD* = .33), $t(24) = 13.69$, and Handwriting (*M* = 9.35, *SD* = .15), $t(24) = 27.58$, all $p < .001$. Physical Qwerty had a significantly higher entry rate than Onscreen Qwerty, $t(24) = 9.62$, Tracing, $t(24) = 14.40$, and Handwriting, $t(24) = 20.71$, all $p < .001$. Tracing and Onscreen Qwerty had significantly higher entry rates than Handwriting, $t(24) = 12.38$, and $t(24) = 6.80$, respectively, both $p < .001$. Finally, Tracing had a significantly higher entry rate than Onscreen Qwerty, $t(24) = 3.40$, $p = .002$ (see Figure 5.1).

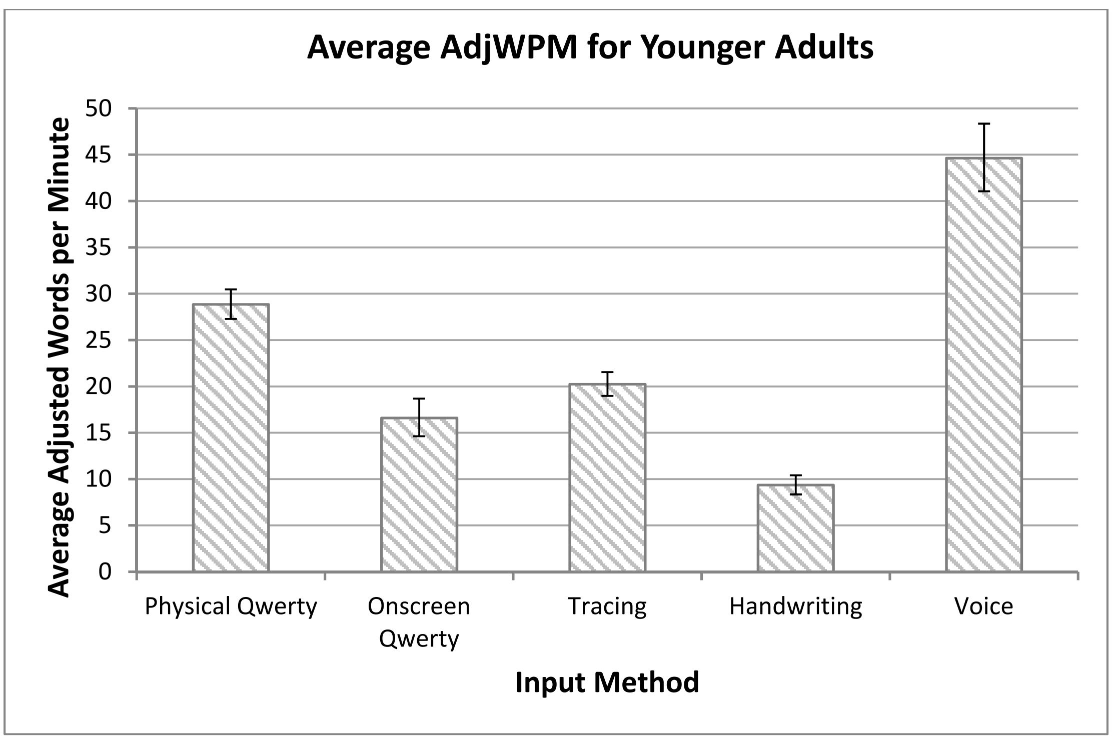


Figure 5.1. Average AdjWPM for younger adults, values are back-transformed. Voice had the highest entry rate, and Handwriting had the lowest. Bars indicate 95% CI.

Error Rates.

The manual input methods (Physical and Onscreen Qwerty, Tracing, and Handwriting) all use methods to produce text at the character-level, thus the Unified Error Metric (Corrected, Uncorrected and Total Error Rates) is used to describe accuracy for these. Since the Voice input method is a word-level entry method, the comparison in accuracy across all input methods is only done using the Word Error Rate. Thus, the accuracy of the input stream is not analyzed for Voice.

***Corrected Error Rate*.**

After applying an arcsine square root transformation to the data, a one-way repeated measures analysis of variance showed a large significant main effect for input method, $F(3, 72) = 115.66$, $p < .001$, $\eta_p^2 = .83$.

Follow-up paired-samples *t* tests showed that the Physical Qwerty (*M* = 1.43%, *SD* = 0.31%) input method had a significantly lower corrected error rate compared to Onscreen Qwerty (*M* = 13.04%, *SD* = 1.01%), $t(24) = 12.98$, Tracing (*M* = 4.67%, *SD* = .74%), $t(24) = 5.10$, and Handwriting (*M* = 19%, *SD* = .66%), $t(24) = 18.90$, all $p < .001$. Tracing had a significantly lower corrected error rate than Onscreen Qwerty, $t(24) = 6.63$, and Handwriting, $t(24) = 12.00$, both $p < .001$. Onscreen Qwerty had a significantly lower corrected error rate than Handwriting, $t(24) = 4.33$, $p < .001$. Of the manual input methods for younger adults, Physical Qwerty clearly had the lowest corrected error rate, sequentially followed by Tracing, Onscreen Qwerty, and Handwriting (see Figure 5.2).

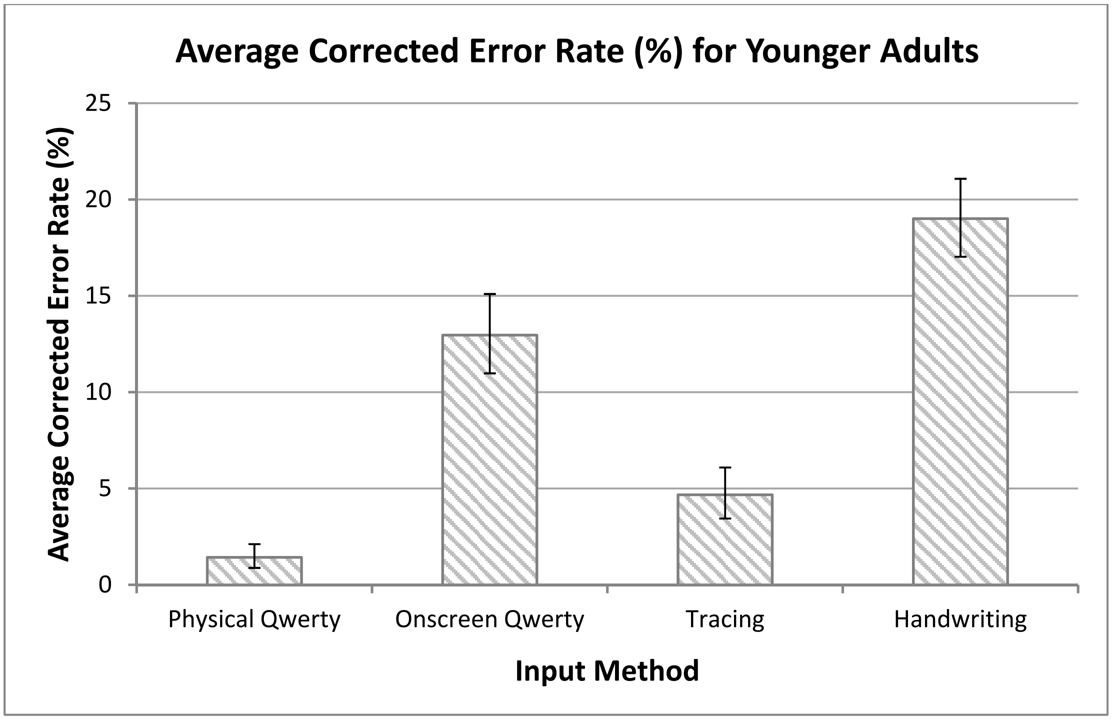


Figure 5.2. Average Corrected Error Rate for younger adults, values are back-transformed. Physical Qwerty had the lowest average Corrected Error Rate. Bars indicate 95% CI.

Uncorrected Error Rate. After applying an arcsine square root transformation to the data, a one-way repeated measures analysis of variance showed a large significant main effect for input method, $F(3, 72) = 12.35$, $p < .001$, $\eta_p^2 = .34$.

Follow-up paired-samples *t* tests indicated that the Physical Qwerty (*M* = .22%, *SD* = .18%) input method had a significantly lower uncorrected error rate compared to Onscreen Qwerty (*M* = .90%, *SD* = .17%), $t(24) = 5.42$, and Handwriting (*M* = 1.13%, *SD* = .27%), $t(24) = 5.18$, both $p < .001$. Tracing (*M* = .38%, *SD* = .18%) had a significantly lower uncorrected error rate than Handwriting, $t(24) = 3.69$, $p = .001$. There were no significant differences between Physical Qwerty and Tracing, $t(24) = 1.18$, $p = .25$, Onscreen Qwerty and Handwriting, $t(24) = 1.06$, $p = .30$, and Tracing and Onscreen Qwerty, $t(24) = 3.08$, $p = .005$. Of the manual input methods for younger adults, Physical Qwerty and Tracing had lower uncorrected error rates, whereas Onscreen Qwerty and Handwriting had higher uncorrected error rates (see Figure 5.3).

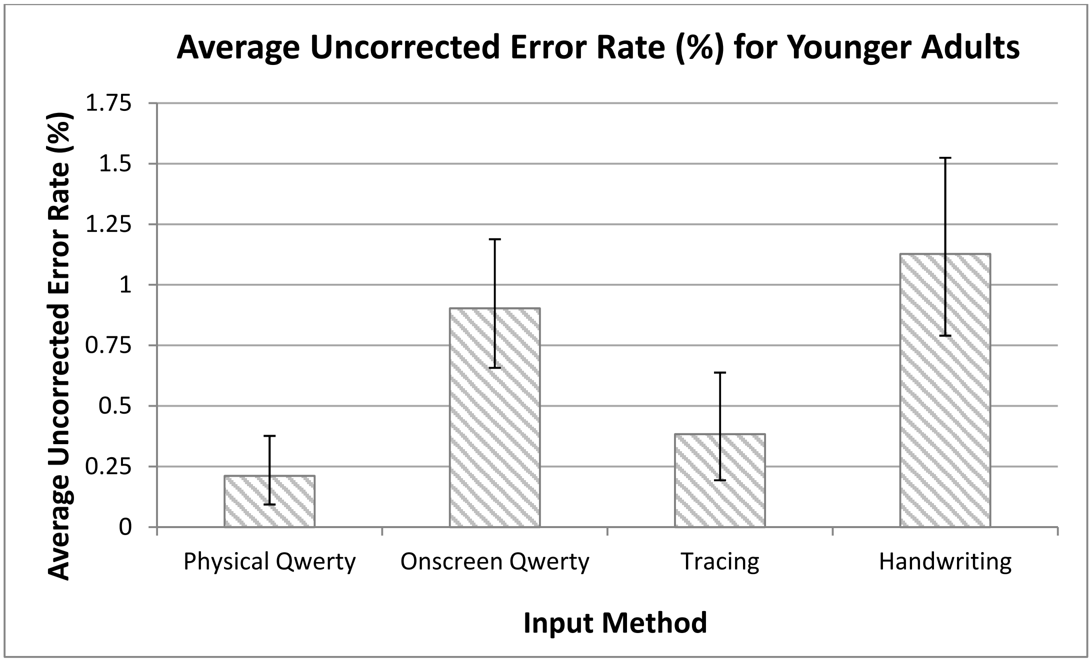


Figure 5.3. Average Uncorrected Error Rate for younger adults, values are back-transformed. Handwriting had the highest UER, Physical Qwerty had the lowest. Bars indicate 95% CI.

Total Error Rate. Following an arcsine square root transformation, a one-way repeated measures analysis of variance showed a large significant main effect for input method, $F(3, 72) = 118.12$, $p < .001$, $\eta_p^2 = .83$.

Follow-up paired-samples *t* tests revealed that the Physical Qwerty ($M = 1.82\%$, $SD = 0.38\%$) input method had a significantly lower average total error rates compared to Onscreen Qwerty ($M = 14.12\%$, $SD = 1.1\%$), $t(24) = 12.20$, Tracing ($M = 5.45\%$, $SD = .49\%$), $t(24) = 5.76$, and Handwriting ($M = 20.38\%$, $SD = .69\%$), $t(24) = 17.45$, all $p < .001$. Tracing had a significantly lower total error rate than Onscreen Qwerty, $t(24) = 6.86$, and Handwriting, $t(24) = 12.54$, both $p < .001$. Onscreen Qwerty had a significantly lower total error rate than Handwriting, $t(24) = 4.43$, $p < .001$. Of the manual input methods for younger adults, Physical Qwerty clearly had the lowest total error rate, sequentially followed by Tracing, Onscreen Qwerty, and Handwriting (see Figure 5.4).

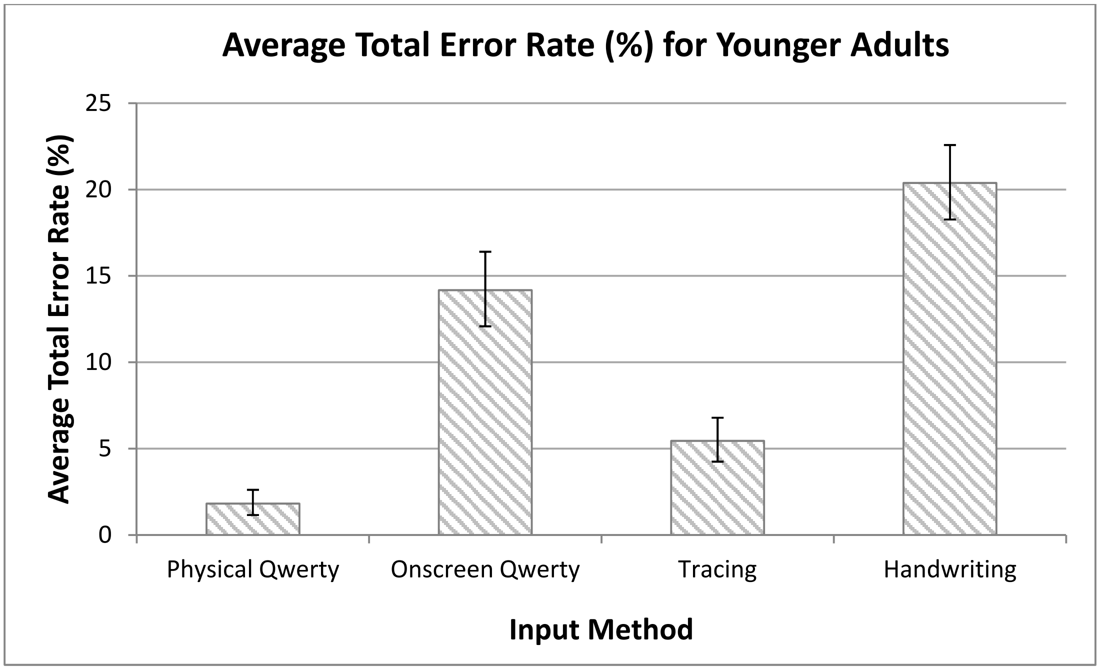


Figure 5.4. Average Total Error Rate for younger adults, values are back-transformed. Physical Qwerty had the lowest average Total Error Rate. Bars indicate 95% CI.

Word Error Rate. Following an arcsine square root transformation, a one-way repeated measures analysis of variance revealed a small, nonsignificant effect for input method, $F(4, 96) = 1.04$, $p = .39$, $\eta_p^2 = .04$, see Figure 5.5.

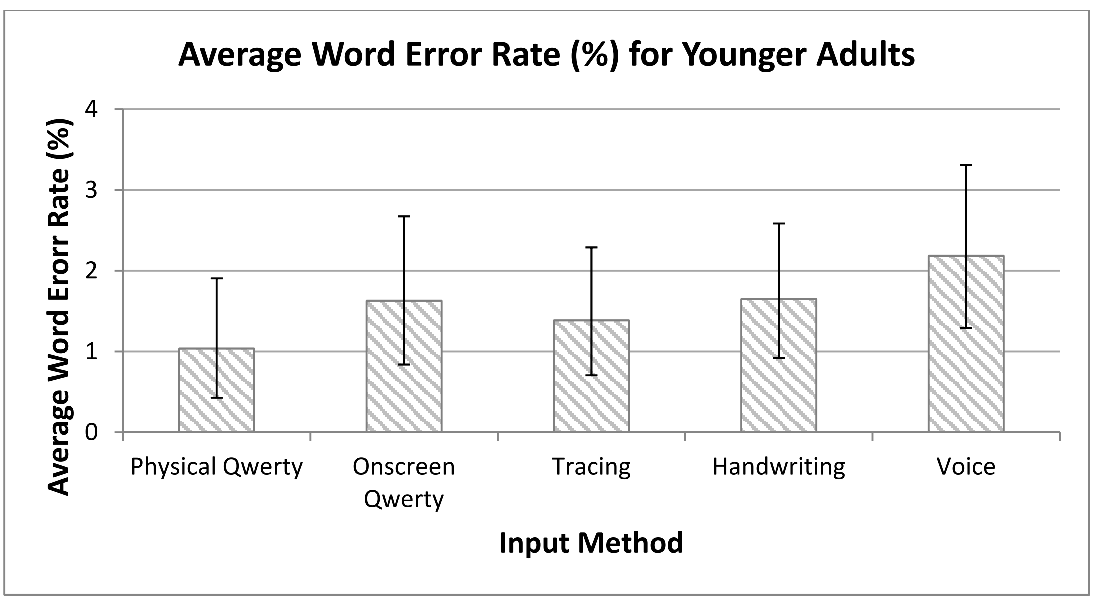


Figure 5.5. Average Word Error Rate for younger adults, values are back-transformed. There was no significant main effect. Bars indicate 95% CI.

Satisfaction.

A one-way repeated measures analysis of variance was conducted to evaluate satisfaction scores across input methods. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(9) = 26.57$, p =.002, therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = .87$). The results showed a large significant main effect for input method, $F(3.49, 83.64) = 29.4$, $p < .001$, $\eta_p^2 = .55$.

Follow-up paired-samples *t* tests indicated that the Voice input method (*M* = 80.4, *SD* = 13.81) was significantly more satisfying than Onscreen Qwerty (*M* = 68.1, *SD* = 20.4), $t(24) = 5.7$, and Handwriting (*M* = 42.2, *SD* = 21.1), $t(24) = 12.38$, both $p < .001$, and Tracing (*M* = 75.5, *SD* = 20.19), $t(24) = 3.37$, $p = .001$. Physical Qwerty (*M* = 88.6, *SD* = 9.11) was more satisfying than Onscreen Qwerty, $t(24) = 8.13$, Tracing, $t(24) = 4$, and Handwriting, $t(24) = 13.56$, all $p < .001$. Onscreen Qwerty was significantly more satisfying than Handwriting, $t(24) = 3.93$, $p < .001$. Tracing was more satisfying than Handwriting, $t(24) = 6.68$, $p < .001$. There were no significant differences in satisfaction between Voice and Physical Qwerty, $t(24) = 1.17$, $p = .25$, and Onscreen Qwerty and Tracing, $t(24) = 1.66$, $p = .10$. It appears that younger adults are more satisfied with the smartphone Physical Qwerty and Voice input methods over the others (see Figure 5.6).

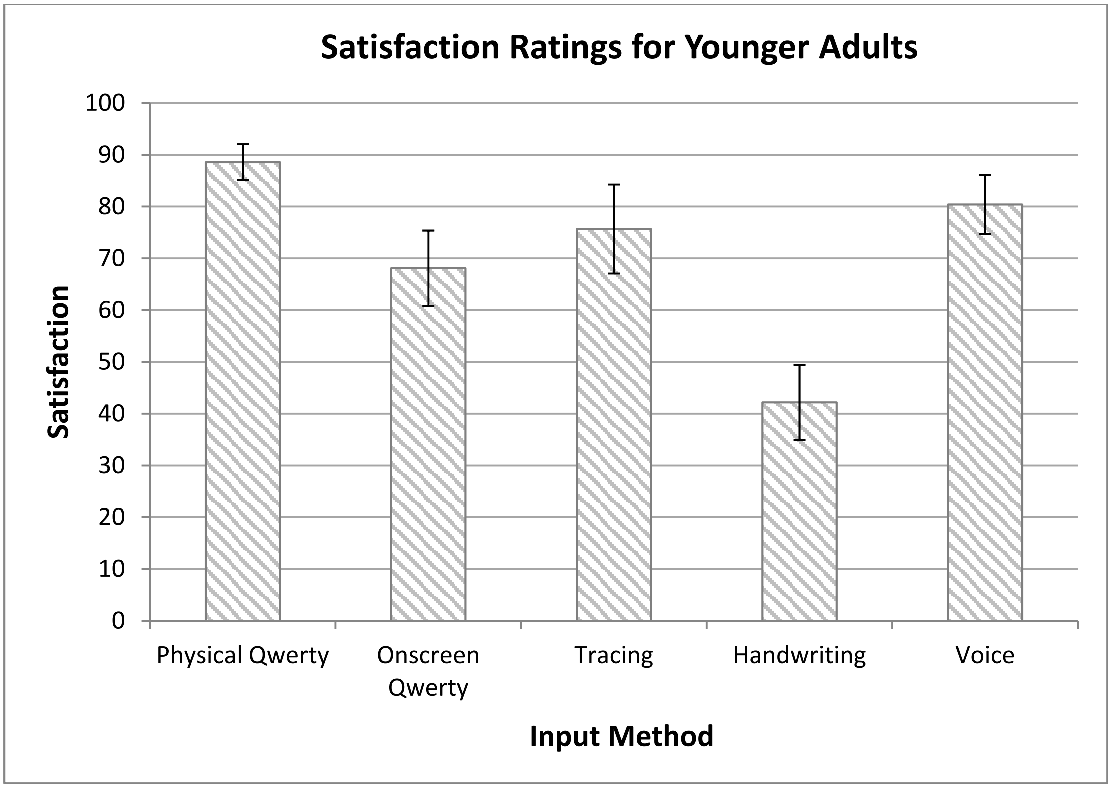


Figure 5.6. Satisfaction ratings for younger adults. Voice and Physical Qwerty elicited higher satisfaction scores, Handwriting was least satisfying. Bars represent 95% CI.

Workload.

Participants rated their subjective level of workload for each of six workload dimensions on a 1 – 21 scale (1 = Very Low, 21 = Very High for every dimension, except for Performance, which was 1 = Perfect, 21 = Failure). Distributions of each dimension were examined independently, and a total of fourteen outliers were recoded (see Appendix M). A reliability analysis was conducted for all of the workload dimensions for each of the input methods. Cronbach's alphas are reported in Appendix A. A composite score for workload for each input method was created by summing the scores for each dimension. Graphs of average ratings for individual workload dimensions for each input method are given in Figures 5.7 through 5.12.

It appears that younger adults perceived the Handwriting input method to be high on each dimension of workload. They reported that the Onscreen Qwerty was also high on the physical, frustration, and performance dimensions, while Tracing was high on the mental workload dimension. Voice and Physical Qwerty scored lower on each dimension of workload, with the exception of Physical Qwerty scoring a little higher on the effort dimension.

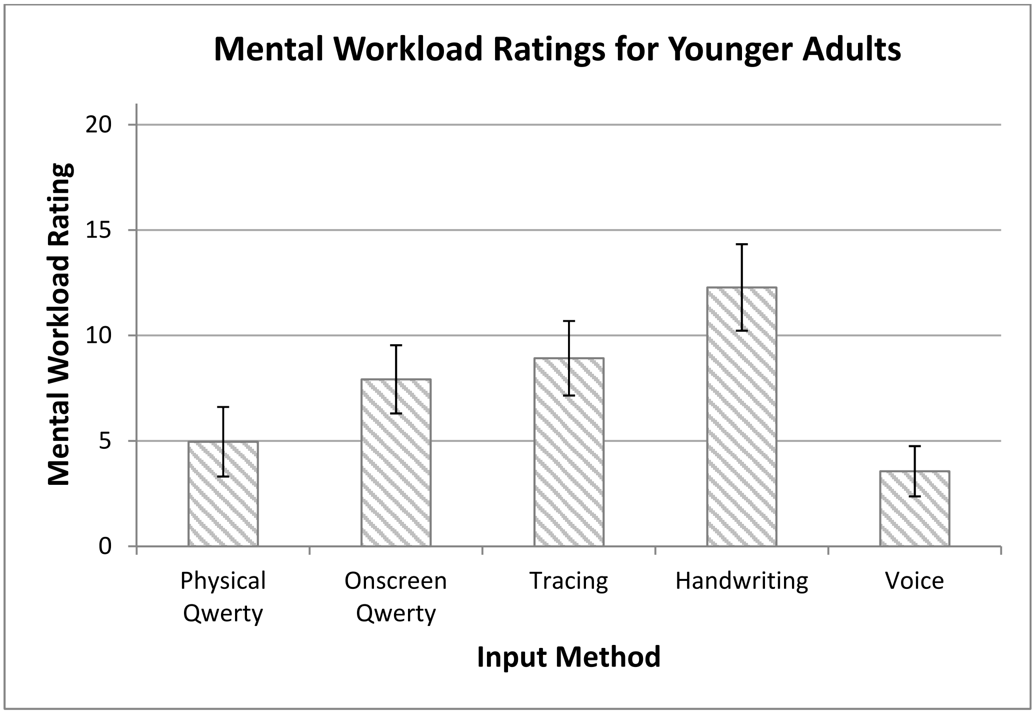


Figure 5.7. Average workload ratings for the mental dimension for younger adults. Lower ratings are better. Error bars indicate 95% CI.

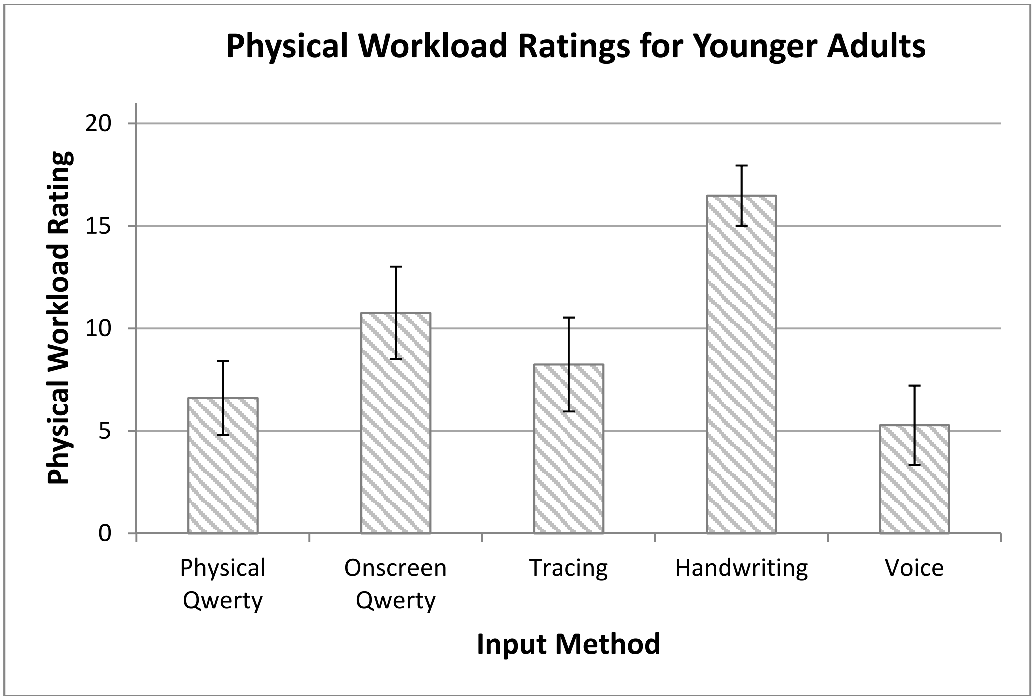


Figure 5.8. Average workload ratings for the physical dimension for younger adults. Lower ratings are better. Error bars indicate 95% CI.

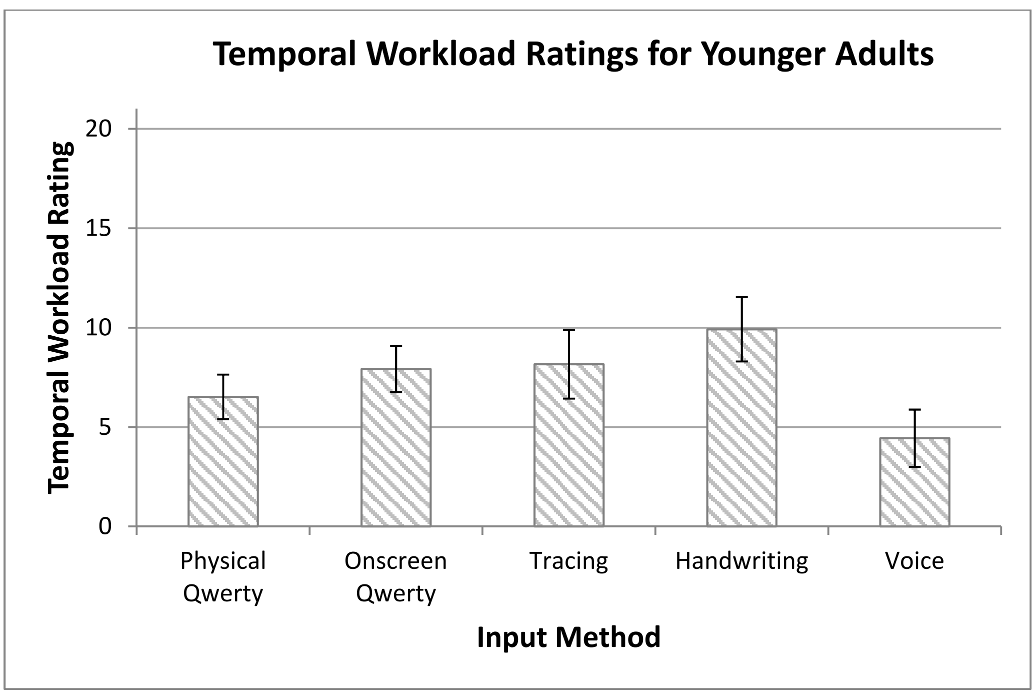


Figure 5.9. Average workload ratings for the temporal dimension for younger adults. Lower ratings are better. Error bars indicate 95% CI.

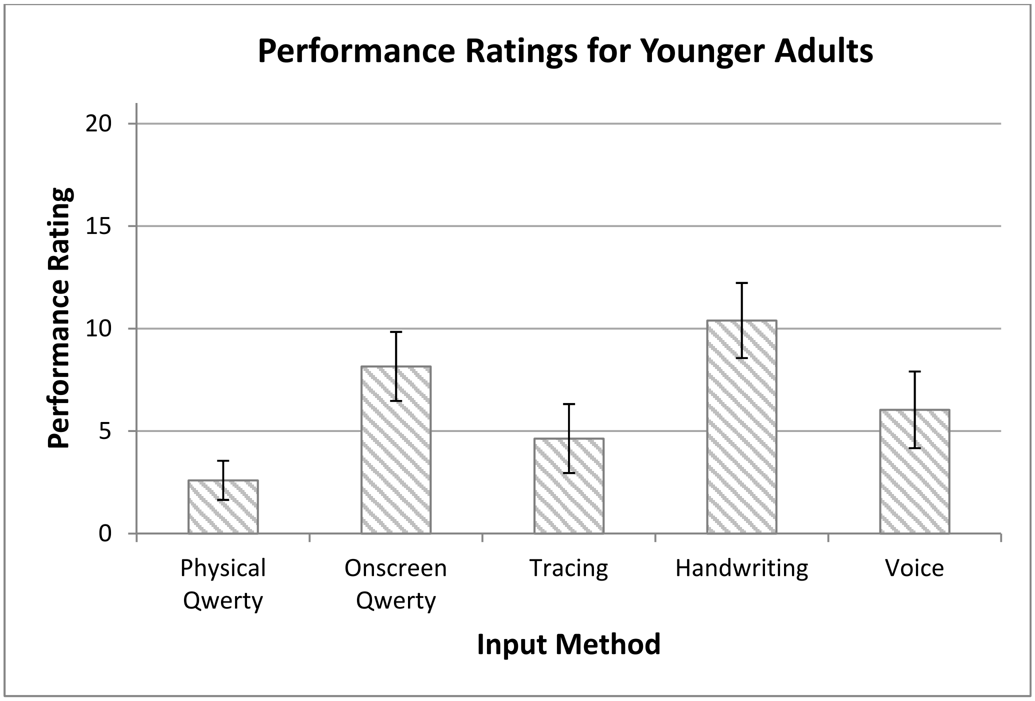


Figure 5.10. Average workload ratings for the performance dimension for younger adults. Lower ratings are better. Error bars indicate 95% CI.

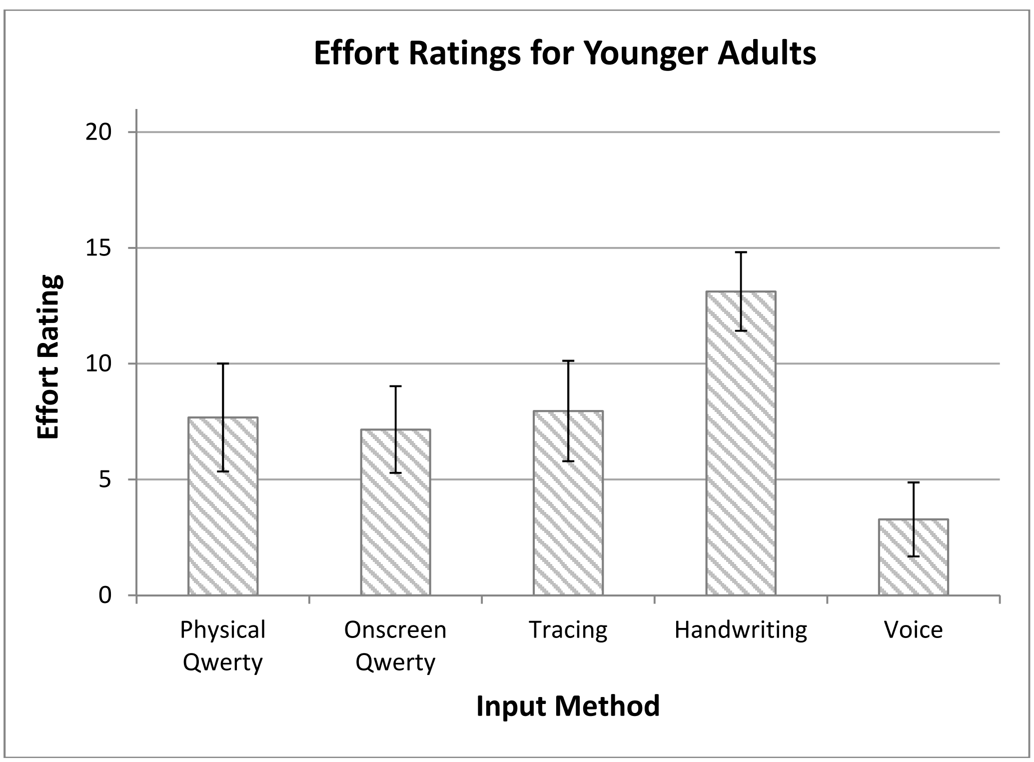


Figure 5.11. Average workload ratings for the effort dimension for younger adults. Lower ratings are better. Error bars indicate 95% CI.

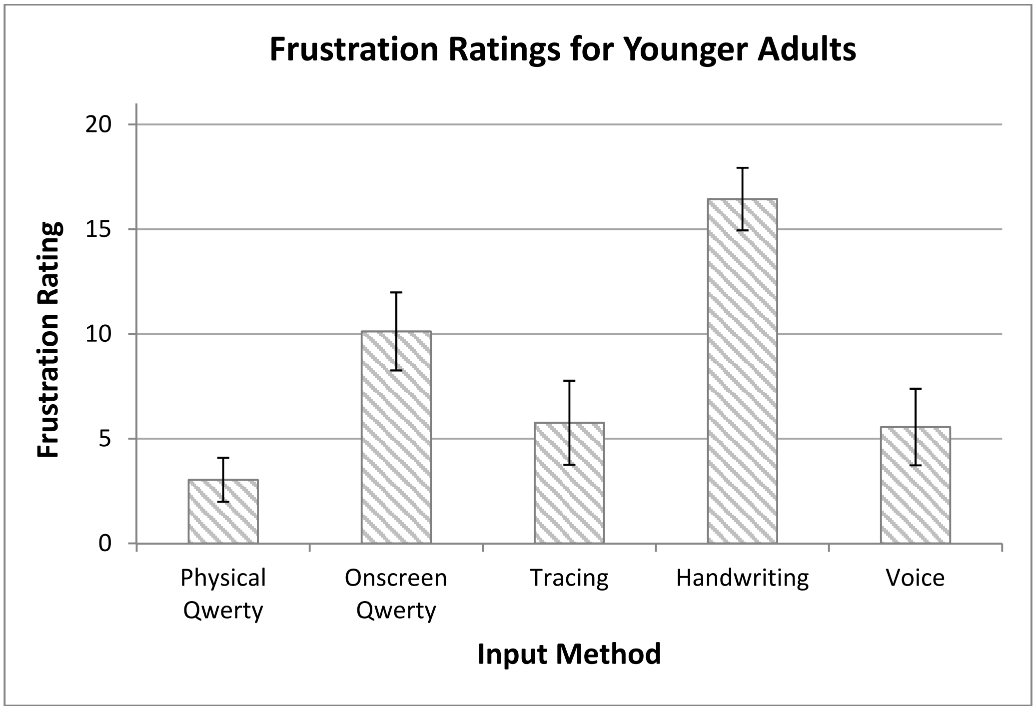


Figure 5.12. Average workload ratings for the frustration dimension for younger adults. Lower ratings are better. Error bars indicate 95% CI.

A one-way repeated measures analysis of variance was conducted to evaluate composite workload scores across input methods. The results showed a large significant main effect for input method, $F(4, 96) = 41.26$, $p < .001$, $\eta_p^2 = .63$.

Follow-up paired-samples *t* tests indicated that the Voice input method ($M = 22.2$, $SD = 11.6$) was perceived as requiring significantly less workload than Onscreen Qwerty ($M = 52.04$, $SD = 22.48$), $t(24) = 6.67$, Handwriting ($M = 78.64$, $SD = 22$), $t(24) = 13.81$, and Tracing ($M = 43.48$, $SD = 23.68$), $t(24) = 3.96$, all $p < .001$. Physical Qwerty ($M = 31.4$, $SD = 19.94$) had lower workload ratings than Onscreen Qwerty, $t(24) = 6.26$, and Handwriting, $t(24) = 10.10$, both $p < .001$. Tracing had lower workload ratings than Handwriting, $t(24) = 6.80$, $p < .001$. Onscreen Qwerty also had lower workload ratings than Handwriting, $t(24) = 5.54$, $p < .001$.

There were no significant differences between Onscreen Qwerty and Tracing, $t(24) = 1.47$, $p = .15$, Physical Qwerty and Tracing, $t(24) = 2.13$, $p = .04$, and Physical Qwerty and Voice, $t(24) = 2.45$, $p = .02$, see Figure 5.13.

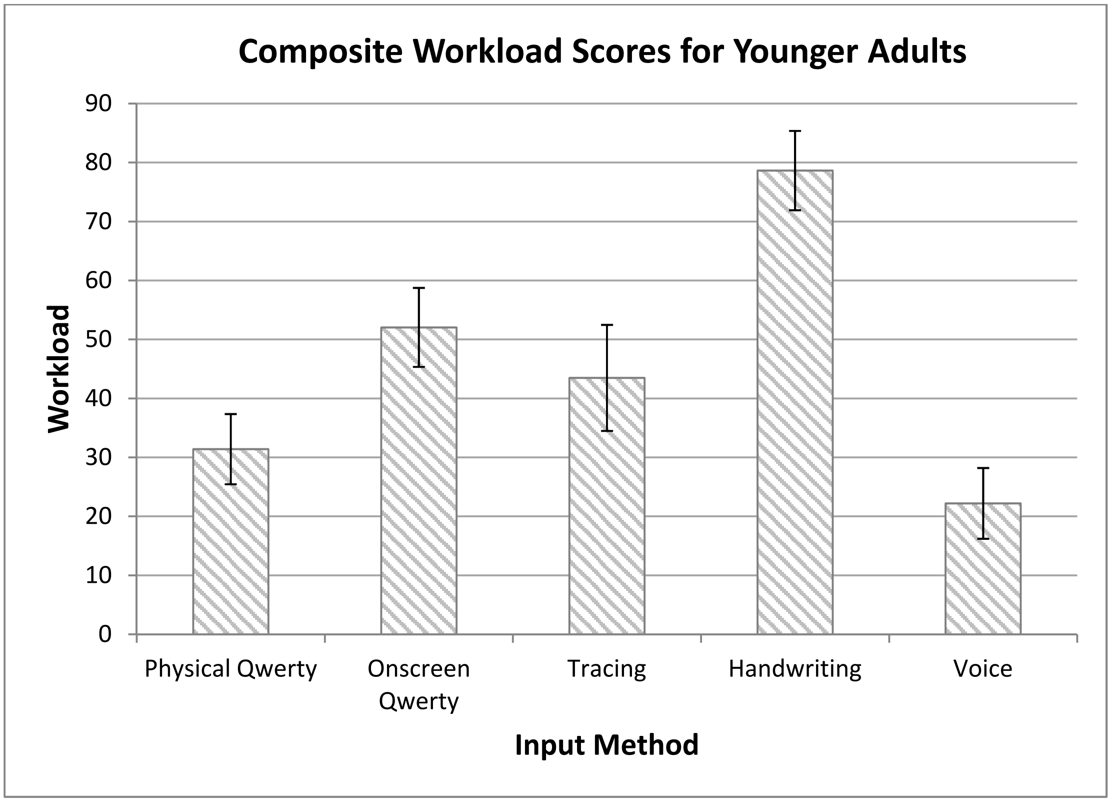


Figure 5.13. Composite workload scores for younger adults. Handwriting elicited the highest workload scores, and Voice the lowest workload scores. Error bars represent 95% CI.

Preference.

Pre-Test. Prior to using the input methods, participants were asked to rate each input method on a 0 – 50 point scale (0 = least preferred, 50 = most preferred) based on how desirable or valuable they perceived it to be for entering text on a smartphone device. The initial ratings are given in Figure 5.14. It appears that younger adults tended to prefer the Physical and Onscreen Qwerty input methods, and tended to least prefer the Handwriting input method.

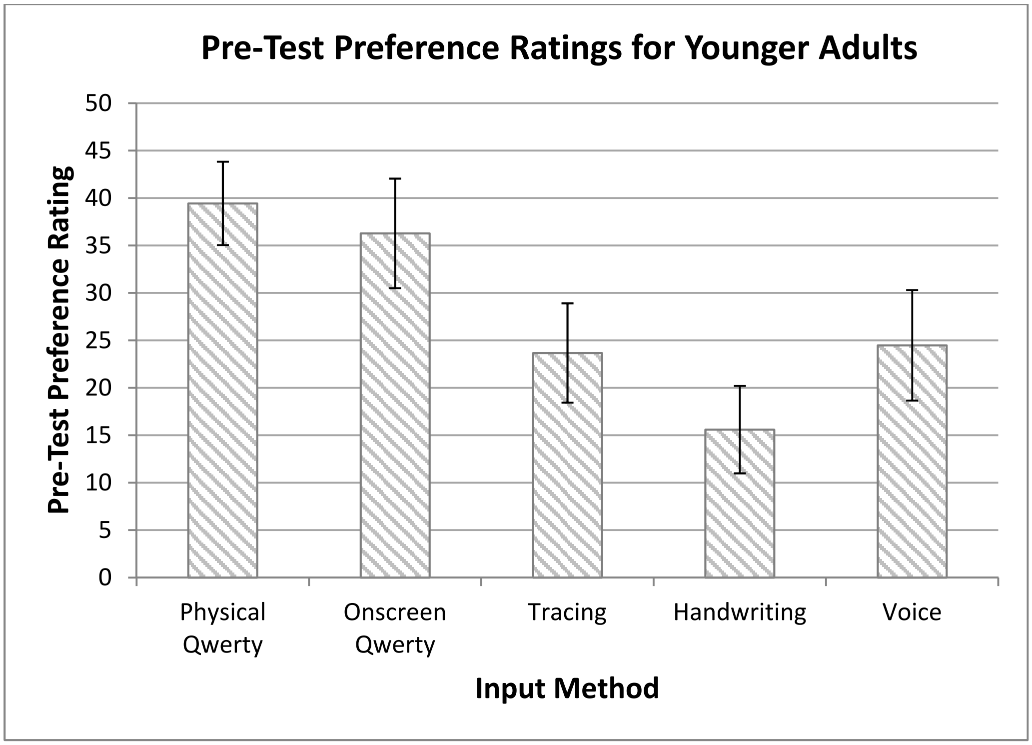


Figure 5.14. Pre-test preference ratings for younger adults. Both Qwerty methods were initially more preferred. Error bars represent 95% CI.

Post-Test. A one-way repeated measures analysis of variance was conducted to evaluate post-test preference ratings across input methods. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(9) = 22.21$, $p = .008$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .66$). The results showed a large significant main effect for input method, $F(2.63, 63.17) = 16.20$, $p < .001$, $\eta_p^2 = .40$.

Follow-up paired-samples *t* tests indicated that the Physical Qwerty ($M = 43.24$, $SD = 6.64$) input method had significantly higher preference ratings than Handwriting ($M = 11.6$, $SD = 11.53$), $t(24) = 11.78$, Onscreen Qwerty ($M = 29.56$, $SD = 14.93$), $t(24) = 4.19$, both $p < .001$. Voice ($M = 34.4$, $SD = 13.69$) and Onscreen Qwerty both elicited significantly higher ratings than Handwriting, $t(24) = 7.85$, and $t(24) = 4.85$, respectively, both $p < .001$. Tracing ($M = 29.08$, $SD = 17.89$) was rated higher than Handwriting, $t(24) = 3.55$, $p = .002$. There were no significant differences in post-test preference ratings between Onscreen Qwerty and Tracing, *t*(24) = .09, *p* = .93, Tracing and Voice, *t*(24) = .97, *p* = .34, Onscreen Qwerty and Voice, *t*(24) = 1.28, *p* = .21, Physical Qwerty and Tracing, *t*(24) = 3.27, *p* = .003, and Physical Qwerty and Voice, *t*(24) = 3.04, *p* = .006. It appears that younger adults most preferred the Physical Qwerty input method for entering text on a smartphone, though Voice, Onscreen Qwerty, and Tracing were also rated positively. Handwriting was clearly the least preferred (see Figure 5.15).

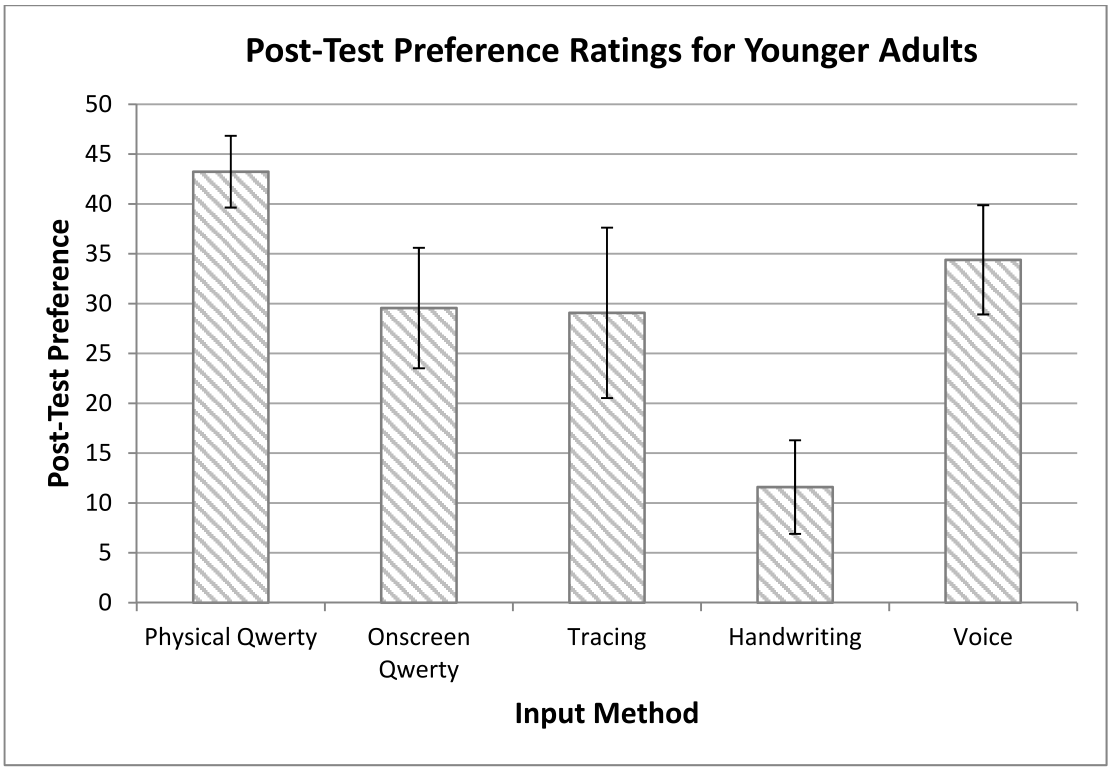


Figure 5.15. Post-test preference ratings for younger adults. Physical Qwerty is most preferred, and Handwriting is least preferred. Error bars represent 95% CI.

Preference Rating Change. Observing the differences in average preference ratings from pre- to post-test, it appears that younger adults had a more favorable view of the Physical Qwerty, Tracing, and Voice after using all of the input methods. Their preference for both Onscreen Qwerty and Handwriting decreased (see Figure 5.16).

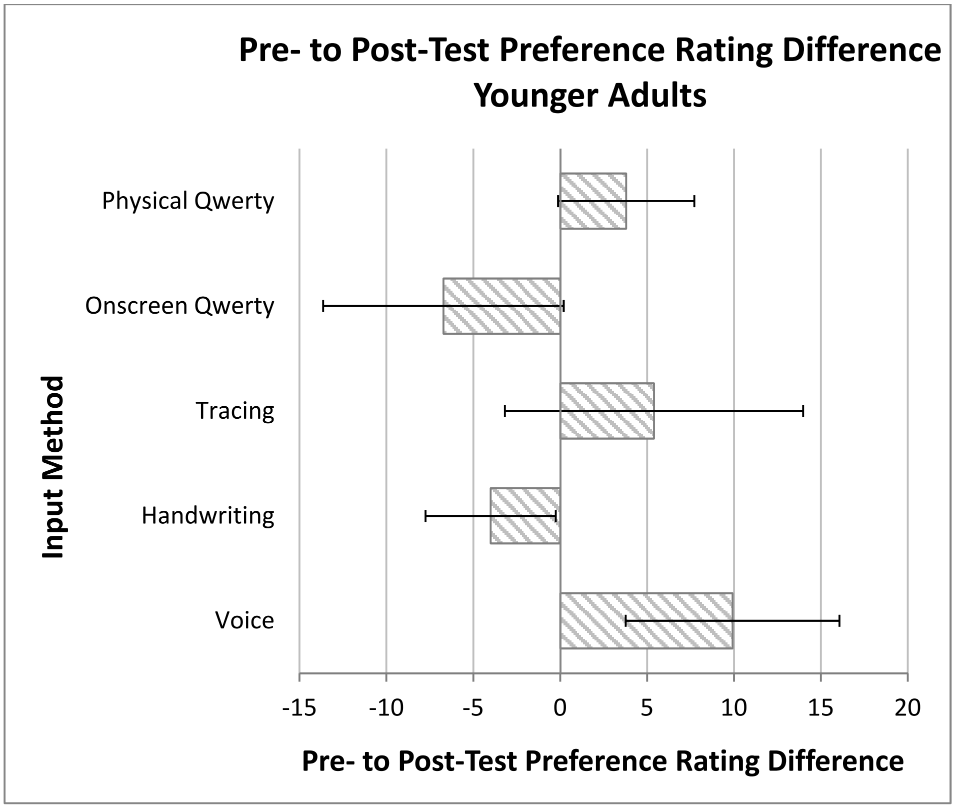


Figure 5.16. Average pre- to post-test preference rating change for each input method. The greatest change was seen with the Voice method. Error bars represent 95% CI.

Participant Comments.

Participant comments regarding initial impressions of each of the input methods were gathered pre-test during the preference ratings (see Appendix B for comment summaries). After completing trials for each method, comments regarding personal likes, dislikes, and suggested changes for each input method were gathered (see Appendix C for summaries of likes, dislikes, and suggested changes). Comments regarding final impressions were gathered post-test (see Appendix D for comment summaries). The most frequently reported impressions for each text input method during the pre- and post-test preference ratings are given in Table 5.1.

TABLE 5.1

PARTICIPANT COMMENT SUMMARY

	Pre-Test	Post-Test
Physical Qwerty	Liked familiarity of the Qwerty layout.	Felt comfortable with the Qwerty layout.
	Physical keys were appreciated for accuracy and tactile quality.	Liked having the feedback afforded by the physical keys.
	Felt they would have more control.	Felt this method was forgiving and accurate, but a little slow.
Onscreen Qwerty	Qwerty layout is familiar.	Buttons were too small, poor accuracy.
	No button pushing may make this method faster, less physically demanding, but also concerns about accuracy and feedback.	It did not meet their expectations. No physical feedback, did not like Autocorrect, and not forgiving. Slower than anticipated.
	Software keyboard looks “cool”.	Pop-up graphics were distracting.
Tracing	Seems foreign and labor intense.	Not entirely intuitive, but easy to learn.
	Concerns about the software “knowing” which letters to include/accuracy.	Autospacing was appreciated, liked the accuracy and low physical demands.
	It seems difficult; would take time to learn.	Took concentration. Frustrated when gestures not recognized by the software.
Handwriting	Concerns about accuracy for those with poor handwriting legibility.	Not very accurate. Many had to change their writing style.
	Skeptical about the quality of the software.	It felt very slow.
	Mixed comments about feeling familiar, comfortable and the perceived entry rates.	Novel, but not very practical.
Voice	Should be fast and easy to use.	Simple, easy, fast and convenient.
	Good for situations requiring multitasking.	Would only use while driving, multi-tasking, or for composing longer messages.
	Concerns about background noise, enunciation, and personal privacy.	Concern regarding background noise and privacy.

Hand Function and Handwriting.

After applying a square root transformation to the Jebsen-Taylor data distribution, a Pearson’s correlation was computed between the Jebsen-Taylor handwriting times for the dominant hand and the average Adjusted Words per Minute for the Handwriting condition. The results of the correlation show a slight positive relationship, which was not statistically significant, $r(23) = .05$, $p = .82$.

Keyboarding and Smartphone Qwerty Speed.

A square root transformation was applied to scores for both conditions. Pearson's correlation coefficients were computed between participants' average Keyboarding Speed (Words per Minute) and average Adjusted Words per Minute for the smartphone Physical and Onscreen Qwerty input methods. The results show a moderate, though not statistically significant, correlation between both the Keyboarding Speed and Adjusted Words per Minute in the Physical Qwerty condition, $r(23) = .49$, $p = .01$, and between Keyboarding Speed and Adjusted Words per Minute in the Onscreen Qwerty condition, $r(23) = .34$, $p = .10$. Participants, who were faster at keying the pangrams using a standard desktop keyboard, tended to also have higher entry rates using the smartphone Physical and Onscreen Qwerty keyboards, though not statistically so.

Hand Function, Anthropometry and Total Error Rate.

Pearson correlation coefficients were computed among averages for each of the hand function, anthropometric measures and average Total Error Rates for each of the physical input methods for each gender. The distribution for the Jebsen-Taylor measure was square root transformed, and the distributions for Total Error Rate were all arcsine square root transformed. Since this was an exploratory investigation of these relationships, the alpha level is not adjusted. The results of these analyses for female participants are given in Table 5.2, and Table 5.3 for the male participants. It appears that longer finger dimensions tend to be related to increased Onscreen Qwerty Total Error Rates for female participants, and that larger hand dimensions and reach tend to be related to decreased Physical and Onscreen Qwerty Total Error Rates for male participants. Additionally, greater Functional Range of Motion tended to be related to increased error rates for males.

TABLE 5.2

HAND FUNCTION, ANTHROPOMETRY AND TOTAL ERROR RATE FOR FEMALES

	Physical Qwerty	Onscreen Qwerty	Tracing	Handwriting
Average Hand Length	-.34	**.54***	.31	.01
Average Hand Breadth	-.30	.36	.23	.16
Average Thumb Length	-.27	.48	.35	.22
Average Thumb Width	-.10	.37	.25	.32
Average Thumb Circumference	-.08	.42	.36	.21
Average Index Length	-.28	**.54***	.11	-.07
Average Index Width	-.09	.37	.25	.24
Average Index Circumference	-.21	.31	.23	.31
Average Thumb Reach	.28	.24	.17	.11
Average Functional Range of Motion	-.12	-.07	.20	.14
Jebsen-Taylor (dominant)	.27	.27	.23	.01
Grooved Pegboard (dominant)	.17	.03	.04	-.35

Note: * indicates significance at the .05 level, two-tailed. *n* = 17.

TABLE 5.3

HAND FUNCTION, ANTHROPOMETRY AND TOTAL ERROR RATE FOR MALES

	Physical Qwerty	Onscreen Qwerty	Tracing	Handwriting
Average Hand Length	-.51	-.52	-.06	.15
Average Hand Breadth	-.60	-.09	-.02	.22
Average Thumb Length	-.51	**-.79***	.12	-.17
Average Thumb Width	**-.74***	.11	.26	-.02
Average Thumb Circumference	-.49	.21	.08	-.10
Average Index Length	-.24	-.69	-.25	.04
Average Index Width	**-.74***	.08	.07	.16
Average Index Circumference	-.45	.21	-.10	.16
Average Thumb Reach	-.26	**-.76***	.08	-.01
Average Functional Range of Motion	.38	**.71***	-.65	.35
Jebsen-Taylor (dominant)	-.31	.37	-.29	.39
Grooved Pegboard (dominant)	.09	.39	.24	-.42

Note: * indicates significance at the .05 level, two-tailed. *n* = 8.

Hand Function, Anthropometry and Satisfaction.

The Jebsen-Taylor distribution was square root transformed, due to a moderate amount of skewness. Pearson correlation coefficients were computed among averages for each of the hand function, anthropometric measures and satisfaction scores for each of the physical input methods for both male and female participants. Since this was an exploratory investigation of these relationships, the alpha level is not adjusted. The results of these analyses are given in Tables 5.4

and 5.5. In general, it appears that larger female hand dimensions have moderate to large relationships to lower satisfaction scores for the Onscreen Qwerty and Handwriting input methods. In addition, it appears that greater Functional Range of Motion for female participants is strongly related to increased satisfaction with the Handwriting input method. Also, increased manipulation dexterity tended to have a strong relationship with increased satisfaction scores for female participants. For male participants, a larger thumb width and lesser thumb reach were strongly associated with increased satisfaction scores.

TABLE 5.4

HAND FUNCTION, ANTHROPOMETRY AND SATISFACTION FOR FEMALES

	Physical Qwerty	Onscreen Qwerty	Tracing	Handwriting
Average Hand Length	.04	-.45	-.13	-.36
Average Hand Breadth	.22	-.40	.03	-.31
Average Thumb Length	.13	-.18	-.18	.09
Average Thumb Width	.10	**-.48***	-.01	**-.53***
Average Thumb Circumference	.09	**-.59***	.04	-.41
Average Index Length	.08	-.46	-.10	-.29
Average Index Width	.03	**-.57***	.19	-.39
Average Index Circumference	.23	**-.52***	.13	-.30
Average Thumb Reach	-.04	-.08	-.05	.03
Average Functional Range of Motion	-.06	-.07	-.25	**.50***
Jebsen-Taylor (dominant)	.19	.37	-.32	-.26
Grooved Pegboard (dominant)	**.50***	.19	-.24	.26

Note: * indicates significance at the .05 level, two-tailed. *n* = 17.

TABLE 5.5

HAND FUNCTION, ANTHROPOMETRY AND SATISFACTION FOR MALES

	Physical Qwerty	Onscreen Qwerty	Tracing	Handwriting
Average Hand Length	-.39	-.11	.21	-.56
Average Hand Breadth	-.31	-.20	.53	-.42
Average Thumb Length	-.11	.14	.05	-.58
Average Thumb Width	-.11	-.23	**.71***	-.38
Average Thumb Circumference	.33	.09	.52	-.24
Average Index Length	-.04	.23	-.13	-.51
Average Index Width	.24	.24	.63	-.33
Average Index Circumference	.58	.53	.45	-.09
Average Thumb Reach	-.06	.01	-.63	**-.83***
Average Functional Range of Motion	.62	.61	.28	.68
Jebsen-Taylor (dominant)	.24	.62	.02	-.24
Grooved Pegboard (dominant)	-.22	-.67	.23	.08

Note: * indicates significance at the .05 level, two-tailed. *n* = 8.

Individual Manual Input Techniques.

While participants were entering phrases, the facilitator recorded the digits used (see Table 5.6) and the way that participants supported the device while entering text (see Table 5.7) with each input method. Most younger adults used both thumbs with the Physical and Onscreen Qwerty input methods, holding the device with both hands, but using the tabletop to support their hands, wrists or forearms. For the Tracing and Handwriting input methods, most participants used a single finger, but also tended to hold the device securely in one hand while using the tabletop for hand, wrist or forearm support. Typically, when a single finger was used to input text, it was the finger on the participants' dominant hand (n = 24). Photographs depicting common hand and device posture for the manual input methods are given in Appendix O.

TABLE 5.6

DIGITS YOUNGER ADULTS USED TO ENTER TEXT WITH EACH METHOD

	Physical Qwerty	**Onscreen Qwerty**	**Tracing**	**Handwriting**
Single Thumb	0%	0%	8%	0%
Both Thumbs	92%	68%	0%	0%
Single Index	0%	28%	84%	96%
Both Index	8%	0%	0%	0%
Single Middle	0%	4%	8%	4%

TABLE 5.7

YOUNGER ADULT MANUAL DEVICE SUPPORT

	Physical Qwerty	**Onscreen Qwerty**	**Tracing**	**Handwriting**
Held in both hands	92%	68%	8%	0%
Held in one hand	0%	28%	64%	64%
Stabilized with one hand	0%	0%	20%	28%
None	8%	4%	8%	8%
Used tabletop	80%	60%	76%	72%
Used lap	4%	8%	4%	8%
No support	16%	32%	20%	20%

Speech and Voice Characteristics and Voice Recognition.

A logarithmic transformation was applied to the Frequency and Amplitude Tremor Intensity distributions, and square root transformation was applied to Pitch and Amplitude Period Perturbation Quotient distributions. Pearson's correlation coefficients were computed among Speech Rate and voice characteristic measures with average Adjusted Words per Minute and Word Error Rate for the Voice condition for both males and females. Since this was an exploratory investigation of these relationships, the alpha level is not adjusted. The results of these analyses are given in Table 5.8 and 5.9. It does not appear that voice characteristics or speech rate had a significant difference on Adjusted Words per Minute or Word Error Rate for the Voice condition for males. It does appear that Amplitude Perturbation (shimmer) did have a large, positive effect on Adjusted Words per Minute for females.

TABLE 5.8

SPEECH AND VOICE CHARACTERISTICS AND VOICE RECOGNITION FOR FEMALES

	Average Voice AdjWPM	Voice Word Error Rate
Speech Rate (WPS)	.18	.25
Average Fundamental Frequency (Hz)	-.10	-.02
Frequency Tremor Intensity Index (%)	.28	-.22
Amplitude Tremor Intensity Index (%)	-.17	.08
Pitch Perturbation Quotient (%)	-.15	.29
Amplitude Perturbation Quotient (%)	**.57***	-.38
Voice Turbulence Index	.39	-.41

Note: * indicates significance at the .05 level, two-tailed. *n* = 17.

TABLE 5.9

SPEECH AND VOICE CHARACTERISTICS AND VOICE RECOGNITION FOR MALES

	Average Voice AdjWPM	Voice Word Error Rate
Speech Rate (WPS)	.57	-.30
Average Fundamental Frequency (Hz)	.17	-.13
Frequency Tremor Intensity Index (%)	.22	-.56
Amplitude Tremor Intensity Index (%)	-.58	.41
Pitch Perturbation Quotient (%)	-.05	.25
Amplitude Perturbation Quotient (%)	.29	.13
Voice Turbulence Index	.44	-.16

Note: n = 8.

Discussion

The results from Study 1 lend support to the hypotheses that the type of smartphone input method (Physical Qwerty, Onscreen Qwerty, Tracing, Handwriting, and Voice) would impact performance metrics, as well as subjective satisfaction, workload and preference ratings for novice younger adults. The results also suggest that anthropometrics and voice qualities may play an important role in determining aspects of user performance and perceptions of these input methods, however, reasoning for the observation of these relationships is speculative.

Performance.

***Adjusted Words per Minute*.** Input method type had a large effect on Adjusted Words per Minute, accounting for most of the variability in the scores. Voice clearly had the highest entry rate, followed by Physical Qwerty, then Tracing, then Onscreen Qwerty and, lastly, Handwriting.

It makes sense that Voice was the fastest, as the human speech rate of roughly 200 WPM is much faster than any form of manual input (Cox et al., 2008). It was also not surprising that Handwriting was the slowest method, as natural handwriting typically does not typically exceed 20 WPM (Bailey, 1996). However, the significant difference between Physical and Onscreen Qwerty was unexpected since the act of text input with each method was anticipated to be similar. One reason that this may have occurred is that the Physical Qwerty provided kinesthetic cues, including button “travel” (i.e., movement) on each key press and bumps on the “f” and “j” keys, which may have provided some assistance in quickly locating the correct keys. The finding that the entry rate for Tracing was faster than Onscreen Qwerty was also unexpected, as gesturing to enter text is a new concept and could have potentially taken longer to learn, though research has shown that, with training, novices can reach an entry rate of 25 WPM with the Tracing input method (Kristensson, 2007).

It was expected that there would be significant differences, and though differences between all input methods were significant, they were not entirely in line with the results of Castellucci and MacKenzie (2011) and the meta-analysis done by Arif and Stuerzlinger (2009). The results by Castellucci and MacKenzie (2011) showed, of the three methods they evaluated, Onscreen Qwerty had the highest Words per Minute, followed by Tracing and then Handwriting. The meta-analysis figured that a thumb mini-Qwerty would be faster than an Onscreen Qwerty or Handwriting ,which agrees with the results from this study, however, the WPM are quite different than values obtained in this study. Possible explanations for this are that the metric used in this study was Adjusted Words per Minute, which corrects for errors left in the final transcription, and results from the other studies report Words per Minute. Another possible explanation is that Castellucci and MacKenzie’s study (2011) used a different device and input method software than what was used in this study, as well as participants who were familiar with the Physical Qwerty method. Additionally, the meta-analysis by Arif and Stuerzlinger included participants with mixed levels of expertise, differences between devices, software, protocols, and tasks.

Error Rates. At the character-level, the results support the hypotheses that input method type has a significant impact on character-level error rates, especially for Corrected Error Rate and, thus, Total Error Rate. Results indicated that Handwriting had significantly higher Corrected and Total Error Rates, followed by Onscreen Qwerty, Tracing and Physical Qwerty, with the lowest Corrected and Total Error Rates. This trend was also seen in the Uncorrected Error Rates, though there were no statistically significant differences between Physical Qwerty and Tracing, and Onscreen Qwerty and Handwriting. The difference between Tracing and Onscreen Qwerty was also nonsignificant, however, this was due to the alpha adjustment for multiple comparisons, thus, a significant difference may be discovered in studies where the alpha level is not adjusted. The same pattern for these error rates was found by Castellucci and MacKenzie (2011) for Handwriting, Onscreen Qwerty and Tracing. The difference in error rates between the Physical and Onscreen Qwertys is, again, somewhat surprising given the similar nature of text input for the two methods. An explanation could be that, even though novice typists have been shown to rely more on visual than proprioceptive feedback on a regular keyboard (Klemmer, 1971), the detection of errors on the touchscreen Qwerty may have been more difficult. Indeed, when proprioceptive cues are not available, a greater reliance on visual cues occurs (Barrett & Krueger, 1994).

A caveat to Corrected Error Rate is that the equation does not differentiate between corrected errors for items that were initially correct and those which were in error. Thus, it can only be interpreted that participants deleted more characters with Handwriting and Onscreen Qwerty than with Tracing and Physical Qwerty. The trend in Uncorrected Error Rates implies that the input methods with higher error rates (Handwriting and Onscreen Qwerty) might have been more laborious to make corrections than those which had lower Uncorrected Error Rates (Tracing and Physical Qwerty). However, there may have been differences inherent to these input methods that made errors less noticeable. For example, participants were observed to write characters in the Handwriting input method which resembled the intended character, but that returned an erroneous character ("l" instead of "i") on occasion. Additionally, it was noticed that participants would, at times, spend more time attending to the input method rather than the output. This may have been due to the increased workload required to enter text with these methods, or simply the participant's trust that the software would interpret their input correctly. In these instances, factors such as erroneous autocorrected words may have had more of an influence on the frequency of uncorrected errors.

Results from Study 1 showed that there were no significant differences among input methods for Word Error Rates. This is interesting, since there were significant differences between uncorrected errors at the character-level. One possible explanation for this is that novice younger adults may have been more conscientious about correcting errors that changed the interpretation of their output. Another explanation is that once they made an error, they may have been more likely to make another error in the same word, which would have counted as multiple character-level errors, but only a single word error. At any rate, the percentage of word errors in all categories was very low, ranging from roughly 1 – 2%, which may have been viewed as an acceptable level of error by the participants.

***Satisfaction*.** Results from Study 1 lend support for the hypothesis that input method type had a significant effect on reported Satisfaction levels. Results suggest that Physical Qwerty and Voice are the most satisfying input methods to use for entering text on a smartphone, followed by Tracing and Onscreen Qwerty. Handwriting was clearly the least satisfactory. One reason that these results may have occurred is that, as the participants were all novices, the Tracing, Onscreen Qwerty and Handwriting input methods may have seemed more foreign to them. Another reason may be that they associated their satisfaction with their high levels of perceived accuracy and low amount of workload with the Physical Qwerty and Voice input methods. Additionally, since the participants in this study mostly had experience with a physical numeric keypad, they may have been more comfortable with the act of pushing buttons to enter text, and the familiarity of this method may have made it appear more appealing.

Workload. The results from Study 1 lend support for the hypothesis that input method had a significant effect on perceived workload. The Voice input method had significantly lower workload ratings than all of the manual input methods, except Physical Qwerty which was rated similarly, and Handwriting had significantly higher levels of reported workload compared to all of the other input methods. Though these input methods have never been directly compared before, some of these results were expected. Handwriting has a high physical investment in each letterform, since each letter could be comprised from one to four strokes (as is the case for a lowercase “l” or a lowercase “w”), depending on writing style. Often, participants reported they had to “learn” the way that the software “wanted” them to write. It was observed that participants would correct letters over and over again until they would figure out an acceptable method for getting their intended letter, even if it was not intuitive. Voice has a very low physical investment, in that speech is not produced at the character or stroke-level. The significant difference in workload ratings between Physical Qwerty and Onscreen Qwerty was not expected. Participants frequently commented about how they liked the tactile feedback, but also mentioned they disliked the high key resistance of the physical keyboard. Comments regarding the Onscreen Qwerty included appreciation for not having to depress any keys. Thus, though the two Qwerty methods scored similar on the Effort dimension, it is interesting that the Onscreen Qwerty scored higher on the Physical dimension of workload. Other contributing factors to high workload ratings for the Onscreen Qwerty, compared to its physical counterpart, include higher perceptions of frustration, mental workload and the participants' perceptions that they performed more poorly with this method. Tracing performed similarly to both Qwertys, which is surprising considering participant comments about hands and fingers obscuring the screen while they were trying to trace and that many mentioned that they disliked having their finger in contact with the screen for extended periods of time.

Preference. Participant ratings pre-test showed that they favored the Qwerty methods over the others mainly for the familiarity of the layout and that they felt confident using desktop keyboards and comfortable using buttons to enter text (i.e., phones with a numeric keypad). They reported they were skeptical of technologies that did too much "behind the scenes" processing, which made the input method difficult to conceptualize and which they perceived as taking control over the output away from the user.

After the users had experienced each of the input methods, they shifted their opinions somewhat. They clearly preferred Physical Qwerty above all other methods, and they expressed their perception that this was the most accurate method. Many mentioned they liked the tactile feedback, which may be related to their familiarity of entering text with a physical button press with their phone's numeric keypad or with full-sized keyboards. However, they also commented that the physical keys created the opportunity to do other tasks while typing without having to look at the keyboard. Onscreen Qwerty and Handwriting were rated lower than they were during the pre-test preference ratings. This could be due to a more disappointing experience with each of these methods than anticipated. Indeed, many participants reported that they struggled with both methods because they believed they made more mistakes and that it was difficult to correct mistakes with them. Impressions of the Onscreen Qwerty were that the keyboard was inconsistently responsive in across areas of the screen, and that there were numerous small keys, spaced too closely together. In fact, it was observed that participants would often miss the space bar and instead hit the bezel of the device, which gave them the impression that the space bar was not responding to their touch. The feedback that the Handwriting method provided participants was fairly ambiguous, and kept participants guessing about the proper way to make letterforms. Tracing and Voice were both seen in a more positive light post-test. These results could be due to a more positive experience with the technology than anticipated.

Hand Function and Handwriting. The results from the Pearson's correlation between the Jebsen-Taylor hand function test for the dominant hand and Handwriting Adjusted Words per Minute showed the relationship was not significant. This was somewhat surprising, since hand function was thought to have been related to handwriting performance. However, many of the participants had difficulty with the Jebsen-Taylor test because they were not used to writing in cursive. Additionally, the Handwriting input method used a finger to input text, rather than a stylus, so it is a different type of interaction and requires different muscles, force, and control. Entry rate may have been more dependent upon letter form rather than composition speed, which was not accounted for by the Jebsen-Taylor test. Generally, observations of personal style did seem to make a difference, qualitatively, and legibility did not. Participants who used different stroke orders when forming their characters, used more ligatures between letters, wrote their characters too large on the screen, waited for their phrase to process one letter at a time, and who forgot to hit the space bar frequently tended to have more errors. So it seemed as though personal handwriting style and proficiency with the conventions of the interface tended to increase accuracy with this method, rather than simply hand function.

Keyboarding and Smartphone Qwerty Speed. Results from the Pearson's correlations between Keyboarding Speed and both of the Qwerty input methods indicated non-significant, though moderate, positive relationships. The relationship was slightly stronger between Keyboarding Speed and the Physical Qwerty. Even though it was not significant, it suggests the trend that those who have a higher entry rate with a computer keyboard also have a higher entry rate for smaller Qwerty devices. This was not surprising, as familiarity and experience with the Qwerty layout may give users an advantage over those who are not. Having a non-significant relationship was surprising, however, and could be explained by the reduced size of the smartphone Qwertys. Many participants expressed that they thought they would be able to transfer their training with desktop Qwerty keyboard to the smaller keypads. However, they reported that they were unable to use any muscle memory since the keypads were only large enough to use one or two digits to input text. Indeed, smartphone Qwertys are typically designed to accommodate "thumbing" (thumb input), which is quite different from utilizing all ten fingers with a full-sized keyboard. Additionally, the Onscreen Qwerty did not provide any tactile feedback.

Hand Function, Anthropometry and Total Error Rate. The results from the Pearson's correlations between Hand Function (Jebsen-Taylor, Grooved Pegboard, Thumb Reach and Functional Range of Motion), Anthropometry (hand length and breadth, thumb length, width and circumference, and index length, width and circumference) and error rates for all of the manual input methods showed two significant relationships with the female participants, and five significant relationships with the male participants.

The large, positive relationships between hand and index finger length seem to suggest that female participants with longer hand dimensions tend to create more errors with the Onscreen Qwerty method. This is interesting because roughly one third of participants used their index finger to type with this method. One possible explanation for this is that female participants with larger hands felt more comfortable typing the phrase with both thumbs since they could reach all areas of the screen better. In fact, seven of the eight female participants with longer hands did type using both thumbs. This may have decreased their coordination with key selection. Whereas, females with smaller hands might have decided to use a single digit to enter text, allowing more control over the interface. Females with shorter hands did use a single index finger (four of nine) than the group with longer hands, however, more than half used both thumbs.

Based on the large negative relationships between thumb length, thumb reach and the Onscreen Qwerty input method, it appears that male participants had fewer errors if they had longer digit dimensions and had greater extension of their thumbs. This may imply that entering text with this input method may be more successful for people who are able to easily reach all areas of the keypad. However, this explanation may not hold since the large positive relationship of Functional Range of Motion with males' Total Error Rates for Onscreen Qwerty implies that those with increased joint mobility experienced increased error rates. Also interesting to note, the large negative relationships between thumb and index width with Physical Qwerty Total Error Rates indicates that male participants, who had wider digits, tended to have lower error rates with this input method. One possible explanation for this is that people with smaller hands tended to report that they felt the physical keypad was too large, and that they had problems reaching the center keys. This result may make sense when thinking about overall hand dimensions (which is also reflected in the trends of the coefficients) where male participants with smaller hands created more errors with the Physical Qwerty.

Hand Function, Anthropometry and Satisfaction. Results from the Pearson's correlations between Hand Function (Jebsen-Taylor, Grooved Pegboard, Thumb Reach and Functional Range of Motion), Anthropometry (hand length and breadth, thumb length, width and circumference, and index length, width and circumference) and error rates for all of the manual input methods revealed seven significant relationships with the female participants and only two significant relationships with the males.

It appears that female participants who had greater digit width and circumference were significantly (large effect) less satisfied with the Onscreen Qwerty and, and those who had larger thumb width were less satisfied with Handwriting. This is interesting because the key size for the Onscreen Qwerty method was actually larger than the key size for the Physical Qwerty, and yet digit size does not seem to play a role when determining satisfaction with the physical keyboard. However, most of the female participants with larger hands used both thumbs for entering text, so perhaps it can be explained by the technique used to enter text rather than by the size of the digits. The result that larger female thumb width is related to lower satisfaction scores for Handwriting is puzzling, since many participants reported that they did not feel constrained by digit size when using this input method, and none of the participants entered text with Handwriting using their thumbs. Similarly, the male thumb width seemed to have a large positive relationship to satisfaction with the Tracing input method, indicating it was more satisfactory to use for males with wider thumbs. This may indicate that the males' larger thumbs made it more awkward to interact with the other input methods.

Additionally, Functional Range of Motion seems to have a large, positive relationship with the Handwriting input method for the female participants. As range of motion measures the ability of a joint to move, this finding may imply that those with greater ability to move their digits felt more satisfied using the input method for some reason. Range of motion does not appear to be related to Total Error Rate, but it may be related to entry rate, pacing of the task, more complete visual feedback from the trace on the interface, or other factors. Range of motion was not significantly related to the other input methods, which may mean that participants did not have to utilize much joint movement while doing those tasks. Indeed, some participants would hold their fingers stationary while entering text with these methods, using their wrists or forearms to move the pad of their finger along the screen. What is surprising is that many participants used both thumbs when interacting with the Physical Qwerty method, thus, it was expected that Functional Range of Motion would have had a stronger relationship with that input method.

There was a significant large negative relationship discovered between Thumb Reach and the Handwriting input method for male participants. This was surprising, since most participants used their index fingers to interact with Handwriting. This result may be related to the willingness the participants had to extend their digits to complete the task. Perhaps those who had greater reach were more capable of creating larger letterforms, which could have been interpreted by the software as capital letters, or may have even caused the finger to go out of the bounds of the input area on the interface. These would have created more errors and input not recognized by the system, which may have made it more frustrating to use. This might be a plausible explanation, except that it was not reflected in the Pearson's correlations between reach and Total Error Rates for Handwriting.

Results from the Pearson's correlation between the Grooved Pegboard and the Physical Qwerty satisfaction scores also revealed a large, positive effect for female participants. This implies that the less manipulative dexterity, or coordination, the participant had, the more satisfied they were with the Physical Qwerty input method. This is surprising because the type of interaction with the Physical Qwerty is roughly the same as with the Onscreen Qwerty, for those who use both thumbs. The Physical Qwerty may have required less dexterity to operate than the Onscreen Qwerty, or perhaps the proprioceptive cues provided by the Physical Qwerty made it more satisfying to use for those who had lower levels of dexterity.

Speech and Voice Characteristics and Recognition. Results from Study 1 suggest that Amplitude Perturbation has a significant positive relationship with the Voice input method Adjusted Words per Minute for females. Amplitude perturbation, shimmer, is the variation in the amplitude of a waveform and contributes to the perception of a "rough" voice. It is interesting that increased shimmer would be related to a higher entry rate, however, this might have occurred for a couple of reasons. First, an increased presence of shimmer can indicate a decrease in vocal effort (Huang et al., 1995). If this applies to the obtained data, then it may indicate that the female participants who used less effort or spoke more naturally may have had better results with the software, and thus were able to enter text at a faster rate. Second, shimmer has been shown to remain relatively constant across a variety of emotions and neutral speech, but tends to decrease with "fear" (Li & Jo, 2005). This might indicate that those who were more afraid, or worried, may have had more difficulties with the system, resulting in lower entry rates. A third explanation could be that the voice quality measures were taken as a separate measure, before the users interacted with the voice input method, and they may not truly be related to actual performance since voice quality was not assessed during the trials. Also, the measured voice parameters do not occur in isolation, thus, it is important to further explore potential relationships using more complex models and a larger sample size. No significant relationships were found with the younger adult males, which could mean that the relationships (ranging from small to large in magnitude) implied by the correlation coefficients are simply due to chance.

CHAPTER 6

STUDY 2: SMARTPHONE INPUT METHOD PERFORMANCE, SATISFACTION, WORKLOAD, AND PREFERENCE WITH NOVICE ADULTS AGED 60 - 82

Purpose

Considerable physical and cognitive changes related to aging may impede aspects of smartphone text entry performance, and subjective perceptions of smartphone text entry methods. Previous research regarding smartphone input methods, and input methods in general, have neglected older adult novices. As older adults discover the benefits of owning smartphones, and as societal demographics are projecting a growing percentage of people in the 65 – 84 age group in the coming decades, it is imperative to develop a greater understanding of the strengths and weaknesses of current standard text entry modes for these devices. Capturing initial usability through performance, as well as subjective impressions of the users' experiences and attitudes, is key for developing design recommendations for this age group.

Previous research (e.g., Castellucci & MacKenzie, 2011; Arif & Steurzlinger, 2009; Költringer& Grechenig, 2004; Read et al., 2001) has provided inconsistent results on performance measures across different input methods, and no research examines the impact that smartphone input methods may have on novice older adult's performance and subjective experience. The lack of evidence in the literature made the formulation of directional hypotheses impossible, thus, this study was exploratory in nature. This study investigated possible differences between five current smartphone text entry methods on novice older adult performance, perceived satisfaction, workload and preference. Additionally, it was of interest to explore whether certain assessments of individual characteristics (e.g., attitudes, cognitive and motor abilities, speech characteristics, etc.) may have been correlated with relevant dependent measures. This was a 5-level, within-subjects design. It was hypothesized that:

H_1: Significant main effects for text input method with performance measures (Adjusted Words per Minute, Corrected Error Rate, Uncorrected Error Rate, Total Error Rate, and Word Error Rate) would be found.

H_2: Significant main effects for text input method with subjective measures (reported Satisfaction, Workload and Post-Test Preference) would be found.

H_3: Significant relationships between individual assessments and the dependent measures would be discovered, where:

- H_{3a}: Baseline hand function (Jebsen-Taylor sentences) would be significantly correlated with Handwriting Adjusted Words per Minute.
- H_{3b}: Baseline keyboarding speed would be significantly correlated with Physical and On-Screen Qwerty Adjusted Words per Minute.
- H_{3c}: Larger hand and digit dimensions would be significantly correlated with Total Error Rate, for physical modes of input.
- H_{3d}: Larger hand and digit dimensions would be significantly correlated with satisfaction, for physical modes of input.
- H_{3e}: Speech characteristics (e.g., speech rate, Jitter, Shimmer, etc.) would be significantly correlated with Voice Adjusted Words per Minute and Voice Word Error Rate.

Method

Participants.

Thirty participants were recruited through the Wichita State University SONA system and Exercise Science classes. Flyers were posted at Wichita State University, Barton County Community College, and the Larksfield Place retirement community. Additionally, snowball sampling occurred through Facebook, email, and word of mouth. Participants were compensated $30 cash. Five participants' data was discarded, as three participants were unable to adequately interact with the smartphone device and two participants asked to withdraw from the study before data collection was complete.

Twenty-five participants (16 female, 9 male) were retained for analysis. Participants had an age range of 60-82 years (M = 68.8, SD = 7.4). Many participants reported they had at least attained a Bachelor's Degree (n = 17) as their highest level of completed education. Most were retired (n = 15), and of the participants still employed, many worked in an academic setting (n = 8). Most were right-handed (n = 21), and did not report any major current finger, hand or wrist problems. Some reported mild arthritis (n = 7), however, all of these participants expressed the problem was managed by over-the-counter pain relievers, when necessary. None of the participants had current speech disorders or smoked tobacco. All participants were novices to each of the smartphone input methods investigated in this study (for demographics, see Appendix N). All participants exceeded the acceptable threshold of 23 set for cognitive function determined by the SMMSE (Standardized Mini-Mental State Examination), had acceptable near visual acuity, and fingertip sensitivity.

Materials.

Materials were the same in this study as in Study 1, with the exception of the Standardized Mini-Mental State Examination (SMMSE) (Folstein et al., 1975), which consisted of a paper form on which responses were recorded, a blank sheet of paper, a sheet of paper on which the participant responded to certain questions, a wristwatch, and a pencil (see Appendix H). Materials are described in Chapter 4.

Procedure.

The procedure for Study 2 was nearly the same as for Study 1, described in Chapter 4. Differences included the administration of the SMMSE, which was done after assessing binocular visual acuity and before testing fingertip sensitivity. In addition, though the baseline assessments remained at roughly 45 minutes in duration, the older participants typically took 2 to 3 hours to complete the performance section of the study, including time taken for breaks.

Five participants elected to split this portion of the study into two sessions, while most (n = 20) completed the study in one block of time. Those who completed the study in two sessions were asked to enter a couple of phrases with the methods they had used during the previous session before rating their preferences post-test. However, generally baseline measures were gathered during one session and the performance, satisfaction, workload, and preference measures at another session. No more than one week lapsed between sessions.

Results

All data screening, cleaning, and alpha adjustment procedures, as well as guidelines for effect size interpretations, are described in Chapter 4. A list of outliers for each dependent measure is given in Appendix M.

Adjusted Words per Minute.

After the application of a square root transformation, a one-way repeated measures analysis of variance showed a large significant main effect for input method, $F(4, 96) = 453.36$, $p < .001$, $\eta_p^2 = .95$.

Follow-up paired-samples *t* tests indicated that Voice ($M = 41.70$, $SD = .17$) had a significantly higher entry rate than Physical Qwerty ($M = 16.58$, $SD = .30$), $t(24) = 21.09$, Onscreen Qwerty ($M = 8.68$, $SD = .21$), $t(24) = 29.50$, Tracing ($M = 10.99$, $SD = .21$), $t(24) = 30.06$, and Handwriting ($M = 6.50$, $SD = .35$), $t(24) = 30.95$, all $p < .001$. Physical Qwerty had a higher entry rate than Onscreen Qwerty, $t(24) = 9.68$, Tracing, $t(24) = 10.13$, and Handwriting, $t(24) = 17.54$, all $p < .001$. Tracing and Onscreen Qwerty both had higher entry rates than Handwriting, $t(24) = 8.51$, and $t(24) = 4.11$, respectively, both $p < .001$. Finally, Tracing had a higher entry rate than Onscreen Qwerty, $t(24) = 4.03$, $p < .001$ (see Figure 6.1).

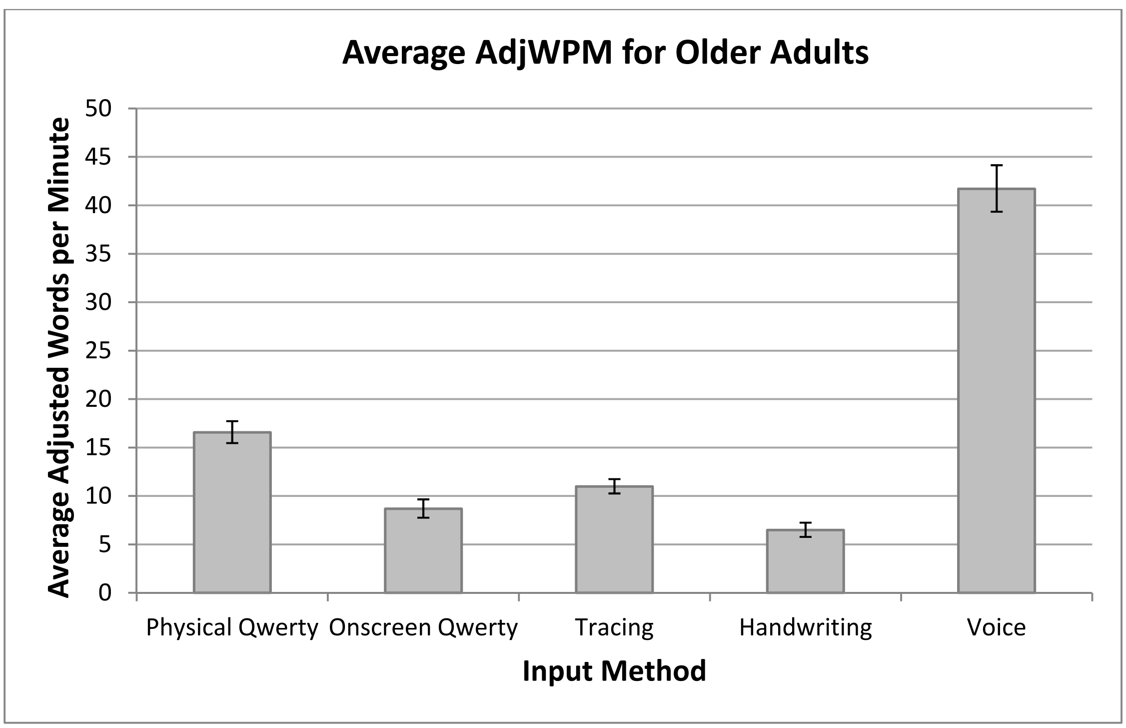


Figure 6.1. Average Adjusted Words per Minute for older adults, values are back-transformed. All conditions were significantly different from each other. Bars indicate 95% CI.

Error Rates.

The manual input methods (Physical and Onscreen Qwerty, Tracing, and Handwriting) all use methods to produce text at the character-level, thus the Unified Error Metric (Corrected, Uncorrected, and Total Error Rates) is used to describe accuracy for these input methods. Since the Voice input method is a word-level entry method, the comparison in accuracy across all input methods is only done using the Word Error Rate. Thus, the accuracy of the input stream is not analyzed for Voice.

Corrected Error Rate. After an arcsine square root transformation was applied, a one-way repeated measures analysis of variance revealed a large significant main effect for input method, $F(3, 72) = 46.68$, $p < .001$, $\eta_p^2 = .66$.

Follow-up paired-samples *t* tests indicated that the Physical Qwerty ($M = 2.59\%$, $SD = 1.06\%$) input method had a significantly lower corrected error rate compared to Onscreen Qwerty ($M = 12.32\%$, $SD = 1.84\%$), $t(24) = 6.97$, Tracing ($M = 6.88\%$, $SD = .89\%$), $t(24) = 4.04$, and Handwriting ($M = 19.57\%$, $SD = 1.52\%$), $t(24) = 14.69$, all $p < .001$. Tracing had a significantly lower corrected error rate than Handwriting, $t(24) = 7.32$, $p < .001$, and Onscreen Qwerty, $t(24) = 3.56$, $p = .002$. The difference between Onscreen Qwerty and Handwriting was not significant, $t(24) = 3.34$, $p = .003$. Of the manual input methods for older adults, Physical Qwerty clearly had the lowest corrected error rate, sequentially followed by Tracing, Onscreen Qwerty, and Handwriting (see Figure 6.2).

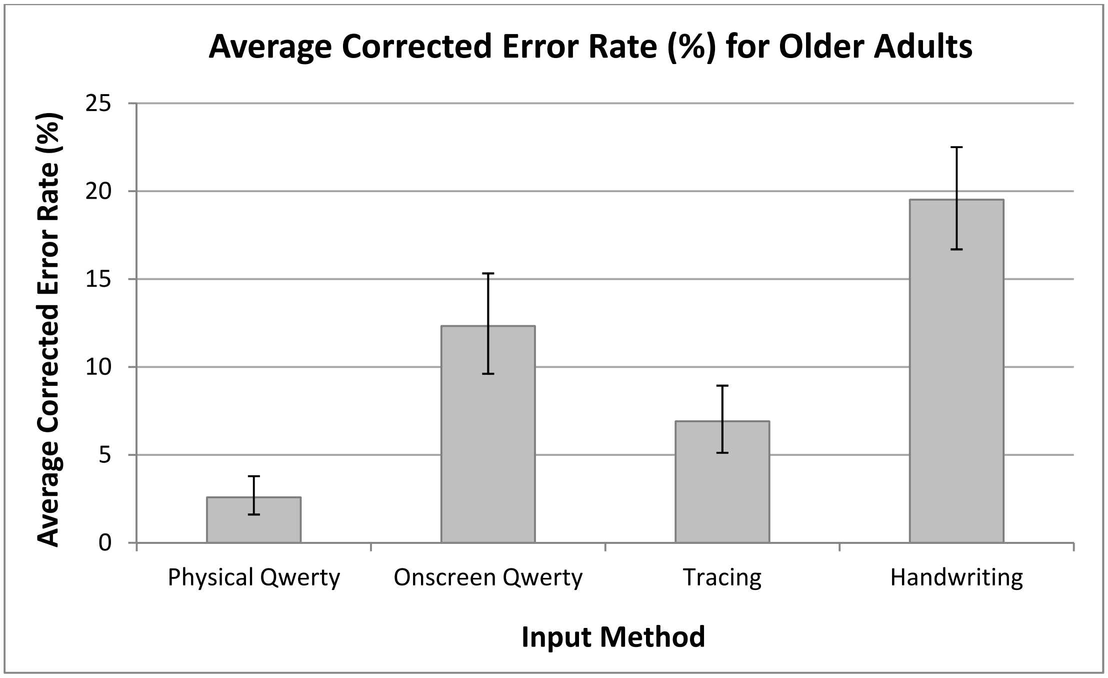


Figure 6.2. Average Corrected Error Rate for older adults, values are back-transformed. Physical Qwerty had the lowest average Corrected Error Rate. Bars represent 95% CI.

Uncorrected Error Rate. Preliminary analyses of average Uncorrected Error Rates revealed four outlying percentages, which were all recoded (see Appendix M). After applying an arcsine square root transformation, results from a one-way repeated measures analysis of variance showed a large significant main effect for input method, $F(3, 72) = 11.88$, $p < .001$, $\eta_p^2 = .33$.

Follow-up paired-samples *t* tests indicated that the Physical Qwerty ($M = .46\%$, $SD = .25\%$) input method had a significantly lower uncorrected error rate compared to Onscreen Qwerty ($M = 2.06\%$, $SD = .62\%$), $t(24) = 4.43$, Tracing ($M = 1.64\%$, $SD = .55\%$), $t(24) = 4.24$, and Handwriting ($M = 2.78\%$, $SD = .74\%$), $t(24) = 5.12$, all $p < .001$. There were no significant differences in uncorrected error rates between Onscreen Qwerty and Tracing, $t(24) = 1.22$, $p = .23$, Onscreen Qwerty and Handwriting, $t(24) = 1.16$, $p = .26$, and Tracing and Handwriting, $t(24) = 1.99$, $p = .06$. Of the manual input methods for older adults, Physical Qwerty had the lowest uncorrected error rate. Onscreen Qwerty, Tracing, and Handwriting had similar uncorrected error rates (see Figure 6.3).

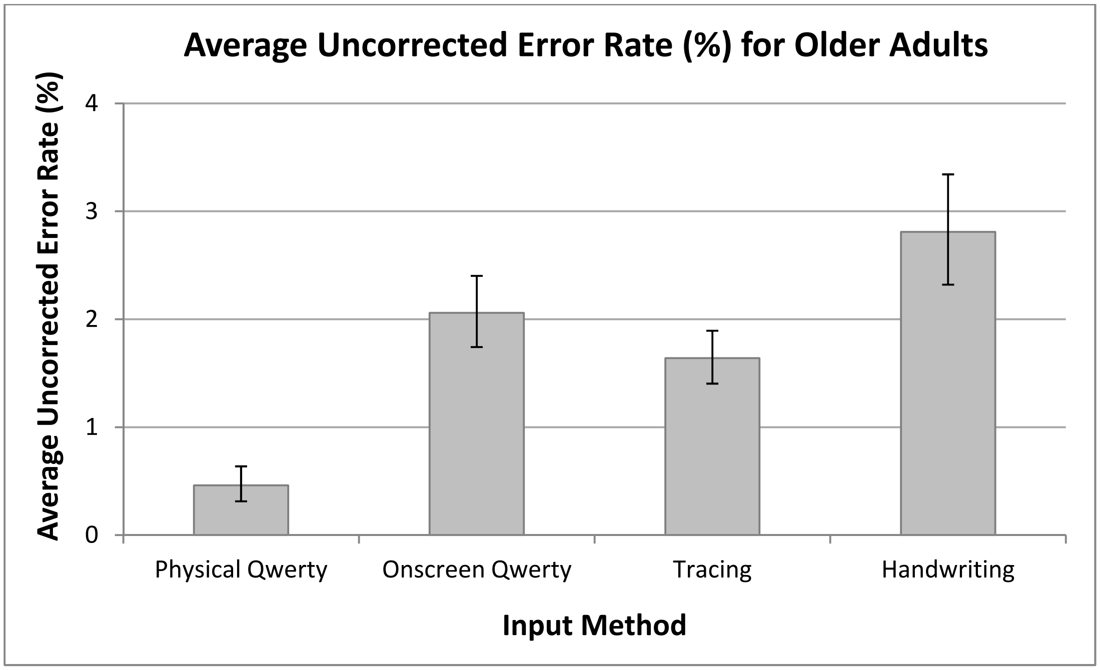


Figure 6.3. Average Uncorrected Error Rate for older adults, values are back-transformed. Physical Qwerty had the lowest average Uncorrected Error Rate. Bars indicate 95% CI.

Total Error Rate. An arcsine square root transformation was applied to the data, and results from a one-way repeated measures analysis of variance revealed a large significant main effect for input method, $F(3, 72) = 62.84$, $p < .001$, $\eta_p^2 = .72$.

Follow-up paired-samples *t* tests showed that the Physical Qwerty (M = 3.38%, SD = 1.09%) input method had a significantly lower average total error rates compared to Onscreen Qwerty (M = 15.53%, SD = 1.59%), $t(24) = 9.28$, Tracing (M = 9.41%, SD = .90%), $t(24) = 5.24$, and Handwriting (M = 23.17%, SD = 1.43%), $t(24) = 14.28$, all $p < .001$. Tracing had a significantly lower total error rate than Handwriting, $t(24) = 8.24$, and Onscreen Qwerty, $t(24) = 3.97$, both $p < .001$. Onscreen Qwerty had significantly lower error rates compared to Handwriting, $t(24) = 3.51$, $p = .002$. Of the manual input methods for older adults, Physical

Qwerty clearly had the lowest total error rate, sequentially followed by Tracing, Onscreen Qwerty, and Handwriting (see Figure 6.4).

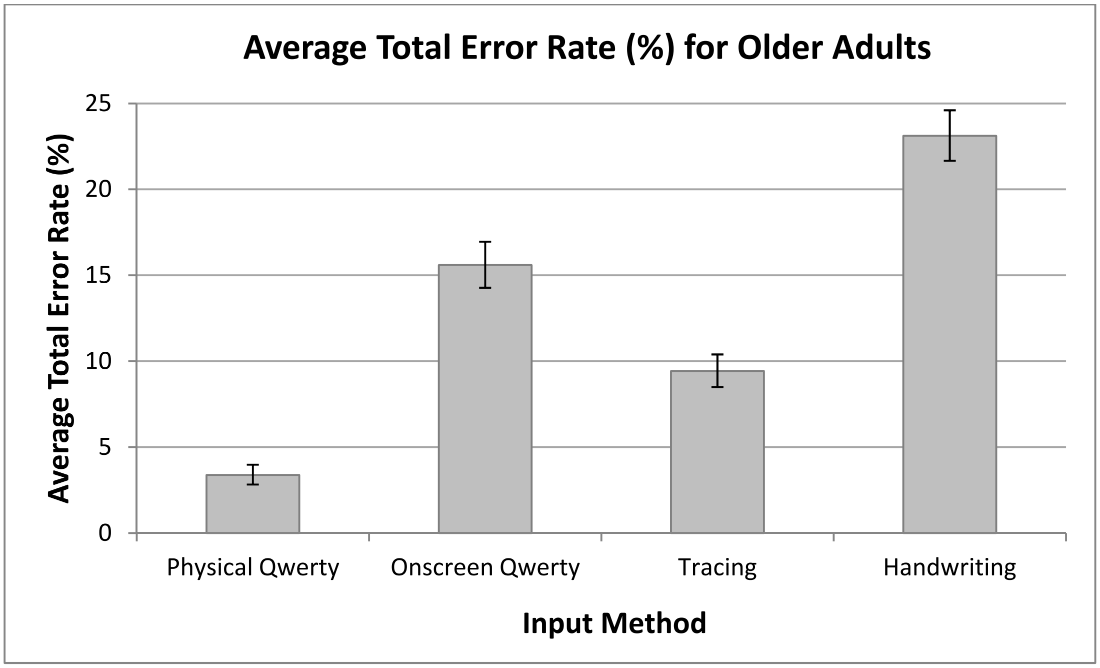


Figure 6.4. Average Total Error Rate for older adults, values are back-transformed. Physical Qwerty had the lowest average Total Error Rate. Bars indicate 95% CI.

***Word Error Rate*.** Preliminary analyses of average Word Error Rate revealed four outlying percentages, which were all recoded (see Appendix X). After applying an arcsine square root transformation, a one-way repeated measures analysis of variance revealed a large significant main effect for input method, $F(4, 96) = 5.49$, $p = .001$, $\eta_p^2 = .19$.

Follow-up paired-samples *t* tests indicated that the Voice ($M = 1.93\%$, $SD = .85\%$) input method had signficantly lower word error rates compared to Handwriting ($M = 6.43\%$, $SD = 2.59\%$), $t(24) = 3.63$, $p = .001$. There were no significant differences between Physical Qwerty ($M = 2.04\%$, $SD = 1.48\%$) and Voice , $t(24) = .12$, $p = .91$, Onscreen Qwerty ($M = 4.29\%$, $SD = 2.19\%$) and Tracing, $t(24) = .29$, $p = .77$, Onscreen Qwerty and Handwriting, $t(24) = 1.36$, $p = .19$, Tracing and Handwriting, $t(24) = 1.49$, $p = .15$, Physical Qwerty and Onscreen Qwerty, $t(24) = 2.12$, $p = .05$, Onscreen Qwerty and Voice, $t(24) = 2.13$, $p = .04$, Voice and Tracing (*M*

= 4.55%, *SD* = 1.20%), *t*(24) = 2.86, *p* = .009, Physical Qwerty and Handwriting, *t*(24) = 3.02, *p* = .006, and Physical Qwerty and Tracing, *t*(24) = 3.14, *p* = .004. For older adults, the Voice and Physical Qwerty input methods tended to have lower word error rates, while Tracing and Handwriting had higher word error rates. Onscreen Qwerty was in the middle. However, only the Voice and Handwriting input methods differed significantly (see Figure 6.5).

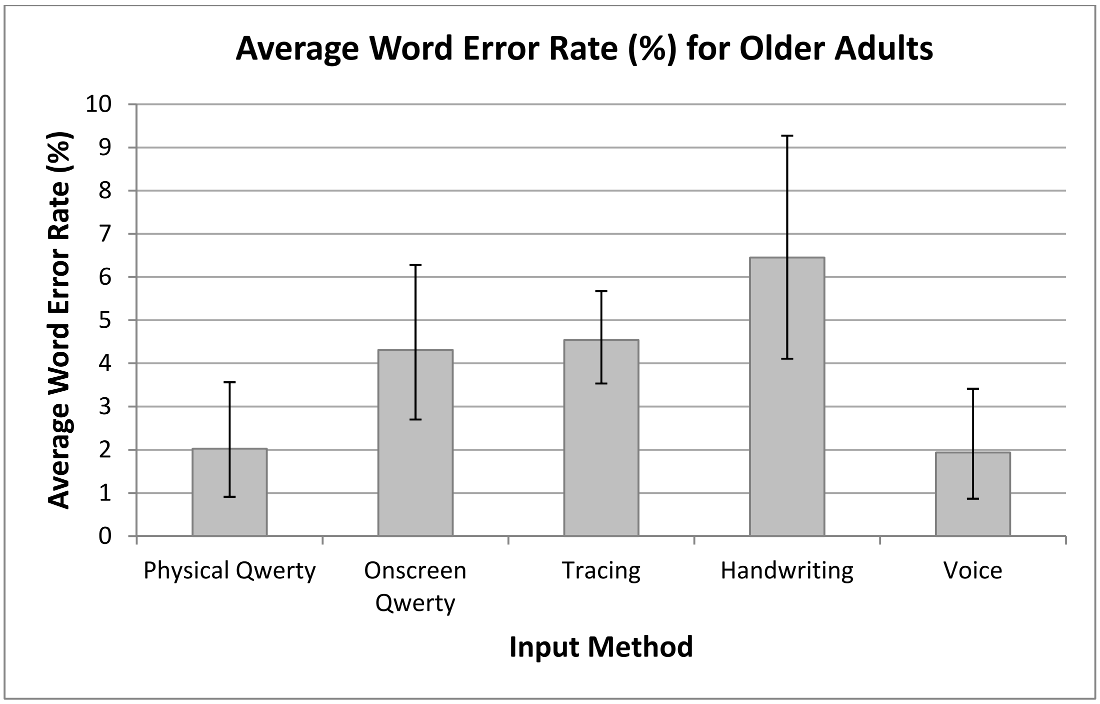


Figure 6.5. Average Word Error Rate for older adults, values are back-transformed. Voice and Physical Qwerty had the lowest Word Error Rates. Bars indicate 95% CI.

Satisfaction.

Preliminary analyses of satisfaction using the modified System Usability Scale (SUS) revealed two outlying values, both of which were recoded (see Appendix M). A one-way repeated measures analysis of variance was conducted to evaluate satisfaction scores across input methods. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(9)$ = 29.73, *p* =.001, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ε = .57). The results showed a large significant main effect for input method, $F(2.28, 54.77) = 19.01$, $p < .001$, $\eta_p^2 = .44$.

Follow-up paired-samples *t* tests indicated that the Voice input method ($M = 81.38$, $SD = 9.77$) was perceived as significantly more satisfying than Onscreen Qwerty ($M = 47$, $SD = 27.49$), $t(24) = 5.70$, Handwriting ($M = 38.5$, $SD = 21.23$), $t(24) = 8.65$, and Tracing ($M = 57.5$, $SD = 28.76$), $t(24) = 3.79$, $p < .001$. Physical Qwerty ($M = 78.9$, $SD = 18.07$) was reported more satisfying than Onscreen Qwerty, $t(24) = 6.23$, and Handwriting, $t(24) = 8.14$, both $p < .001$. Tracing was more satisfying than Handwriting, $t(24) = 3.79$, $p < .001$. There were no significant differences in reported satisfaction between Voice and Physical Qwerty, $t(24) = .62$, $p = .54$, Onscreen Qwerty and Tracing, $t(24) = 1.14$, $p = .26$, and Onscreen Qwerty and Handwriting, $t(24) = 1.22$, $p = .24$, and Physical Qwerty and Tracing, $t(24) = 2.87$, $p = .008$. It appears that older adults are generally more satisfied with the Voice and Physical Qwerty input methods over the others (see Figure 6.6).

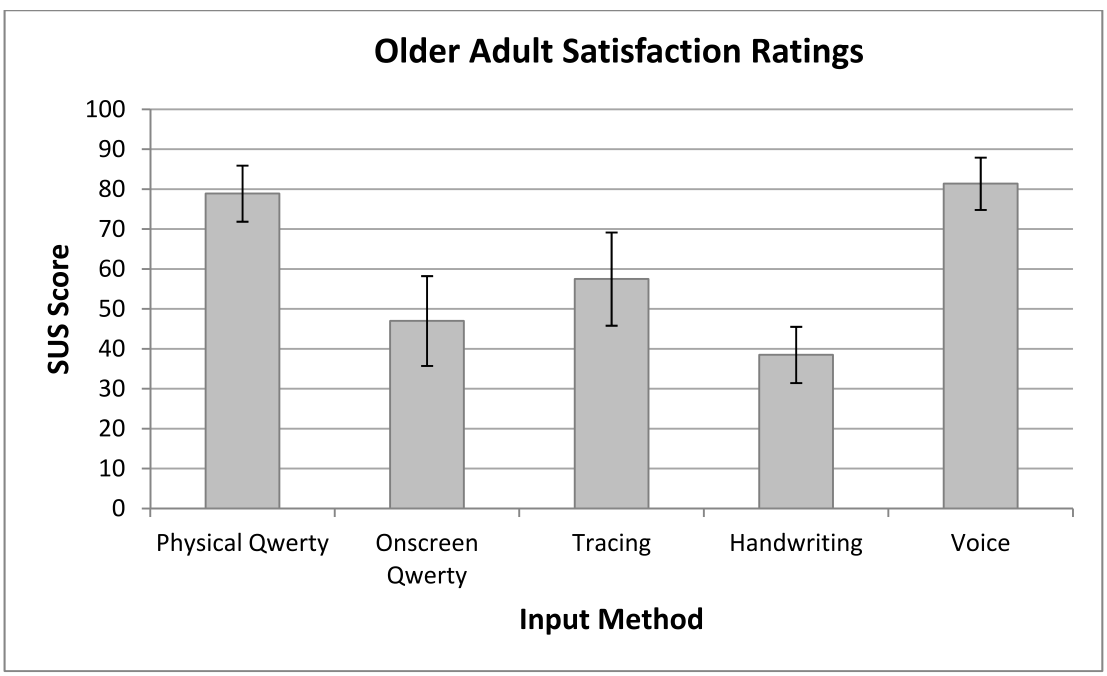


Figure 6.6. Older adult Satisfaction ratings. Voice and Physical Qwerty elicited higher satisfaction scores. Bars indicate 95% CI.

Workload.

Participants rated their subjective level of workload for each of six workload dimensions (i.e., Mental, Physical, Temporal, Performance, Effort, and Frustration) on a 1 – 21 scale (1 = Very Low, 21 = Very High for every dimension except for Performance, which was rated 1 = Perfect, 21 = Failure). Distributions of each dimension were examined independently, and a total of seventeen outliers were recoded (see Appendix M). A reliability analysis was conducted for all of the workload dimensions for each of the input methods. Cronbach's alphas are reported in Appendix A. A workload composite score was created for each input method by summing the scores for each dimension. Graphs of average ratings for individual workload dimensions for each input method are given in Figures 6.7 through 6.12.

Graphically, it appears that the Onscreen Qwerty and Handwriting input methods were more frustrating, made participants feel more hurried or rushed while completing the task, and required more physical workload and effort. Additionally, participants seemed to feel that their performance was worse when using the Handwriting input method. Mental workload seemed to be similarly high among Onscreen Qwerty, Tracing and Handwriting. Voice scored the lowest on all dimensions of workload.

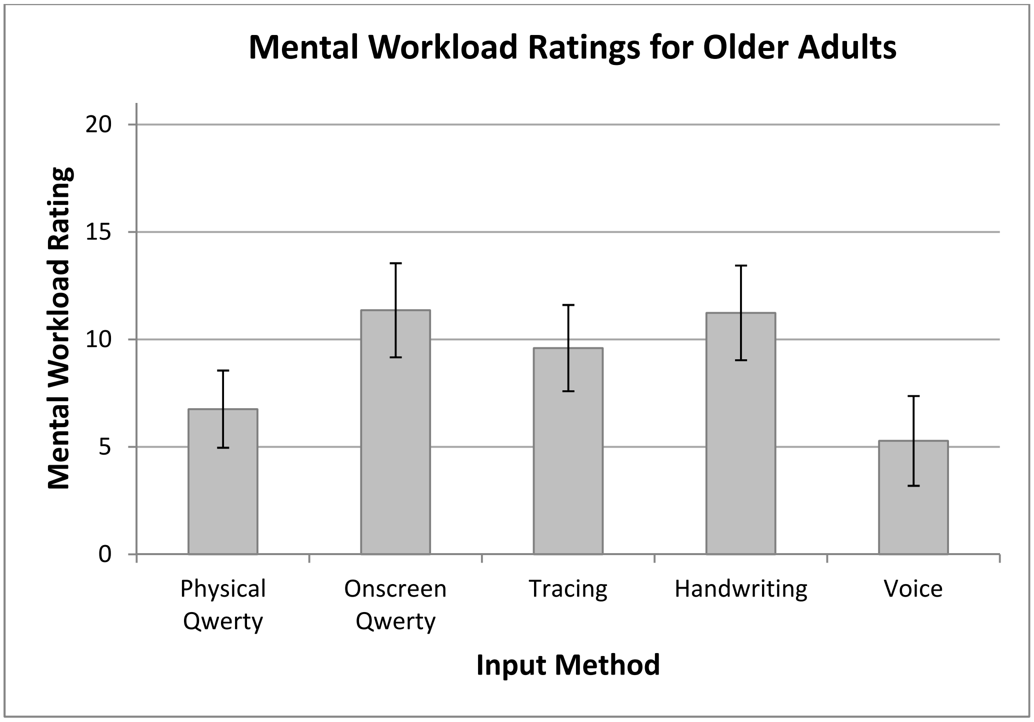


Figure 6.7. Mental workload ratings for older adults. Error bars represent 95% CI.

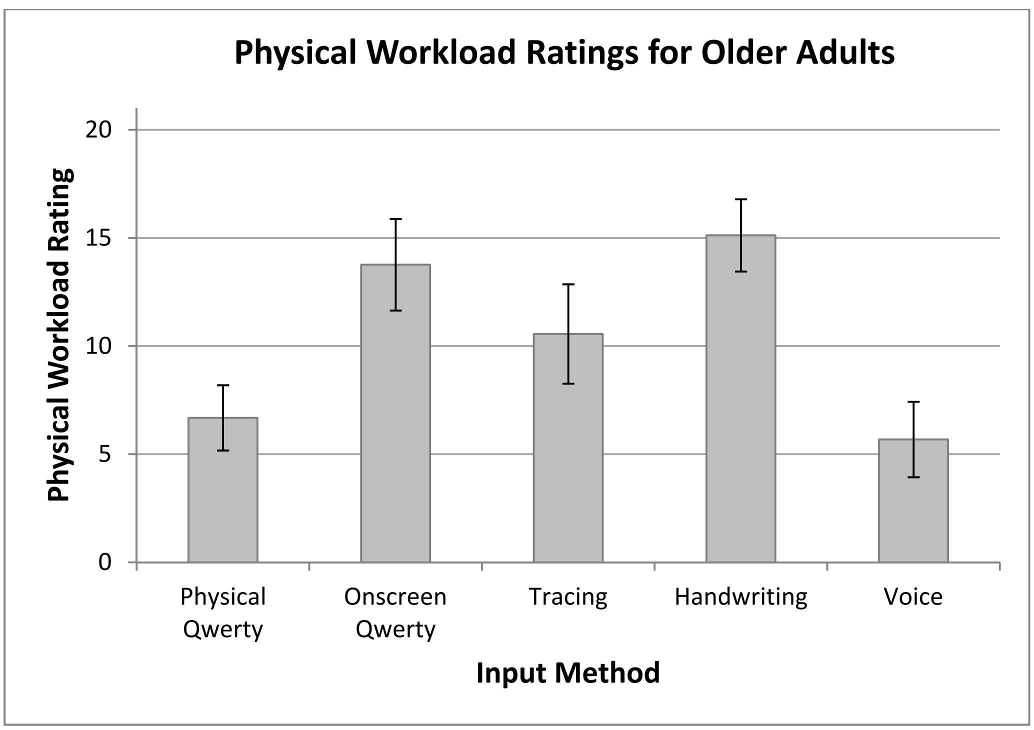


Figure 6.8. Physical workload ratings for older adults. Error bars represent 95% CI.

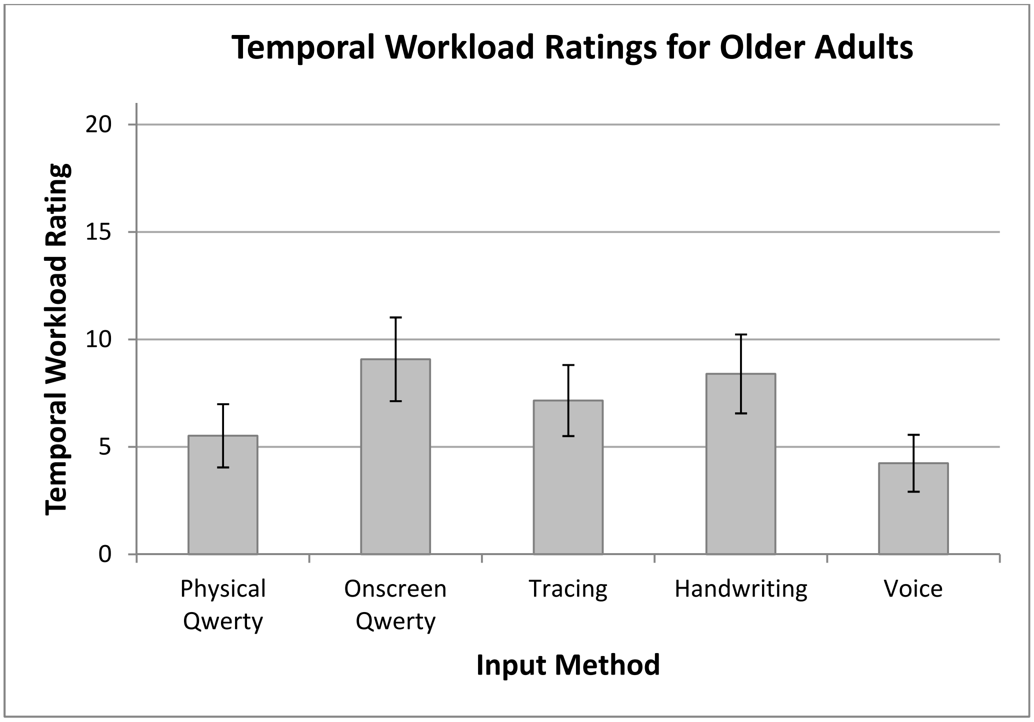


Figure 6.9. Temporal workload ratings for older adults. Error bars represent 95% CI.

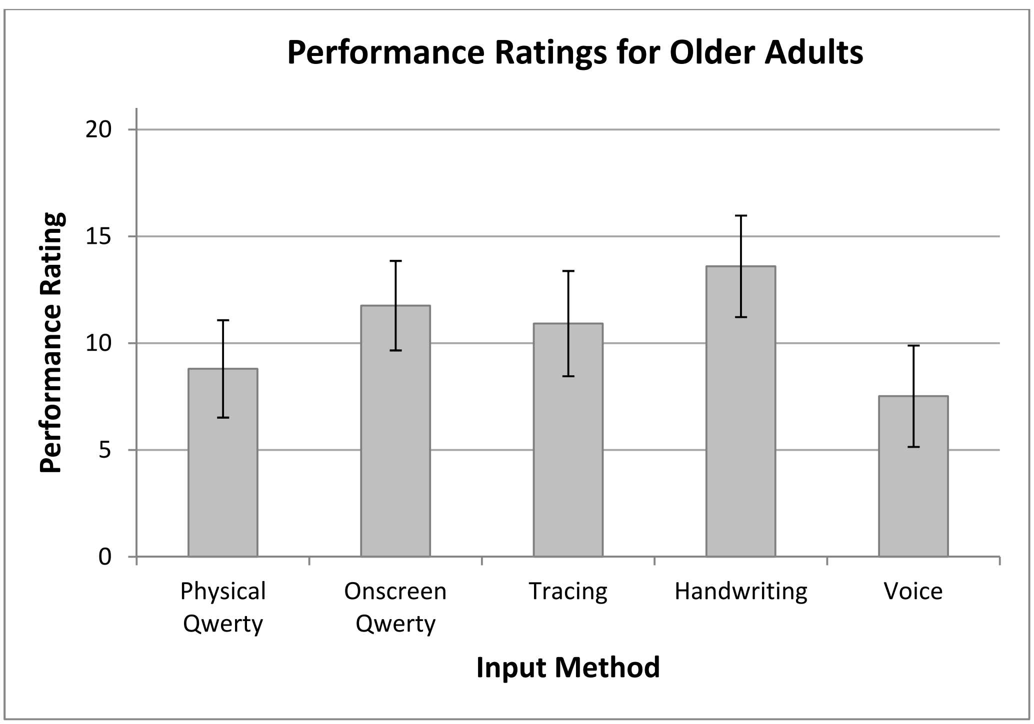


Figure 6.10. Performance dimension ratings for older adults. Error bars represent 95% CI.

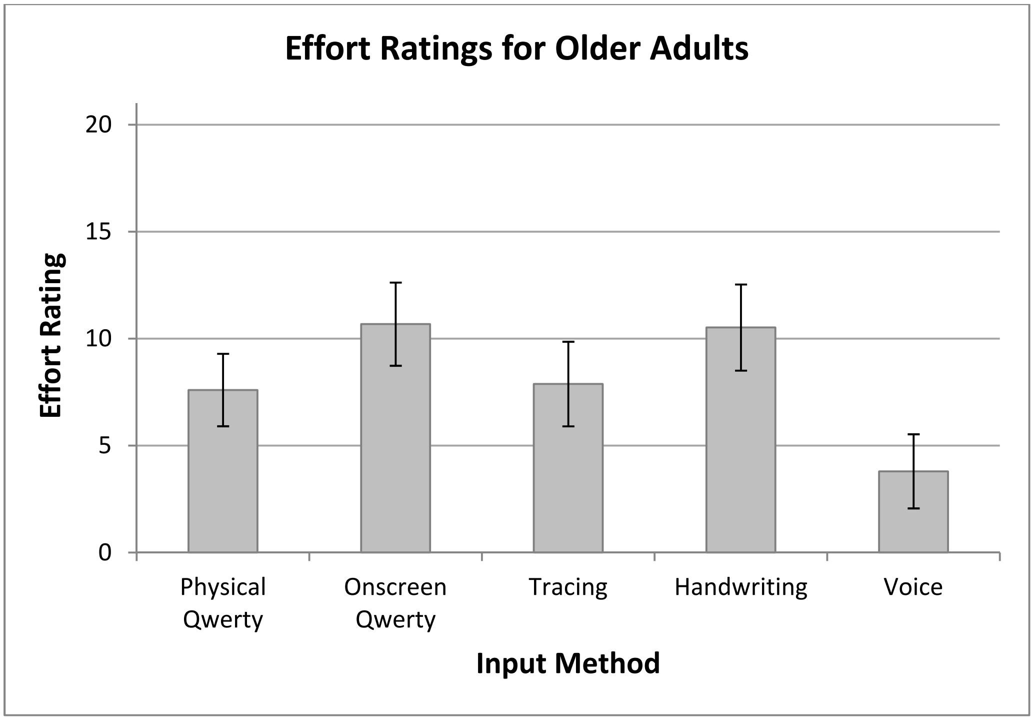


Figure 6.11. Effort dimension ratings for older adults. Error bars represent 95% CI.

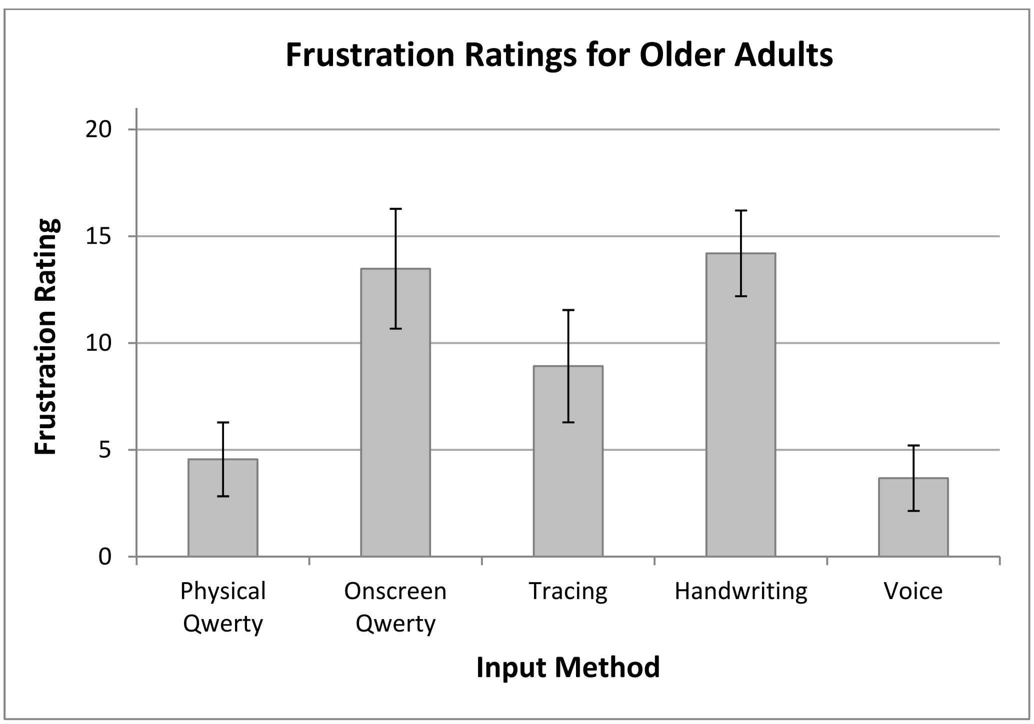


Figure 6.12. Frustration dimension ratings for older adults. Error bars represent 95% CI.

A one-way repeated measures analysis of variance was conducted to evaluate workload scores across input methods. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(9) = 23.45$, p =.005, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .61$). The results showed a large significant main effect for input method, $F(2.42, 58.07) = 19.45$, $p < .001$, $\eta_p^2 = .45$.

Follow-up paired-samples *t* tests indicated that the Voice input method (*M* = 26.36, *SD* = 13.78) was perceived to require significantly less workload than Onscreen Qwerty (*M* = 70.12, *SD* = 26.2), $t(24) = 7.64$, Handwriting (*M* = 73.08, *SD* = 25.48), $t(24) = 8.04$, and Tracing (*M* = 55.04, *SD* = 28.68), $t(24) = 4.46$, all $p < .001$. Physical Qwerty (*M* = 39.92, *SD* = 19.53) had lower reported workload than Onscreen Qwerty, $t(24) = 6.14$, and Handwriting, $t(24) = 5.08$, both $p < .001$. Tracing had less perceived workload than Handwriting, $t(24) = 3.83$, $p < .001$. There were no significant differences between Onscreen Qwerty and Handwriting, $t(24) = .38$, $p = .70$, Onscreen Qwerty and Tracing, $t(24) = 1.71$, $p = .10$, Physical Qwerty and Tracing, $t(24) = 2.16$, $p = .04$, and Physical Qwerty and Voice, $t(24) = 2.80$, $p = .01$ (see Figure 6.13).

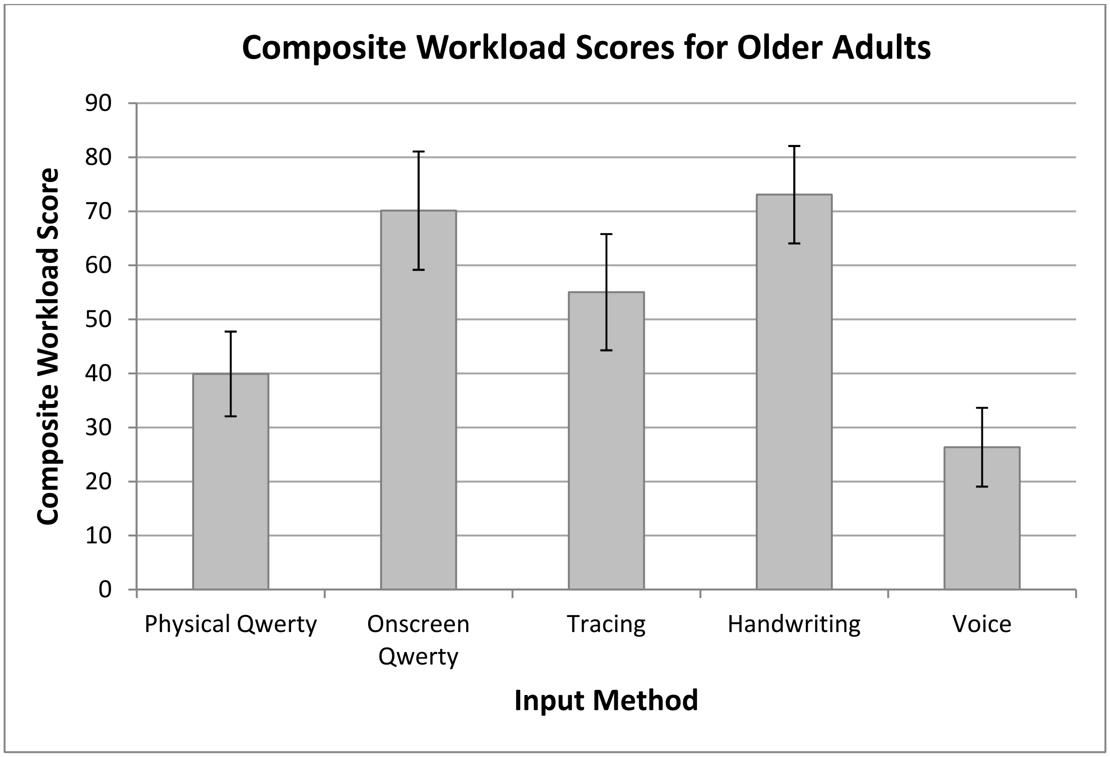


Figure 6.13. Composite workload scores for older adults. Voice elicited the lowest workload scores. Error bars indicate 95% CI.

Preference.

Pre-Test. Prior to using the input methods, participants were asked to rate each input method on a 0 – 50 point scale (0 = least preferred, 50 = most preferred) based on how desirable or valuable they perceived it to be for entering text on a smartphone device. The initial ratings are given in Figure 6.14. It appears that older adults tended to prefer the Voice input method and tended to least prefer the Tracing input method.

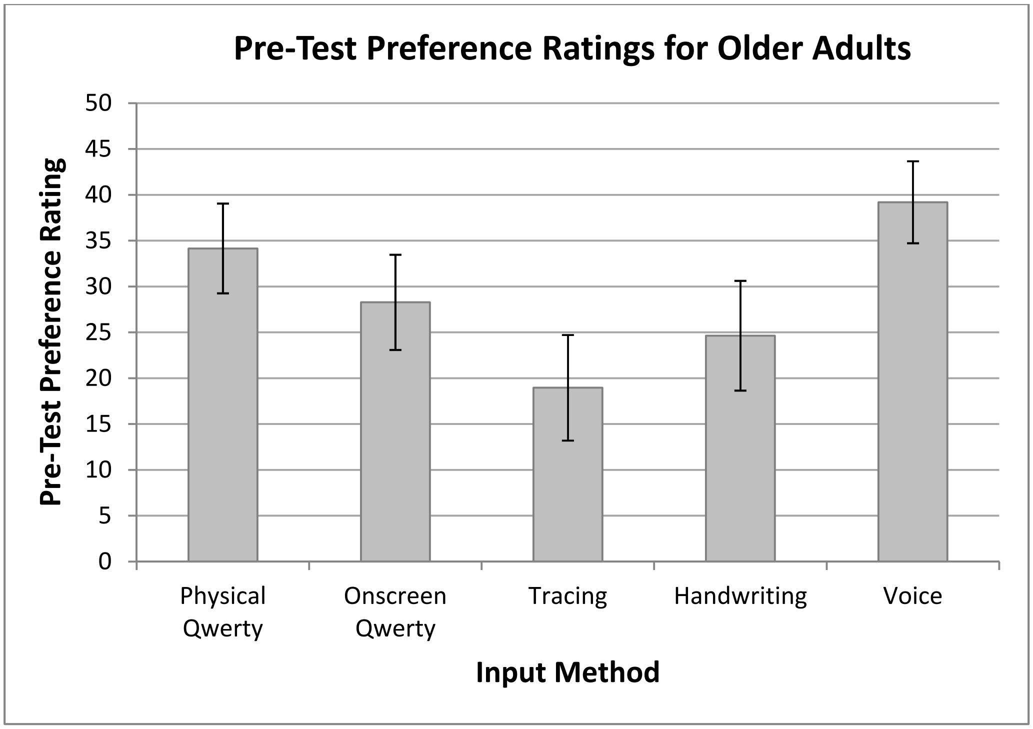


Figure 6.14. Pre-Test Preference ratings for older adults. Error bars represent 95% CI.

Post-Test. Preliminary analyses of post-test preference ratings (scale from 0 – 50, 0 = least preferred, 50 = most preferred) revealed two outlying values, which were recoded (see Appendix M). A one-way repeated measures analysis of variance revealed a large significant main effect for input method, $F(4, 96) = 14.21$, $p < .001$, $\eta_p^2 = .37$.

Follow-up paired-samples *t* tests showed that the Voice ($M = 40.48$, $SD = 8.93$) input method had significantly higher preference ratings than Handwriting ($M = 16.28$, $SD = 12.78$), $t(24) = 8.03$, Onscreen Qwerty ($M = 19.76$, $SD = 15.41$), $t(24) = 4.93$, and Tracing ($M = 24.72$, $SD = 16.73$), $t(24) = 4.42$, all $p < .001$. Physical Qwerty ($M = 36.36$, $SD = 11.26$) had higher preference ratings than Onscreen Qwerty, $t(24) = 4.80$, and Handwriting, $t(24) = 6.28$, both $p < .001$. There were no significant differences in post-test preference ratings between Onscreen Qwerty and Handwriting, $t(24) = .75$, $p = .46$, Onscreen Qwerty and Tracing, $t(24) = .96$, $p = .35$, Physical Qwerty and Voice, $t(24) = 1.28$, $p = .21$, Tracing and Handwriting, $t(24) = 2.27$, $p = .03$, and Physical Qwerty and Tracing, $t(24) = 2.54$, $p = .02$. It appears that older adults most preferred the Voice input method for entering text on a smartphone after having an opportunity to use each of the input methods. Physical Qwerty was rated similar to Voice and tended to be rated higher than the other manual modes of input. Onscreen Qwerty and Handwriting were least preferred. Onscreen Qwerty, Tracing tended to be rated higher than Onscreen Qwerty and Handwriting, however, it was similarly rated compared to Onscreen Qwerty, Handwriting, and Physical Qwerty (see Figure 6.15).

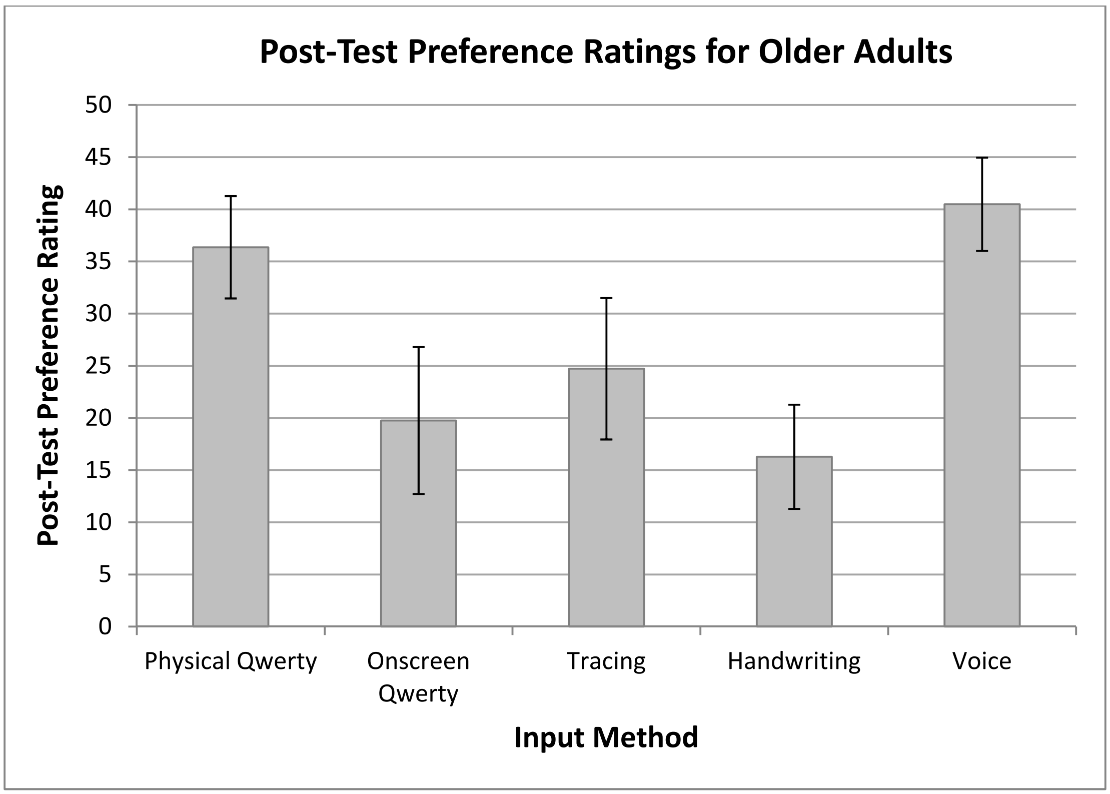


Figure 6.15. Average post-test preference ratings for older adults. Voice and the Physical Qwerty input methods had the highest preference ratings. Bars indicate 95% CI.

Preference Rating Change. It appears that older adults' preference perceptions of the input methods changed slightly from pre-test to post-test for a few of the input methods. Physical Qwerty and Voice were slightly more preferred, but relatively consistent with pre-test ratings, after entering twenty phrases with each input method. Their preference for Tracing seems to have increased, while their preference ratings for Onscreen Qwerty and Handwriting decreased (see Figure 6.16).

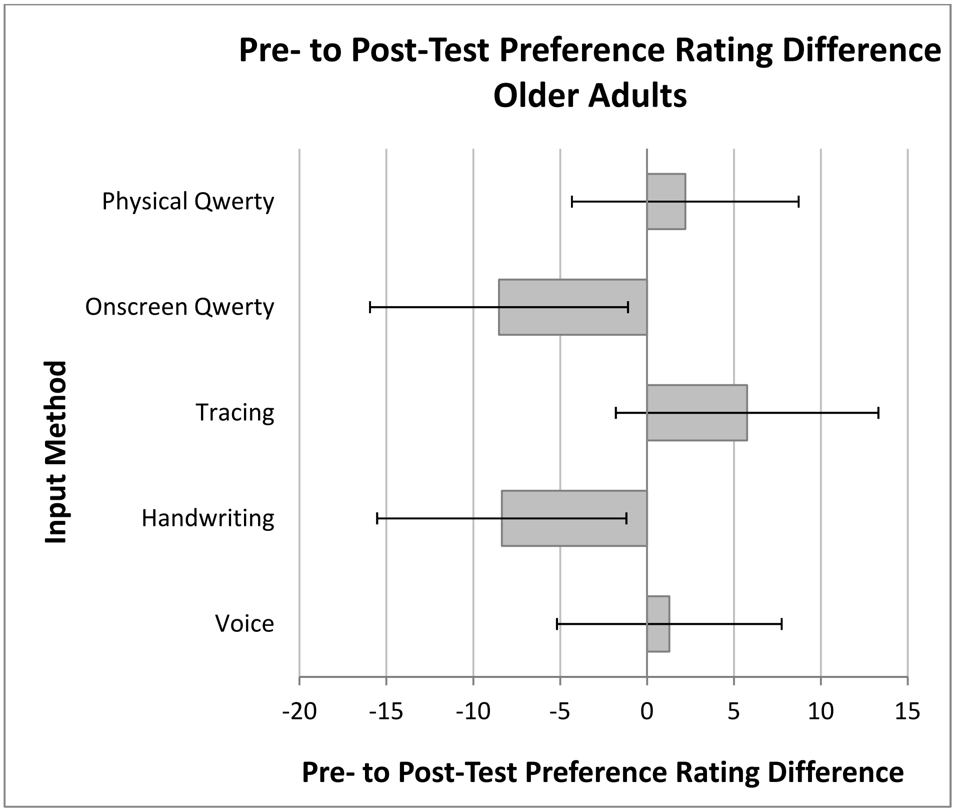


Figure 6.16. Differences in average preference ratings pre- to post-test for older adults. Error bars represent 95% CI.

Participant Comments.

Participant comments regarding initial impressions of each of the input methods were gathered pre-test during the preference ratings (see Appendix B for comment summaries). After completing trials for each method, comments regarding personal likes, dislikes, and suggested changes for each input method were gathered (see Appendix C for summaries of likes, dislikes, and suggested changes). Comments regarding final impressions were gathered post- test (see Appendix D for comment summaries). The most frequently reported impressions for each text input method during the pre- and post-test preference ratings is given in Table 6.1.

TABLE 6.1

PARTICIPANT COMMENT SUMMARY

	Pre-Test	**Post-Test**
Physical Qwerty	Qwerty layout is comfortable and familiar.	Appreciated the tactile and “clickiness” of the physical keys.
	Touch typists like the tactile quality of the keys, and feels it would be fast and accurate	Felt the keys were too small and had too much resistance.
	Concern about the size of the keys on a small device.	Thought it was accurate and comfortable, but slow.
Onscreen Qwerty	Those with tablet experience felt it was a familiar text entry method; others did not.	Struggled with how much force was needed to select the keys.
	No button pushing might be faster.	The keys were small; it was not accurate.
	Concerns about accuracy, having to increase precision and shaky hands.	Took too much energy and concentration, and disliked using one finger.
Tracing	Skeptical about the accuracy of the technology.	Mixed reviews regarding ease of use, entry rate speed, and accuracy.
	Using gestures sounds foreign.	Thought it was fun, enjoyable, novel.
	Concerns about the amount of force needed, and hand shakiness.	Did not like their hand/finger obscuring the keyboard while they were tracing.
Handwriting	Mixed reviews. Those writing legibly felt confident; it would be accurate.	Many struggled; had to change their normal style of writing to increase accuracy.
	Skeptical about the technology.	It was difficult, tiring and frustrating.
	Thought it would be slow; inefficient.	It was a slow way to enter text.
Voice	Felt it would be fast, convenient, beneficial for those with disabilities.	Liked the fast entry rate, ease of use and that it was accurate.
	Thought they would train the software.	Appreciated that it was hands-free.
	Concerns about enunciation, background noise, disturbing others, and privacy.	Concern about privacy and disturbing others, the way it handles homophones.

Hand Function and Handwriting.

A square root transformation was applied to both the Jebsen-Taylor dominant hand function times and the Adjusted Words per Minute for the Handwriting condition distributions. A Pearson’s correlation was computed between the two measures. The correlation coefficient indicates a medium negative relationship, but it was not statistically significant, *r*(23) = -.38, *p* = .06. It appears that participants who wrote slower during the Jebsen-Taylor task tended to have higher text entry rates using the Handwriting input method.

Keyboarding and Smartphone Qwerty Speed.

A square root transformation was applied to the average Keyboarding Speed and Physical and Onscreen Qwerty average Adjusted Words per Minute distributions. Pearson's correlation coefficients were computed between participants' average Keyboarding Speed (Words per Minute) and average Adjusted Words per Minute for the smartphone Physical and Onscreen Qwerty input methods. The results show a large, statistically significant correlation between Keyboarding Speed and Adjusted Words per Minute in the Physical Qwerty condition, *r*(23) = .70, *p* < .001. A small relationship, which was not statistically significant, was shown between Keyboarding Speed and Adjusted Words per Minute in the Onscreen Qwerty condition, *r*(23) = .09, *p* = .66. Participants, who were faster at keying the pangrams using a standard desktop keyboard, tended to also be faster using the smartphone Physical Qwerty.

Hand Function, Anthropometry and Total Error Rate.

Pearson correlation coefficients were computed among averages for each of the hand function, anthropometric measures and average Total Error Rates for each of the physical input methods for each gender. One outlying value on the Grooved Pegboard test was recoded from 156 to 139 (the next highest was 138). The distributions for Total Error Rate were all arcsine square root transformed. Since this was an exploratory investigation of these relationships, the alpha level is not adjusted. The results of these analyses are given in Tables 6.2 and 6.3. In general, it appears that hand anthropometry and the hand function tests that were administered in this study do not have significant relationships with Total Error Rates. However, thumb circumference appears to have a large, negative association with the Onscreen Qwerty Total

Error Rate. As hand dimensions increase, accuracy decreases for the Handwriting input method. In addition, dexterity seems to play a role with the Handwriting input method, where participants who were slower in completing the Grooved Pegboard (worse manipulation dexterity) tended to also have a higher Total Error Rate with the Handwriting input method.

TABLE 6.2

HAND FUNCTION, ANTHROPOMETRY AND TOTAL ERROR RATE FOR FEMALES

	Physical Qwerty	Onscreen Qwerty	Tracing	Handwriting
Average Hand Length	.33	.23	.00	-.02
Average Hand Breadth	-.23	-.11	.19	-.28
Average Thumb Length	.03	.04	-.10	.25
Average Thumb Width	.23	.00	.12	.04
Average Thumb Circumference	-.19	**-.56***	-.35	.14
Average Index Length	.13	.12	-.11	.18
Average Index Width	-.28	.05	.05	.10
Average Index Circumference	.21	.25	.24	.40
Average Thumb Reach	.28	.21	.23	.12
Average Functional Range of Motion	.13	-.14	-.07	.49
Jebsen-Taylor (dominant)	-.13	.07	.17	.40
Grooved Pegboard (dominant)	.25	-.13	.48	**.75****

Note: * indicates significance at the .05 level, and ** indicates significance at the .01 level, (both two-tailed), *n* = 16.

TABLE 6.3

HAND FUNCTION, ANTHROPOMETRY AND TOTAL ERROR RATE FOR MALES

	Physical Qwerty	Onscreen Qwerty	Tracing	Handwriting
Average Hand Length	.14	.17	.37	.29
Average Hand Breadth	-.18	-.03	.02	.00
Average Thumb Length	-.01	.21	.64	.60
Average Thumb Width	.01	.16	.25	.48
Average Thumb Circumference	.11	.15	.17	.34
Average Index Length	.24	.44	.61	**.68***
Average Index Width	.22	.29	**.76***	.62
Average Index Circumference	-.08	.00	.21	.36
Average Thumb Reach	.14	.48	.05	.25
Average Functional Range of Motion	.20	.43	.05	.60
Jebsen-Taylor (dominant)	.21	.10	-.36	-.37
Grooved Pegboard (dominant)	.13	-.01	-.01	-.20

Note: * indicates significance at the .05 level, two-tailed, n = 9.

Hand Function, Anthropometry and Satisfaction.

Pearson correlation coefficients were computed among averages for each of the hand function, anthropometric measures and satisfaction scores for each of the physical input methods for both genders. Since this was an exploratory investigation of these relationships, the alpha level was not adjusted. The results of these analyses are given in Tables 6.4 and 6.5. For female participants, there was a large positive relationship between thumb circumference and Onscreen Qwerty satisfaction scores and a large negative relationship between manipulation dexterity and

Tracing satisfaction scores. For males, there were large negative relationships between index width and the Tracing and Handwriting input methods.

TABLE 6.4

HAND FUNCTION, ANTHROPOMETRY AND SATISFACTION FOR FEMALES

	Physical Qwerty	Onscreen Qwerty	Tracing	Handwriting
Average Hand Length	.00	.11	.07	-.35
Average Hand Breadth	.02	-.12	.38	.10
Average Thumb Length	.28	.38	.00	-.14
Average Thumb Width	.15	.28	.07	-.15
Average Thumb Circumference	.14	**.57***	-.31	.08
Average Index Length	-.09	.22	.05	-.31
Average Index Width	-.17	-.27	-.24	-.34
Average Index Circumference	.01	-.24	-.34	-.19
Average Thumb Reach	.49	.22	-.14	-.08
Average Functional Range of Motion	.27	.27	.02	.28
Jebsen-Taylor (dominant)	-.22	.11	-.13	-.41
Grooved Pegboard (dominant)	-.07	.37	**-.66****	-.23

Note: * indicates significance at the .05 level, and ** indicates significance at the .01 level (both two-tailed), *n* = 16.

TABLE 6.5

HAND FUNCTION, ANTHROPOMETRY AND SATISFACTION FOR MALES

	Physical Qwerty	Onscreen Qwerty	Tracing	Handwriting
Average Hand Length	-.04	-.12	-.36	-.19
Average Hand Breadth	.30	.44	-.43	-.11
Average Thumb Length	.07	-.17	-.09	-.51
Average Thumb Width	-.02	.10	-.36	-.49
Average Thumb Circumference	.18	.33	-.51	-.59
Average Index Length	-.37	-.58	.06	-.25
Average Index Width	.32	.06	**-.67***	**-.76***
Average Index Circumference	.38	.38	-.62	-.66
Average Thumb Reach	-.47	-.14	.37	.16
Average Functional Range of Motion	-.16	.00	-.02	-.47
Jebsen-Taylor (dominant)	-.32	-.10	.44	.44
Grooved Pegboard (dominant)	.42	.54	-.61	-.23

Note: * indicates significance at the .05 level, and ** indicates significance at the .01 level (both two-tailed), n = 9.

Individual Manual Input Techniques.

While participants were entering phrases, the facilitator recorded the digits used (see Table 6.6) and the way that participants supported the device while entering text (see Table 6.7) with each input method. Most older adults used both thumbs with the Physical Qwerty input methods, holding the device with both hands, but using the tabletop to support their hands, wrists or forearms. For the Onscreen Qwerty, Tracing and Handwriting input methods, most participants used a single index finger, and supported the phone on the table while stabilizing the device with fingers on their non-dominant hand. Photographs depicting common hand and device posture for the manual input methods are given in Appendix O.

TABLE 6.6

DIGITS OLDER ADULTS USED TO ENTER TEXT WITH EACH METHOD

	Physical Qwerty	**Onscreen Qwerty**	**Tracing**	**Handwriting**
Single Thumb	4%	0%	0%	0%
Both Thumbs	56%	4%	0%	0%
Single Index	20%	56%	80%	84%
Both Index	12%	12%	0%	4%
Single Middle	0%	12%	8%	4%
Single Ring	0%	4%	4%	8%
Single Pinky	0%	8%	8%	0%
Other	8%	4%	0%	0%

TABLE 6.7

OLDER ADULT MANUAL DEVICE SUPPORT

	Physical Qwerty	**Onscreen Qwerty**	**Tracing**	**Handwriting**
Held in both hands	56%	12%	0%	0%
Held in one hand	12%	28%	36%	44%
Stabilized with one hand	12%	48%	56%	56%
None	20%	12%	8%	0%
Used tabletop	80%	84%	80%	80%
Used lap	8%	0%	8%	8%
No support	12%	16%	12%	12%

Speech and Voice Characteristics and Voice Recognition.

A logarithmic transformation was applied to scores in the Frequency and Amplitude Tremor Intensity distributions, and square root transformation was applied to scores in the Pitch and Amplitude Period Perturbation distributions. Pearson correlation coefficients were computed among Speech Rate and voice quality measures, with average Adjusted Words per Minute and Word Error Rate within the Voice condition for both genders. Results showed no significant relationships between any of the measures for either gender, see Tables 6.8 and 6.9.

TABLE 6.8

SPEECH AND VOICE CHARACTERISTICS AND VOICE RECOGNITION FOR FEMALES

	Average Voice AdjWPM	Voice Word Error Rate
Speech Rate (WPS)	.40	-.46
Average Fundamental Frequency (Hz)	.01	.11
Frequency Tremor Intensity Index (%)	-.31	-.21
Amplitude Tremor Intensity Index (%)	-.02	-.32
Pitch Perturbation Quotient (%)	.01	-.49
Amplitude Perturbation Quotient (%)	-.02	-.24
Voice Turbulence Index	.42	-.22

Note: *n* = 16.

TABLE 6.9

SPEECH AND VOICE CHARACTERISTICS AND VOICE RECOGNITION FOR MALES

	Average Voice AdjWPM	Voice Word Error Rate
Speech Rate (WPS)	.38	-.25
Average Fundamental Frequency (Hz)	.11	-.08
Frequency Tremor Intensity Index (%)	-.17	-.13
Amplitude Tremor Intensity Index (%)	.33	-.27
Pitch Perturbation Quotient (%)	.28	-.44
Amplitude Perturbation Quotient (%)	-.13	.13
Voice Turbulence Index	.05	-.03

Note: *n* = 9.

Discussion

The results from Study 2 lend support to the hypotheses that the type of smartphone input method would impact performance metrics, as well as subjective satisfaction, workload, and preference ratings for novice older adults. The results also suggest that personal characteristics may play an important role in determining aspects of user performance and perceptions of these input methods, however, reasoning for the observation of these relationships is speculative.

Performance.

***Adjusted Words per Minute*.** Input method type had a large effect on Adjusted Words per Minute, accounting for most of the variability in the scores. Voice clearly had the highest entry rate, followed by Physical Qwerty, then Tracing, then Onscreen Qwerty and, lastly, Handwriting. It makes sense that Voice was the fastest, as human rate of speech is much faster than any form of manual input, approaching 200 WPM (Cox et al., 2008). It was also not surprising that Handwriting was the slowest method, as natural handwriting typically does not exceed 20 WPM (Bailey, 1996). However, the significant difference between Physical and Onscreen Qwerty was unexpected since the act of text input with each method would appear to be similar. One reason that this may have occurred is that the Physical Qwerty provided kinesthetic cues, including button “travel” (i.e., movement) on each key press and bumps on the “f” and “j” keys, which may have provided some assistance in quickly locating the correct keys. The finding that the entry rate for Tracing was faster than Onscreen Qwerty was also unexpected, as gesturing to enter text is a new concept and could have potentially taken longer to learn.

It was expected that there would be significant differences, and though differences between all input methods were significant, they were not entirely in line with the results of Castellucci and MacKenzie (2011), and the meta-analysis done by Arif and Stuerzlinger (2009).

The results by Castellucci and MacKenzie (2011) showed of the three methods they evaluated, Onscreen Qwerty had the highest Words per Minute, followed by Tracing and then Handwriting. The meta-analysis figured that a thumb mini-Qwerty would have higher entry rates than an On-Screen Qwerty or Handwriting ,which agrees with the results from this study, however, the Words per Minute are quite different than values obtained in this study. Possible explanations for this are that the metric used in this study was Adjusted Words per Minute, which corrects for errors left in the final transcription, and results from the other studies report Words per Minute. Another possible explanation is that Castellucci and MacKenzie's study (2011) used different hardware and input method software than what was used in this study. Additionally, the meta-analysis by Arif and Stuerzlinger included mixed levels of expertise, types of devices and software, and different protocols and tasks. Another plausible explanation is that the sample used in this study was comprised of older adults (aged 60 – 82), and the other studies utilized convenience samples, typically college undergraduates.

Error Rates. At the character-level, the results support the hypotheses that input method type has a significant impact on the three character-level error rates (corrected, uncorrected and total). Results indicated that Handwriting had significantly higher Corrected and Total Error Rates, followed by Onscreen Qwerty, Tracing and Physical Qwerty, with the lowest Corrected and Total Error Rates. This trend was also seen in the Uncorrected Error Rates, where the Physical Qwerty input method had the lowest error rates. The same pattern for these error rates was found by Castellucci and MacKenzie (2011) for Handwriting, Onscreen Qwerty and Tracing.

A caveat to Corrected Error Rate is that the equation does not differentiate between corrected errors for items that were initially correct and those which were in error. Thus, it can only be interpreted that participants deleted more characters with Handwriting and Onscreen Qwerty than with Tracing and Physical Qwerty. The trend in Uncorrected Error Rates implies that the input methods with higher error rates (Handwriting and Onscreen Qwerty) might have been more laborious to make corrections than those which had lower Uncorrected Error Rates (Tracing and Physical Qwerty). However, there may have been differences inherent to these input methods that made errors less noticeable. For example, direct observations showed that participants would on occasion write characters in the Handwriting input method which resembled the intended character, but that returned an erroneous character ("l" instead of "i"). Additionally, older participants tended to "dot" their lowercase "i's" and "j's" after completing a word, rather than while making the character. This would typically return extraneous punctuation, which led to higher error rates. Another plausible explanation is that the older participants did not attend to the output before entering the phrase. This could have occurred from the participant trusting the feedback was accurate (the pop-up letters and the green trace for Handwriting), or simply not having enough cognitive resources to focus on the output. It may have also happened due to erroneous autocorrections. The difference in error rates between the Physical and Onscreen Qwertys is, again, somewhat surprising given the similar nature of text input for the two methods. An explanation could be that, even though novice typists have been shown to rely more on visual than proprioceptive feedback on a regular keyboard (Klemmer, 1971), the detection of errors on the touchscreen Qwerty may have been more difficult. Indeed, when proprioceptive cues are not available, a greater reliance on visual cues occurs (Barrett & Krueger, 1994).

Results from Study 2 showed that there was only one significant difference among input methods for Word Error Rates. The Voice input method had significantly fewer uncorrected word errors than Handwriting. It is interesting that this is the only significant difference, since there were more significant differences between uncorrected errors at the character-level. One possible explanation for this is that older adults may have been more conscientious about correcting errors that changed the interpretation of the output. Another explanation is that once they made an error, they may have been more likely to make another error in the same word, which would have counted as multiple character-level errors, but only a single word error. The percentage of word errors in all categories was moderately low, ranging from roughly 2 – 7%, which may have been viewed as an acceptable level of error by the participants.

***Satisfaction*.** Results from Study 2 lend support for the hypothesis that input method type had a significant effect on reported Satisfaction levels. Results suggest that Physical Qwerty and Voice are the most satisfying input methods to use for entering text on a smartphone, followed by Tracing, Onscreen Qwerty, and Handwriting. One reason that these results may have occurred is that, as the participants were all novices, the Tracing, Onscreen Qwerty and Handwriting input methods may have seemed more foreign to them. Additionally, participants may have felt that the Physical Qwerty and Voice input methods were more accurate, which likely influenced their experience.

Workload. The results from Study 2 lend support for the hypothesis that input method had a significant effect on perceived workload. The Voice input method had significantly lower workload ratings than all of the manual input methods, except Physical Qwerty which was rated slightly higher. Handwriting had significantly higher levels of reported workload compared to all of the other input methods, except Onscreen Qwerty. Though these input methods have never been directly compared before, some of these results were expected. Handwriting has a high physical investment in each letterform, since each letter could be comprised from one to four strokes (as is the case for a lowercase “l” or a lowercase “w”), depending on writing style. Voice has a very low physical investment, in that speech is not produced at the character or stroke-level. Both of these anticipations were indeed revealed in the average scores for the physical dimension. However, the significant difference in workload between Physical Qwerty and Onscreen Qwerty was not expected. Participants frequently commented on the physical keyboard, about how much they liked having tactile feedback, but also complained about the high key resistance. They also commented on the Onscreen Qwerty and how they appreciated not having to depress any keys. Thus, even though the two Qwerty methods scored similar on the Effort dimension, it is interesting that the Onscreen Qwerty scored higher on the Physical dimension of workload. Other contributing factors to high workload ratings for the Onscreen Qwerty, compared to its physical counterpart, include higher perceptions of frustration, mental workload, and the perception that they performed more poorly with this method. Tracing performed similarly to both Qwertys, which is surprising considering participant comments regarding their hand obscuring the screen while they were trying to trace and that many mentioned that they disliked having their finger in contact with the screen for extended periods of time.

Preference. Participant ratings pre-test showed that they favored the Voice and Physical Qwerty methods over the others, mainly because they seemed to feel more comfortable and familiar with each method. They reported that they were skeptical of technologies that used “behind the scenes” processing, which they perceived as taking the control away from the user, and modes of input that were difficult to conceptualize.

After the users had experienced each of the input methods, they shifted their opinions somewhat. Physical Qwerty and Voice were rated fairly consistently from pre- to post-test, which may indicate that these methods met the participants' expectations. They reported that they liked the tactile quality of the Physical Qwerty and the speed and ease of use of the Voice input method. Onscreen Qwerty and Handwriting were rated lower than they were during the pre-test preference ratings. This could be due to a more disappointing experience with each of these methods than anticipated. Indeed, many participants reported that they struggled with both methods because they easily made mistakes with them. Onscreen Qwerty was said to have inconsistent responsiveness across the keyboard. In fact, on many occasions the participant would hit the bezel of the device rather than the space bar. At times, the participant would accidentally select a "shift" or "mode shift" button instead of a letter key, which would lead to confusion. Some were also confused about the concept of autocorrect, and thought that they had actually typed a different word. Handwriting was frustrating to many participants because they reported they had to change their writing style to use it. Many of the older adults in this sample wrote in cursive or in mixed case naturally, and they were confused if the software did not recognize their letterforms. Tracing was seen in a more positive light post-test, which could be due to a more positive experience with the technology than anticipated.

Hand Function and Handwriting. It was hypothesized that there would be a significant relationship between the Jebsen-Taylor hand function test for the dominant hand and Handwriting Adjusted Words per Minute. The results from the Pearson's correlation revealed a moderate, yet not significant, negative relationship. One explanation may be that participants who wrote slower and more deliberately may have had more accurate results with the Handwriting input method.

Keyboarding and Smartphone Qwerty Entry Rate. It was hypothesized that Keyboard entry rate would be significantly related to smartphone Qwerty entry rates. Results from the

Pearson's correlations between Keyboarding Speed and both of the Qwerty input methods indicated a large positive relationship with the Physical Qwerty method, but a small, non-significant, relationship with Onscreen Qwerty. This was not surprising, as familiarity and experience with the Qwerty layout may have given users an advantage over those who were not familiar. However, it was surprising that desktop entry rates did not seem to be related to Onscreen Qwerty entry rates, in that this input method utilizes the same layout. Perhaps it can instead be explained by the type of hardware used by each method, the physical keyboard and the touchscreen. Though both smartphone methods are quite different from a desktop keyboard, the act of pushing a button and the audio and tactile feedback received with the Physical Qwerty is more similar to keying using a desktop keyboard than the Onscreen Qwerty.

Hand Function, Anthropometry and Total Error Rate. The results from the Pearson's correlations between Hand Function (Jebsen-Taylor, Grooved Pegboard, Thumb Reach, and Functional Range of Motion), Anthropometry (hand length and breadth, thumb length, width, and circumference, and index length, width, and circumference) and error rates for all of the manual input methods showed two significant relationships with the female participants, and two significant relationships with the male participants.

The large, negative relationship between thumb circumference and error rate suggests that female participants with larger thumbs tended to create fewer errors with the Onscreen Qwerty method. This is puzzling because only one older female used both thumbs while entering text on this input method, and it appears that there are no significant relationships with any of the other hand dimensions and error rate with this input method. However, this may not be directly related to the digit used for input but rather for the digits used to support the device. A female with thicker thumbs may have been able to get a better grip on the device, and most females with greater thumb circumference did hold the device in their hand or supported on the table in a type of pinch grip between their thumb and one or two other fingers. Additionally, females who had worse manipulation dexterity appeared to have higher error rates with the Handwriting input method. A possible explanation is that the act of creating letterforms may have required greater dexterity than the other methods.

Based on the large positive relationships between male participants' index length and width and the Tracing and Handwriting input methods, it appears that male participants had more errors if they had longer and wider index dimensions. For the Tracing input method, this might indicate that male participants with wider index fingers may have had less precision, and more frequently selected unintended keys, or that their finger obscured the interface more than those with more narrow index fingers, making it difficult to see the keys. Additionally, longer index fingers were associated with higher error rates for male participants. One explanation could be that those who had longer index fingers were more capable of creating larger letterforms, which could have been interpreted by the software as capital letters. Though capital letters were not counted as errors, many participants corrected them. Thus, these would have inflated the Corrected Error Rate, and also Total Error Rate.

Hand Function, Anthropometry and Satisfaction. Results from the Pearson's correlations between Hand Function (Jebsen-Taylor, Grooved Pegboard, Thumb Reach, and Functional Range of Motion), Anthropometry (hand length and breadth, thumb length, width, and circumference, and index length, width, and circumference) and error rates for all of the manual input methods revealed two significant relationships with the female participants and two significant relationships with the males.

The large, positive relationship between thumb circumference and satisfaction scores suggests that female participants with larger thumbs tended to be more satisfied with the Onscreen Qwerty method. This is in line with the Total Error Rate correlations, which indicated that greater thumb circumference was related to fewer errors with the Onscreen Qwerty method. Additionally, females who had greater manipulation dexterity appeared to have lower satisfaction scores with the Tracing input method. One possible explanation for this finding could be the style of the trace afforded by greater dexterity. It was observed that some participants tended to trace more direct (from letter to letter) while others had more extravagant, loose traces. It is plausible that the participants who did the "freeform" traces may have done so simply because they were able to. A more direct trace movement tended to perform better, in terms of accuracy.

Based on the large negative relationships between male participants' index width and Tracing and Handwriting input methods, it appears that male participants were less satisfied if they had wider index dimensions. For the Tracing input method, this might indicate that male participants with wider index fingers may have had less precision, or that their finger obscured the interface more than those with more narrow index fingers, making it difficult to see the keys. An explanation regarding the Handwriting method may be that those who had wider index fingers were less precise when creating letterforms, which could have led to more errors.

Speech and Voice Characteristics and Recognition. It was predicted that significant relationships would be found between the collected speech and voice measures and the Adjusted Words per Minute and Word Error Rates for the Voice condition. However, results from Study 2 suggest that there were no significant relationships between measures for either gender, or perhaps that the relationship between measures was nonlinear, or that the sample size was too small to detect an effect.

CHAPTER 7

COMPARISON BETWEEN YOUNGER AND OLDER NOVICE PARTICIPANTS

Purpose

Differences between input methods on many of the dependent measures can clearly be seen within the results from Studies 1 and 2. The next logical step would be to assess whether these behavioral patterns, observed on the performance measures and subjective ratings, were consistent to both the younger and older samples or if they differed. These measures have not been examined within the context of smartphone input methods between age groups, therefore, this investigation is exploratory. The purpose is to describe observed behavior between these age groups, and possibly also indicate whether certain input methods may be better suited for either age group.

Results

Since these comparisons are exploratory in nature, no specific *a priori* hypotheses were generated. Type I error was not strictly controlled, therefore, the alpha level remained at .05 for each comparison. Error bars represented within the graphs represent ± 1 Standard Error of the Mean.

Adjusted Words per Minute.

The distributions were square root transformed, and a 2 x 5 mixed factorial analysis of variance was conducted, with age group (young versus older) as the between subjects factor and input method type (Physical and Onscreen Qwerty, Tracing, Handwriting, and Voice) as the within subjects factor. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(9) = 40.68$, $p < .001$, therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = .81$). The results showed significant main effect of input method, $F(3.24, 155.6) = 553.55$, $p < .001$, $\eta_p^2 = .92$, a significant main effect of age group, $F(1, 48) = 53.59$, $p < .001$, $\eta_p^2 = .53$, and a significant interaction between input method and age group, $F(3.24, 155.6) = 14.88$, $p < .001$, $\eta_p^2 = .24$.

Analysis of simple main effects revealed that older participants had a slower average Adjusted Words per Minute when using the Physical Qwerty, $F(1,139) = 64.58$, $p < .001$, Onscreen Qwerty, $F(1,139) = 48.58$, $p < .001$, Tracing, $F(1,139) = 53.64$, $p < .001$, and Handwriting, $F(1,139) = 9.94$, $p = .002$, input methods. There was no effect for age when using the Voice input method, $F(1,139) = 1.89$, $p = .17$ (see Figure 7.1). Additionally, significant differences were discovered between input methods for younger adults, $F(4,192) = 248.58$, $p < .001$, as well as older adults, $F(4,192) = 319.86$, $p < .001$.

Follow-up tests were conducted to evaluate ten pairwise differences among the means for younger adults. All of the input methods were statistically different from each other. The Voice input method had a statistically higher Adjusted Words per Minute rate than each of the other input methods. Physical Qwerty had a lower Adjusted Words per Minute rate than Voice, but higher than Onscreen Qwerty, Tracing, and Handwriting. Tracing had a lower Adjusted Words per Minute rate than Voice and Physical Qwerty, but higher than Onscreen Qwerty and Handwriting. Onscreen Qwerty had a higher rate than Handwriting, but lower than the other three input methods.

Follow-up tests were also conducted to evaluate ten pairwise differences among means for older adults. Again, all of the input methods were statistically different from one another. The Voice input method had a statistically higher Adjusted Words per Minute rate than each of the other input methods. Physical Qwerty had a lower Adjusted Words per Minute rate than Voice, but higher than Onscreen Qwerty, Tracing, and Handwriting. Tracing had a lower Adjusted Words per Minute rate than Voice and Physical Qwerty, but higher than Onscreen Qwerty and Handwriting. Onscreen Qwerty had a higher rate than Handwriting, but lower than the other three input methods.

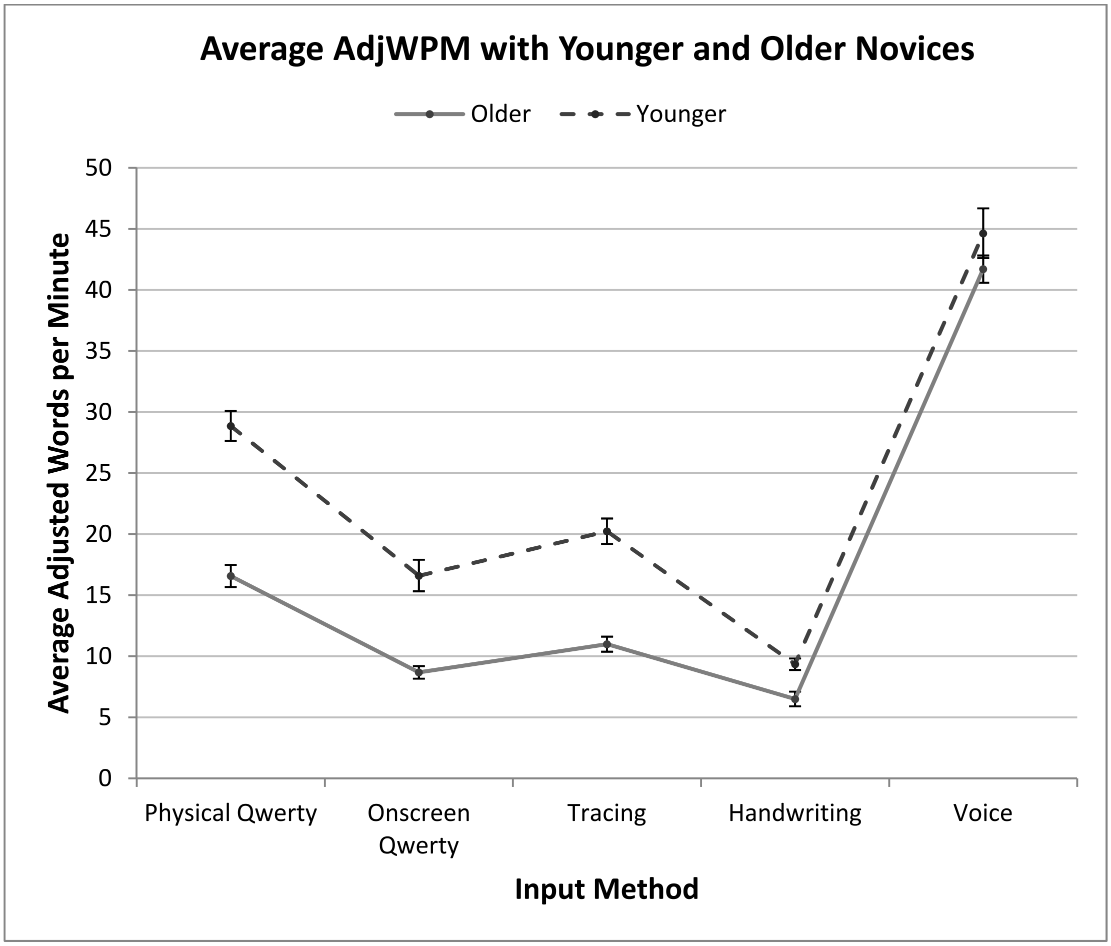


Figure 7.1. Average Adjusted Words per Minute for Younger and Older Novice Adults. No significant differences were found between age groups for Voice. Bars indicate ± 1 SEM.

Error Rates.

The manual input methods (Physical and Onscreen Qwerty, Tracing, and Handwriting) all use methods to produce text at the character-level, thus the Unified Error Metric (Corrected, Uncorrected, and Total Error Rates) is used to describe accuracy for these. Since the Voice input method is a word-level entry method, the comparison in accuracy across all input methods is only done using the Word Error Rate. Thus, the accuracy of the input stream is not analyzed for Voice.

Corrected Error Rate. An arcsine square root transformation was applied to the Physical and Onscreen Qwerty, Tracing, and Handwriting distributions. A 2 x 4 mixed factorial analysis of variance was conducted, with age group (Young versus Older) as the between subjects factor and input method type (Physical Qwerty, Onscreen Qwerty, Tracing, and Handwriting) as the within subjects factor. The results showed a significant main effect of input method, $F(3, 144) = 141.13$, $p < .001$, $\eta_p^2 = .75$. There was not a significant main effect of age group, $F(1, 48) = 1.13$, $p = .29$, $\eta_p^2 = .02$, or a significant interaction between input method and age group, $F(3, 144) = 1.46$, $p = .23$, $\eta_p^2 = .03$.

Follow-up analyses to the input method main effect consisted of all pairwise comparisons among the four types of input methods, ignoring the effect of age. The results of this analysis indicated that there were significant differences between each of the input methods. The Handwriting input method had the highest Corrected Error Rate. The next highest error rate was obtained by the Onscreen Qwerty, which was higher than Tracing and Physical Qwerty. The Physical Qwerty input method had the lowest Corrected Error Rate (all $p < .001$) (see Figure 7.2).

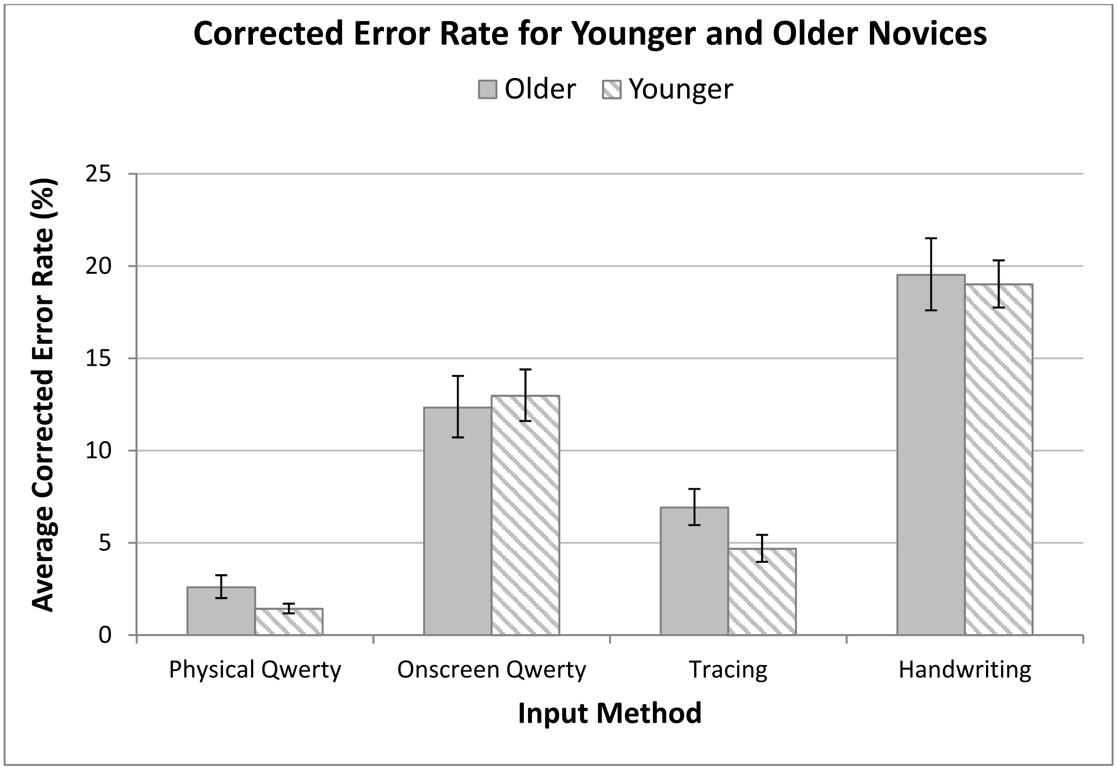


Figure 7.2. Average Corrected Error Rate for Younger and Older Novice Adults.
An input method main effect revealed significant differences between all input methods.
Error bars indicate ± 1 SEM.

Uncorrected Error Rate. Preliminary analyses of Uncorrected Error Rates revealed five outlying percentages, which were all recoded (see Appendix M). Following an arcsine square root transformation, a 2 x 4 mixed factorial analysis of variance was conducted, with age group (young versus older) as the between subjects factor and input method type (Physical and Onscreen Qwerty, Tracing and Handwriting) as the within subjects factor. The results showed a significant main effect of input method for uncorrected error rate across input methods, $F(3, 144) = 22.09$, $p < .001$, $\eta_p^2 = .32$, and a significant main effect for age, $F(1, 48) = 17.81$, $p < .001$, $\eta_p^2 = .27$. There interaction between input method and age was marginally significant, $F(3, 144) = 3.94$, $p = .05$, $\eta_p^2 = .08$.

Analysis of simple main effects revealed that older participants had a higher uncorrected error rate for most of the manual input methods, Onscreen Qwerty, $F(1,155) = 8.09$, $p = .005$, Tracing, $F(1,155) = 15.10$, $p < .001$, and Handwriting, $F(1,155) = 12.40$, $p < .001$. There was no effect for age when using the Physical Qwerty input method, $F(1,155) = 1.35$, $p = .25$ (see Figure 7.3). Additionally, significant differences were discovered between input methods for younger adults, $F(3,144) = 7.20$, $p < .001$, as well as older adults, $F(3,144) = 16.83$, $p < .001$.

Follow-up analyses consisted of six pairwise comparisons among the four types of input methods. Results indicated the uncorrected error rate for Onscreen Qwerty was not significantly lower than for Handwriting input method ($p = .30$), but it did have a significantly higher error rate than both Tracing and Physical Qwerty. Physical Qwerty did not have a significantly lower error rate than Tracing ($p = .25$), but it did have a lower error rate than Onscreen Qwerty and Handwriting.

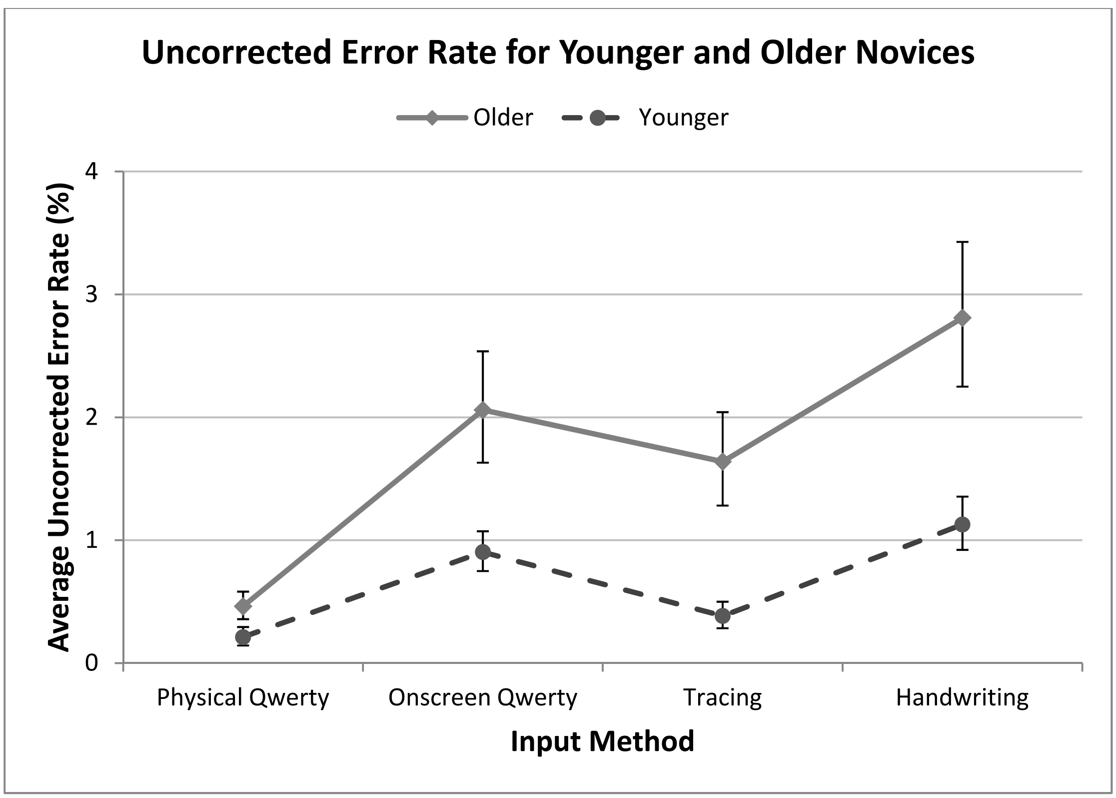


Figure 7.3. Average Uncorrected Error Rate for Younger and Older Novice Adults. Handwriting had the highest Uncorrected Error Rate, Physical Qwerty had the lowest. Bars are ± 1 SEM.

Total Error Rate. An arcsine square root transformation was applied to the distributions. A 2 x 4 mixed factorial analysis of variance was conducted, with age group (young versus older) as the between subjects factor and input method type (Physical and Onscreen Qwerty, Tracing, and Handwriting) as the within subjects factor. The results showed a significant main effect of input method for total error rate across input methods, $F(3, 144) = 167.89$, $p < .001$, $\eta_p^2 = .78$, and a significant main effect for age group, $F(1, 48) = 4.89$, $p = .03$, $\eta_p^2 = .09$. There was not a significant interaction between input method and age group, $F(3, 144) = .01$, $p = .93$, $\eta_p^2 = .00$.

The age main effect indicated that older adults had greater Total Error Rates than younger adults. Follow-up analyses to the input method main effect consisted of all pairwise comparisons among the four types of input methods. All comparisons between input methods were statistically significant ($p < .001$). Handwriting had the highest Total Error Rate, followed by Onscreen Qwerty, then Tracing, and then Physical Qwerty, which was lowest (see Figure 7.4).

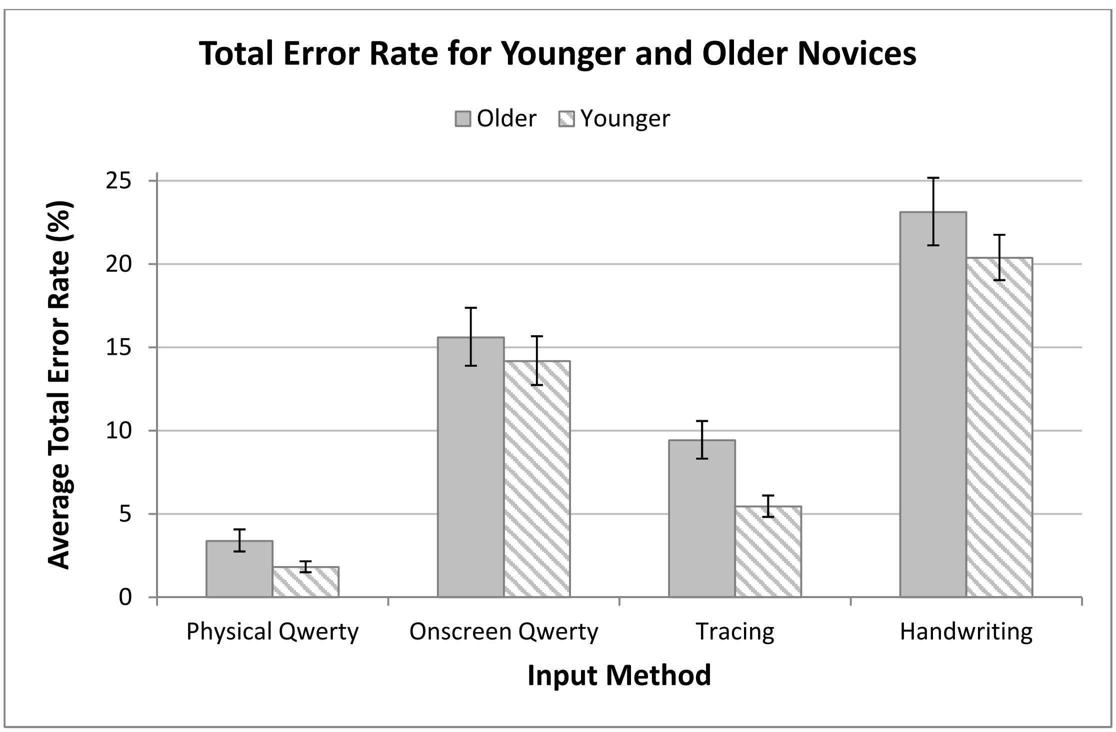


Figure 7.4. Average Total Error Rate for Younger and Older Novice Adults. Handwriting had the highest Total Error Rate, and Physical Qwerty had the lowest. Bars indicate ± 1 SEM.

Word Error Rate. Preliminary analyses of average Word Error Rate distributions revealed nine outlying percentages, which were recoded (see Appendix M). An arcsine square root transformation was applied to each distribution, then a 2 x 5 mixed factorial analysis of variance was conducted, with age group (young versus older) as the between subjects factor and input method type (Physical and Onscreen Qwerty, Tracing, Handwriting, and Voice) as the within-subjects factor. The results revealed a significant main effect of input method for word error rate across input methods, $F(4, 192) = 3.77$, $p = .006$, $\eta_p^2 = .07$, a significant main effect for age group, $F(1, 48) = 10.48$, $p = .002$, $\eta_p^2 = .18$, and a significant interaction between input method and age group, $F(4, 192) = 3.79$, $p = .005$, $\eta_p^2 = .08$.

Analysis of simple main effects revealed that older participants had a greater Word Error Rate than younger participants when using the Onscreen Qwerty, $F(1,182) = 6.34$, $p = .01$, Tracing, $F(1,182) = 9.12$, $p = .003$, and Handwriting, $F(1,182) = 15.85$, $p < .001$, input methods. There was no effect for age when using Voice, $F(1,182) = .07$, $p = .78$, or the Physical Qwerty, $F(1,182) = 1.70$, $p = .19$, input methods. Additionally, significant differences were discovered between input methods for older adults, $F(4,192) = 6.24$, $p < .001$, but not for younger adults, $F(4,192) = .70$, $p = .59$.

Follow-up tests were conducted to evaluate ten pairwise differences among the input methods means for older adults. Results indicated that Handwriting had a significantly higher Word Error Rate than Physical Qwerty and Voice, Tracing had a significantly higher Word Error Rate than Physical Qwerty and Voice, and Onscreen Qwerty had a significantly higher Word Error Rate than Voice. There were no significant differences between Physical Qwerty and Voice, Onscreen Qwerty and Tracing, Onscreen Qwerty and Handwriting, and Tracing and Handwriting (see Figure 7.5).

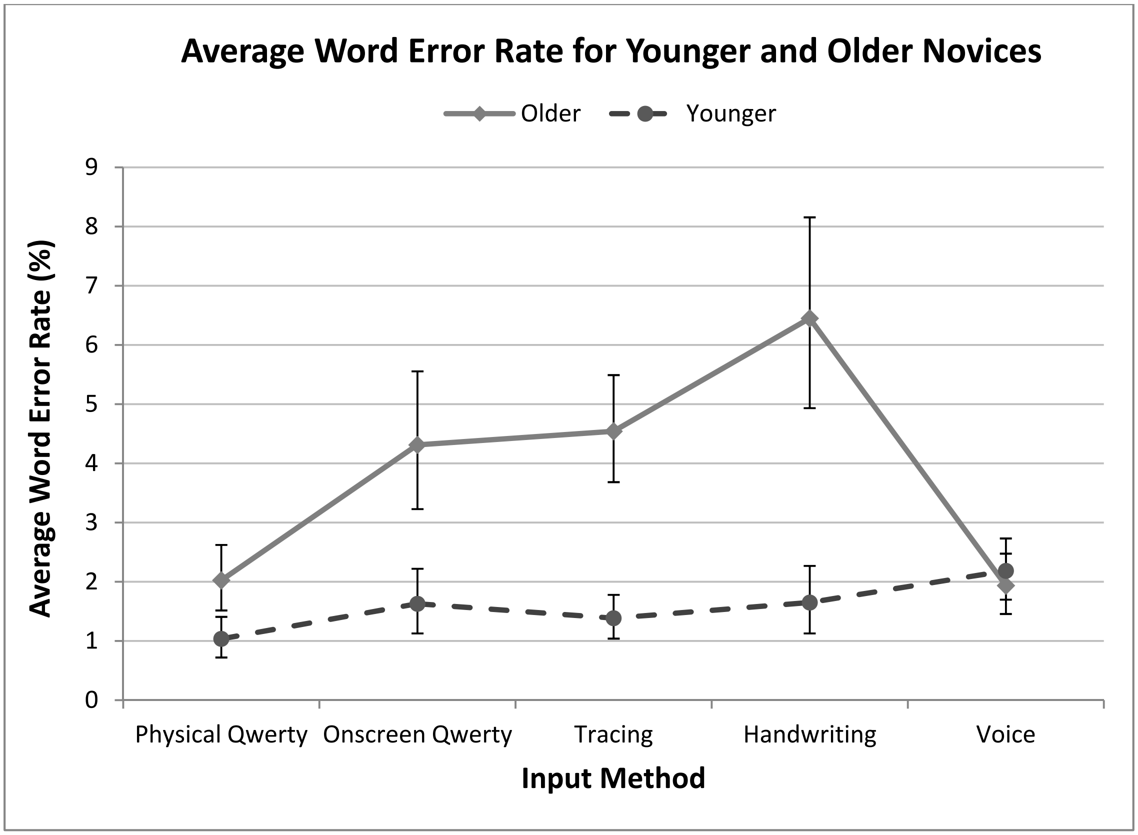


Figure 7.5. Average Word Error Rate for Younger and Older Adults. Bars indicate ± 1 SEM.

Satisfaction.

A preliminary investigation revealed four outliers, which were recoded (see Appendix M). A 2 x 5 mixed factorial analysis of variance was conducted, with age group (young versus older) as the between subjects factor and input method type (Physical and Onscreen Qwerty, Tracing, Handwriting, and Voice) as the within subjects factor. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(9) = 41.32$, $p < .001$, therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = .75$). The results showed significant main effect of input method for satisfaction, $F(3, 144.05) = 42.56$, $p < .001$, $\eta_p^2 = .47$, a significant main effect of age group, $F(1, 48) = 12.95$, $p = .001$, $\eta_p^2 = .21$. The interaction between satisfaction and age group was also significant, $F(3, 144.05) = 2.92$, $p = .036$, $\eta_p^2 = .06$.

Analysis of simple main effects revealed that younger participants reported significantly higher satisfaction scores than older participants for the Onscreen Qwerty, $F(1, 215) = 17.51$, $p < .001$ and Tracing, $F(1, 215) = 12.74$, $p < .001$ input methods. There was no effect for age when using Voice, $F(1, 215) = .04$, $p = .85$, Handwriting, $F(1,215) = .54$, $p = .46$, or the Physical Qwerty input method, $F(1, 215) = 3.68$, $p = .06$. Additionally, significant differences were discovered between input methods for younger adults, $F(4,192) = 21.11$, $p < .001$, and for older adults, $F(4,192) = 24.37$, $p < .001$ (see Figure 7.6).

Follow-up tests were conducted to evaluate ten pairwise differences among input method means for younger adults. Results indicated that the Physical Qwerty was statistically perceived as the most satisfactory input method to use, followed closely by Voice, then Tracing and Onscreen Qwerty, and lastly Handwriting. The only comparisons which were not significant were between Physical Qwerty and Voice ($p = .25$) and Onscreen Qwerty and Tracing ($p = .10$).

Follow-up tests were also conducted to evaluate ten pairwise differences among input method means for older adults. Likewise, the alpha level remained unadjusted at .05. The results indicated that Voice was statistically perceived as the most satisfactory input method to use, closely followed by Physical Qwerty, then Tracing, Onscreen Qwerty, and Handwriting. The only comparisons which were not significant were between Physical Qwerty and Voice ($p = .54$), Onscreen Qwerty and Tracing ($p = .26$), and between Onscreen Qwerty and Handwriting ($p = .24$).

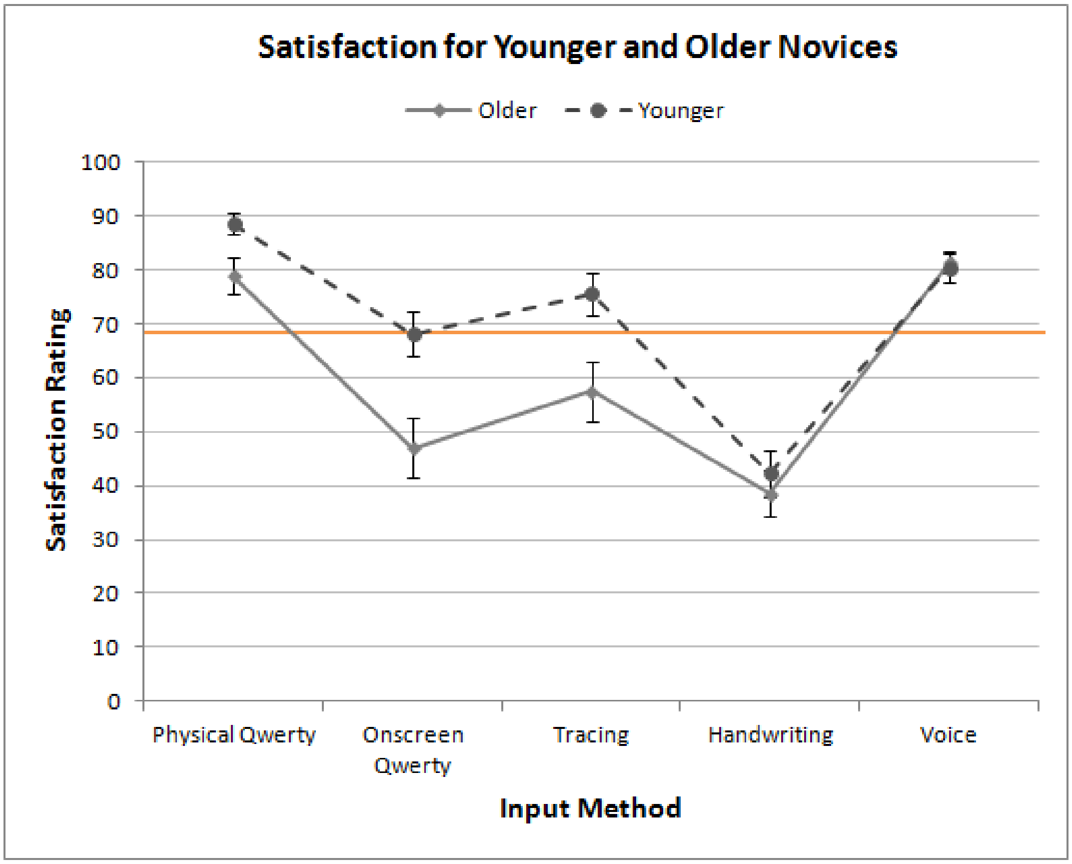


Figure 7.6. Satisfaction ratings for Younger and Older Novice Adults. The orange line reflects the average score for the SUS instrument. Bars indicate ± 1 SEM.

Workload.

Each of the six workload dimensions was examined for each age group. Fourteen outlying values for the younger adult group and thirteen outlying values for the older adult group were identified, and all were recoded (see Appendix M). The NASA-TLX dimension scales were used to obtain the data, however, participants did not complete the paired-comparisons portion of the instrument. Rather than analyze each dimension independently, the reliability of the six dimensions was tested for each input method. A composite score for workload for each input method was created by summing the scores for each dimension. Cronbach's alphas (unstandardized) for the workload dimensions for each input method for each age group are given in Table 7.1.

TABLE 7.1

RELIABILITY OF WORKLOAD COMPOSITE SCALES

	Physical Qwerty	Onscreen Qwerty	Tracing	Handwriting	Voice
Younger Adult Group	.851	.771	.891	.831	.713
Older Adult Group	.812	.837	.889	.827	.741

Note: Cronbach's alphas are for unstandardized scores.

A comparison of scores between age groups for each dimension are given in Figures 7.7 through 7.12. Overall, it appears that the younger adults rated the Handwriting method as requiring higher workload to enter text than the older adults. Additionally, Older adults tend to rate the Onscreen Qwerty higher. The older adults typically rate the input methods as requiring greater workload than the younger adults, especially on the performance dimension.

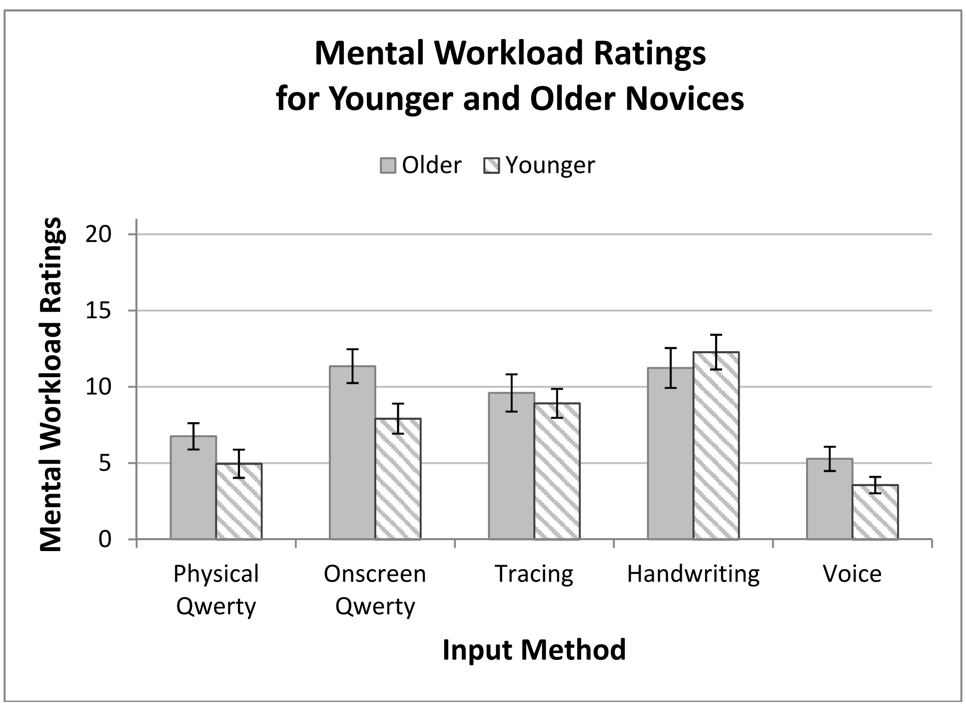


Figure 7.7. Mental workload ratings for younger and older adults.
Error bars indicate ± 1 SEM.

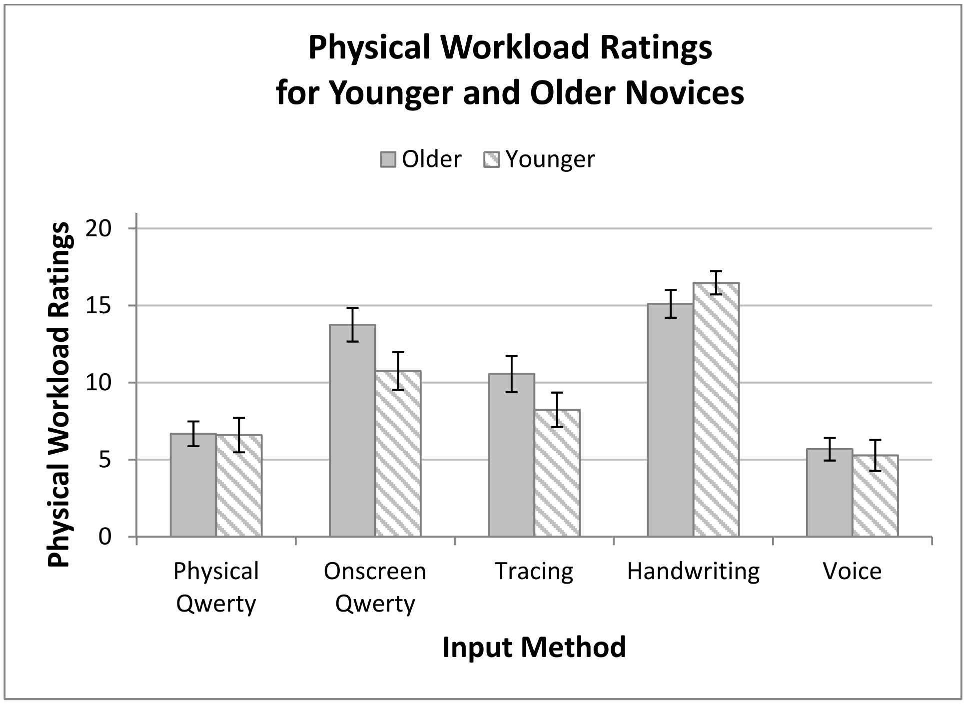


Figure 7.8. Physical workload ratings for younger and older adults. Error bars indicate ± 1 SEM

.

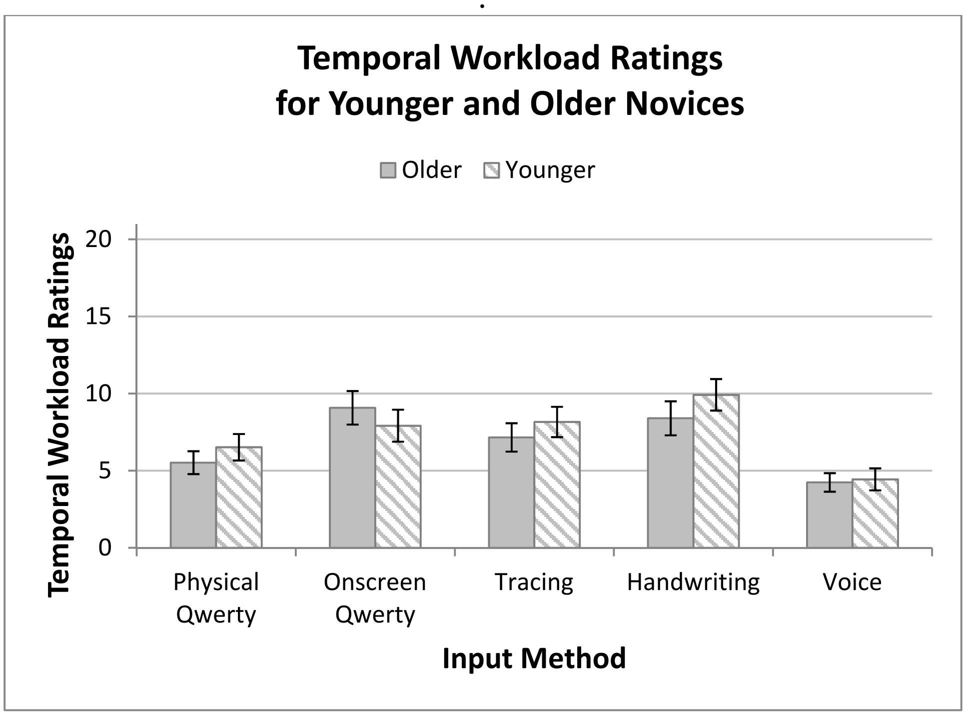


Figure 7.9. Temporal workload ratings for younger and older adults. Error bars indicate ± 1 SEM.

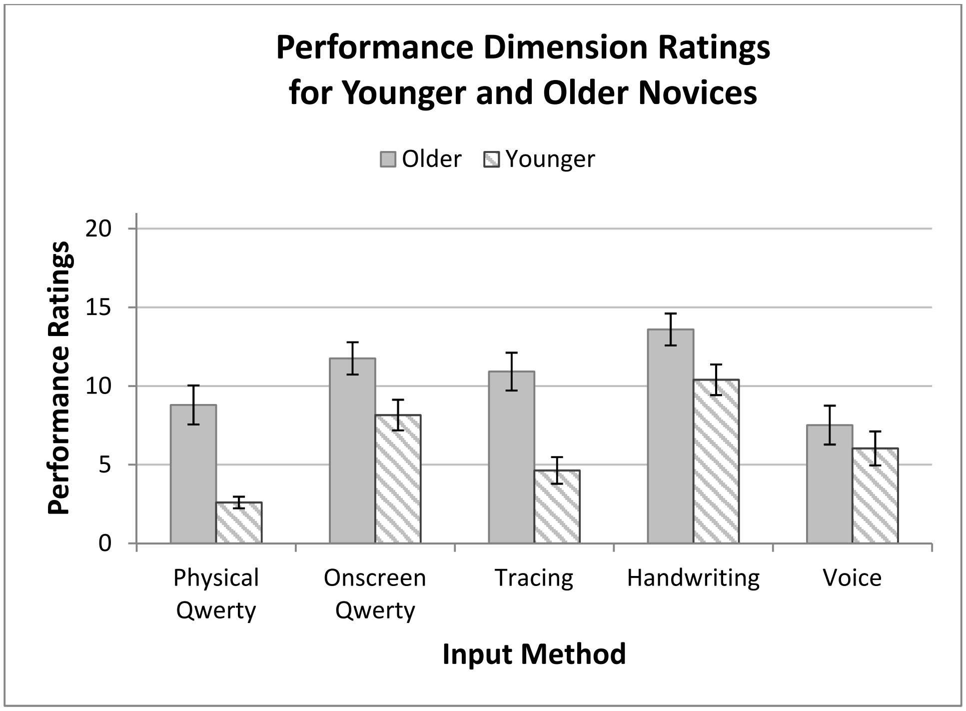


Figure 7.10. Performance ratings for younger and older adults.
Error bars indicate ± 1 SEM.

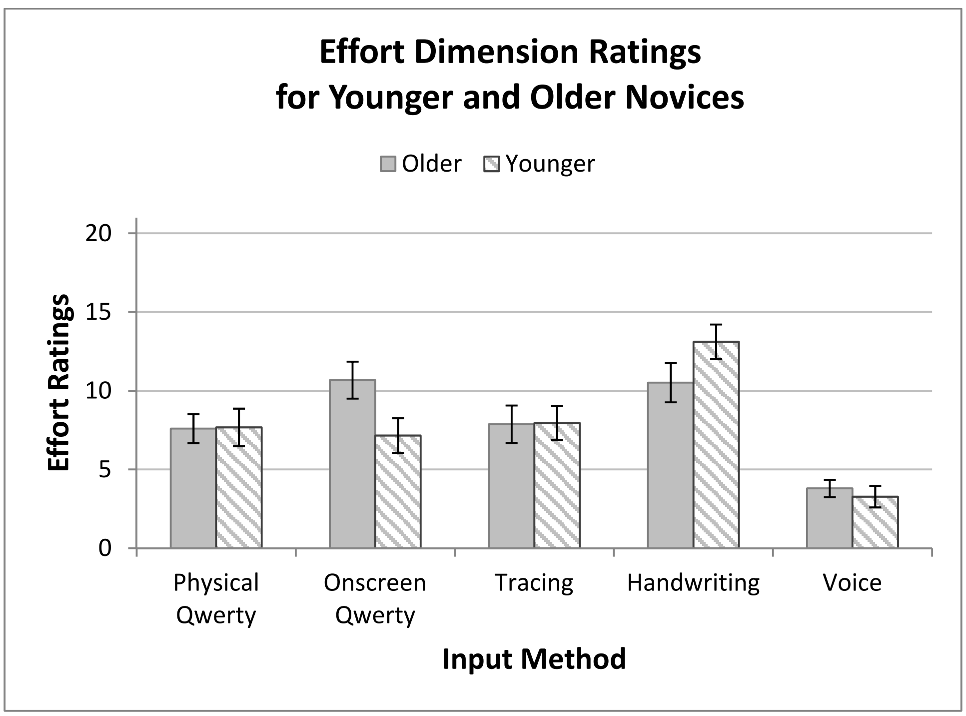


Figure 7.11. Effort ratings for younger and older adults.
Error bars indicate ± 1 SEM.

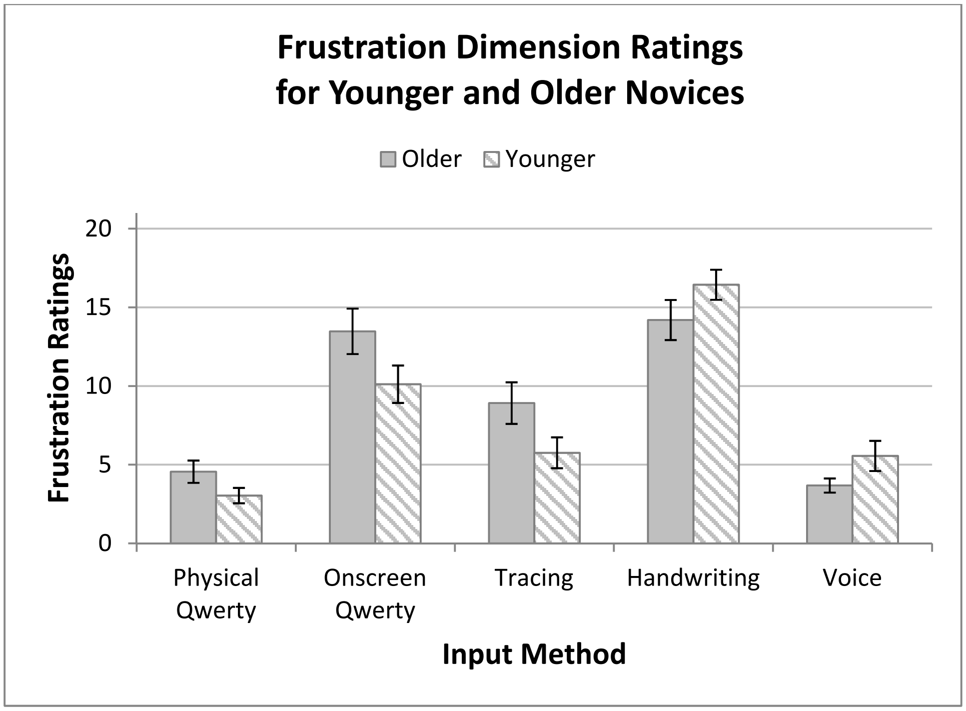


Figure 7.12. Frustration ratings for younger and older adults.
Error bars indicate ± 1 SEM.

A 2 x 5 mixed factorial analysis of variance was conducted, with age group (young versus older) as the between subjects factor and input method type (Physical and Onscreen Qwerty, Tracing, Handwriting, and Voice) as the within subjects factor. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(9) = 33.84$, p <.001, therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = .76$). The results showed significant main effect of input method for input method, $F(3.02, 145.06) = 52.15$, $p < .001$, $\eta_p^2 = .52$, a significant main effect of age group, $F(1, 48) = 4.13$, $p = .048$, $\eta_p^2 = .08$. The interaction between input method and age group was not significant, $F(3.02, 145.06) = 2.45$, $p = .07$, $\eta_p^2 = .05$.

The age main effect indicated that older adults tended to have higher workload ratings than younger adults. Follow-up analyses to the input method main effect consisted of ten pairwise comparisons among input methods. All comparisons were statistically significant. The results of this analysis indicated that the Handwriting input method was perceived as requiring the most workload for entering text. Voice was rated as requiring the least workload to use. Physical Qwerty was rated as more demanding than Voice, but less demanding than the other input methods. Tracing was rated as requiring more workload than Voice or the Physical Qwerty, but it required less than Onscreen Qwerty or Handwriting (see Figure 7.13).

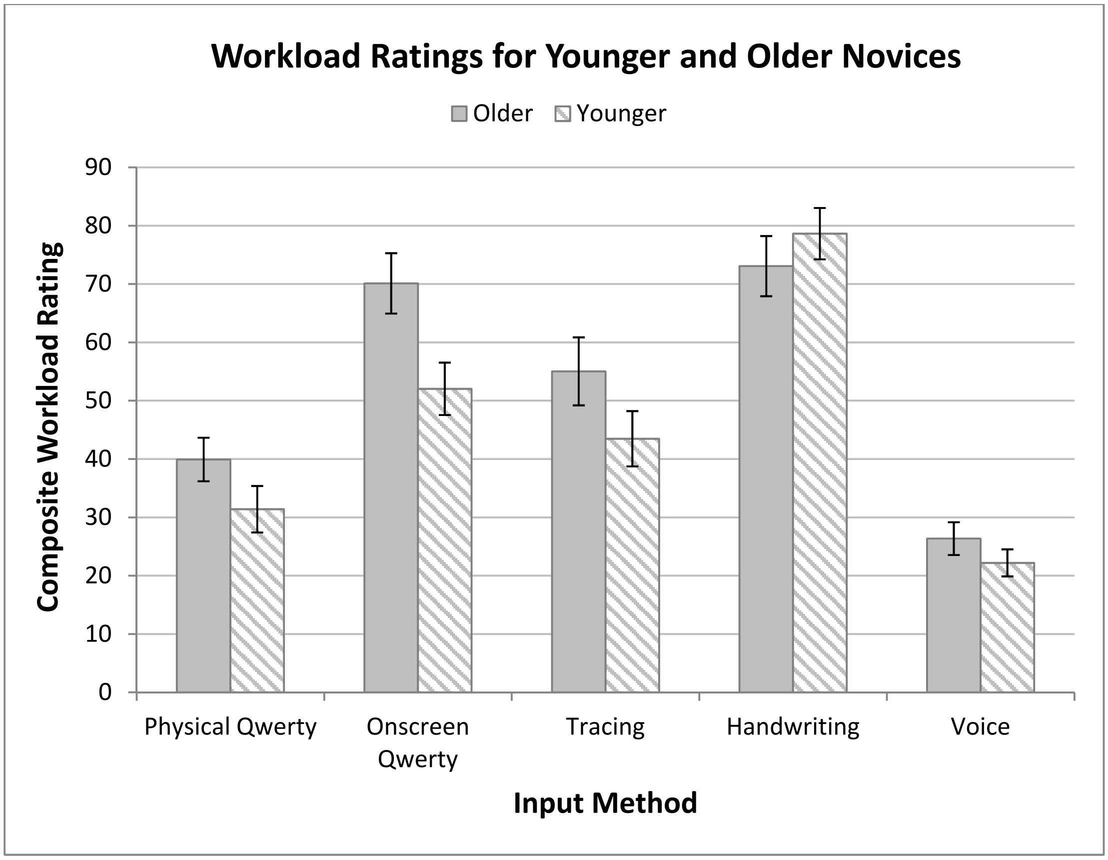


Figure 7.13. Composite workload ratings for Younger and Older Novice Adults. Error bars indicate ± 1 SEM.

Preference.

Pre-Test. It appears that the younger adults thought they would prefer both of the Qwerty methods, while the older adults perceived they would like the Physical Qwerty and Voice input methods. Both groups did not have favorable opinions of the Tracing or Handwriting input methods (see Figure 7.14).

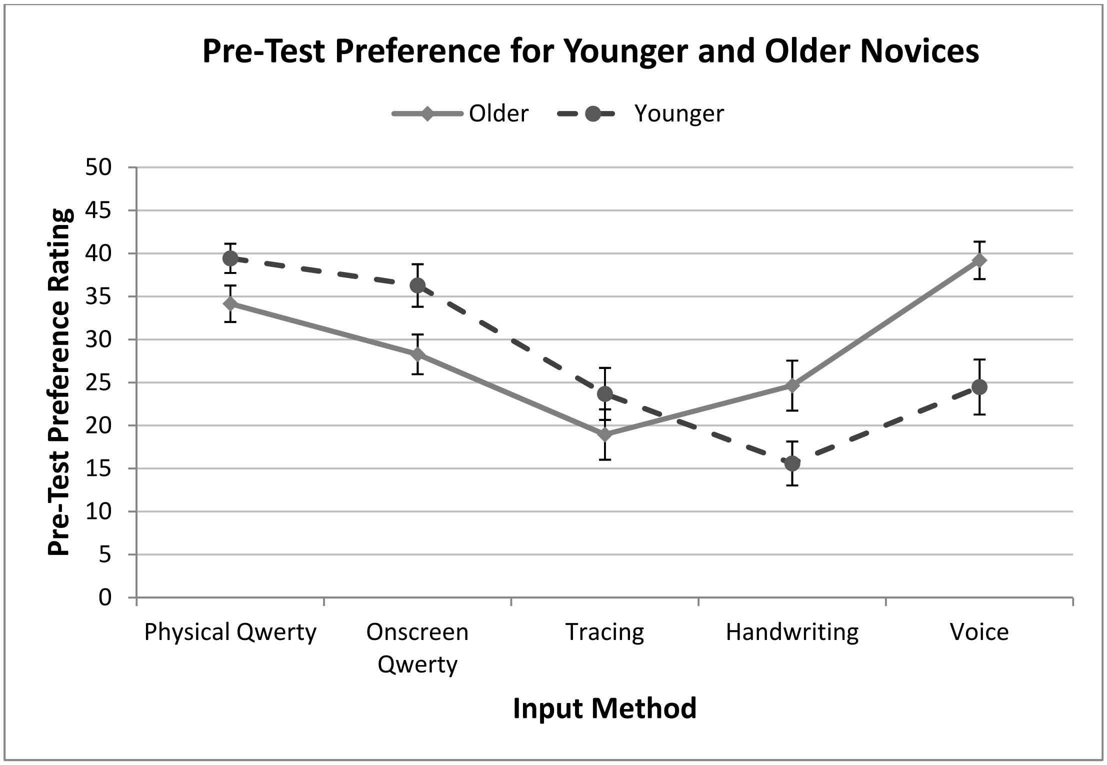


Figure 7.14. Pre-Test Preference ratings for younger and older adults.
Error bars indicate ± 1 SEM.

Post-Test. Preliminary analyses of post-test preference ratings (scale from 0 – 50, 0 = least preferred, 50 = most preferred) revealed five outlying values, which were recoded (see Appendix M). A 2 x 5 mixed factorial analysis of variance was conducted, with age group (young versus older) as the between subjects factor and input method type (Physical and Onscreen Qwerty, Tracing, Handwriting, and Voice) as the within subjects factor. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(9) = 28.87$, p =.001, therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = .84$). The results showed a significant main effect of input method for input method, $F(3.36, 161.48) = 27.34$, $p < .001$, $\eta_p^2 = .36$, and a significant interaction between post-preference ratings and age group, $F(3.36, 161.48) = 3.13$, $p = .02$, $\eta_p^2 = .06$. There was not a significant main effect of age group, $F(1, 48) = 2.70$, $p = .11$, $\eta_p^2 = .05$.

Analysis of simple main effects revealed that younger participants reported significantly higher post-preference scores for the Onscreen Qwerty input method, $F(1,228) = 6.69$, $p = .01$, than the older participants. There was no effect for age for the Physical Qwerty, $F(1,228) = 3.30$, $p = .07$, and Tracing, $F(1,228) = 1.32$, $p = .25$, Voice, $F(1,228) = 2.58$, $p = .11$, or the Handwriting input method, $F(1,228) = 1.53$, $p = .22$. Additionally, significant differences were discovered between input methods for younger adults, $F(4,192) = 16.69$, and for older adults, $F(4,192) = 13.78$, both $p < .001$, (see Figure 7.15).

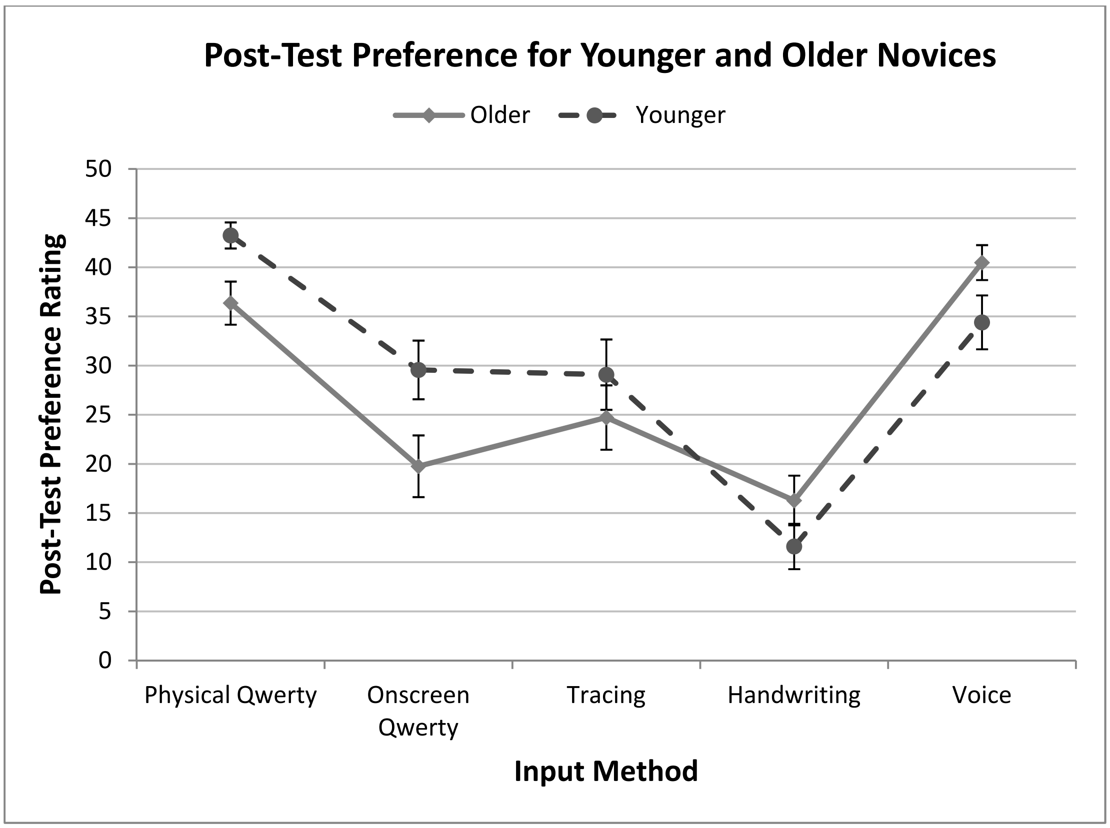


Figure 7.15. Post-Test Preference ratings for younger and older adults. Bars indicate ± 1 SEM.

Follow-up analyses to the input method main effect consisted of ten pairwise comparisons among input methods, ignoring the effect of age. The results of this analysis indicated that the Physical Qwerty and Voice input methods were significantly more preferred over the other input methods post-test, but neither was statistically preferred more than the other, (*p* = .55). Tracing and Onscreen Qwerty obtained similar preference ratings (*p* = .32). Handwriting was significantly least preferred.

Pre- to Post-Test Rating Change. It appears that younger adults had the largest positive change in preference ratings for the Voice input method, and had the largest negative deviation for Onscreen Qwerty. Their ratings remained fairly consistent from pre- to post-test for Physical Qwerty and Handwriting. The largest positive change in ratings for older adults was for Tracing, while Onscreen Qwerty and Handwriting were both seen more negatively (see Figure 7.16).

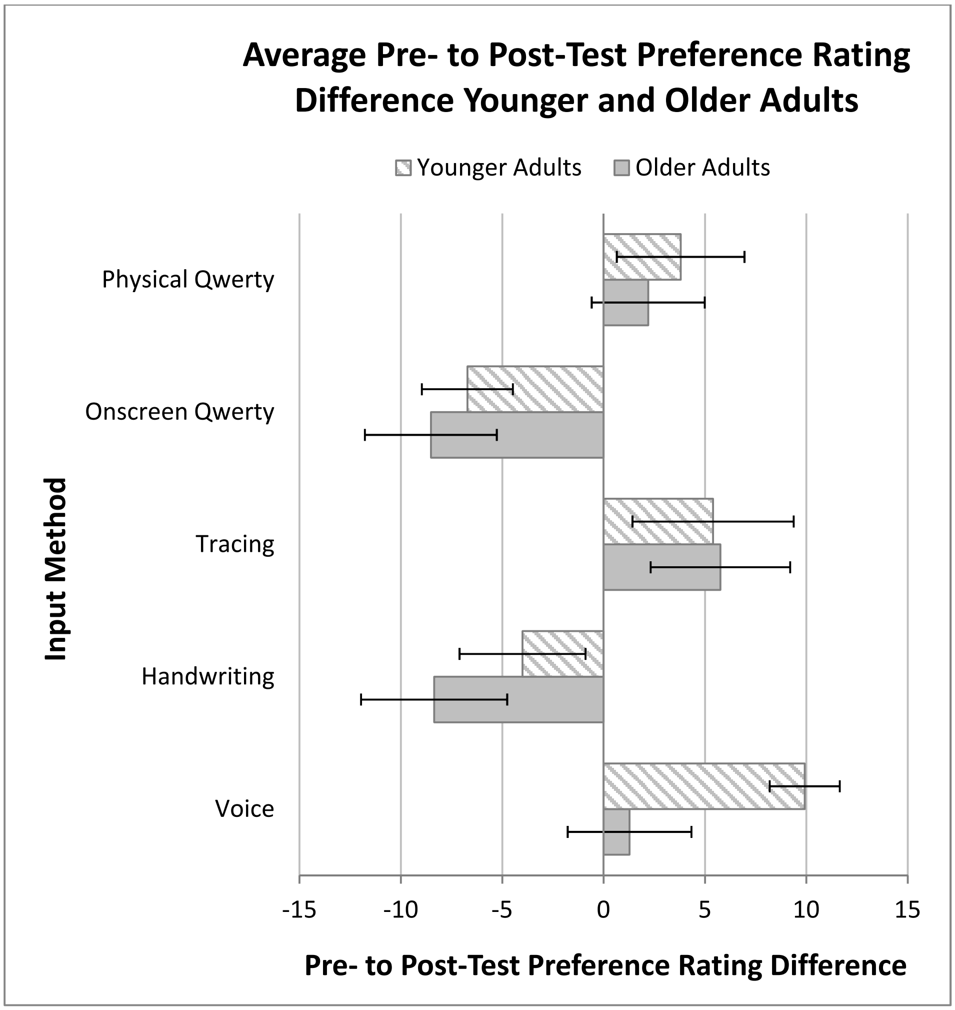


Figure 7.16. Pre- to Post-Test Preference change for younger and older adults.
Error bars are ± 1 SEM.

Individual Assessments.

A series of independent *t* tests was conducted to assess whether significant differences existed on the obtained measures between the younger and older participant samples. Means, standard deviations, and *t* statistics are given in Tables 7.2 and 7.3. It appears that the younger and older adult samples were statistically different regarding the scores on the Digit Symbol Substitution test, Attitudes Towards Computers Questionnaire, speed on the Grooved Pegboard task, Keyboarding entry rate, thumb and index circumferences, index width, speech rate, frequency tremor intensity, and amplitude perturbation.

TABLE 7.2

BETWEEN-GROUPS PHYSICAL, COGNITIVE AND ATTITUDE ASSESSMENT MEASURES

	Younger *M* (*SD*)	Older *M* (*SD*)	*t* statistic
Digit Symbol Substitution (# correct)	89.16 (17.97)	67.12 (12.59)	**5.02*** **
Attitudes Towards Computers	89.16 (6.99)	77.72 (9.59)	**4.72*** **
Finger Tapping Test (average taps)	56.24 (7.34)	54.25 (7.59)	.94
Jebsen-Taylor (dominant, in sec)	16.56 (.46)	13.46 (.25)	2.37
Grooved Pegboard (dominant)	85.33 (6.9)	100.49 (16.42)	**4.25*** **
Keyboarding (WPM)	54.04 (17.31)	36.94 (18.50)	**3.37** **
Hand Length (mm)	179.3 (10.57)	177.9 (10.17)	.47
Hand Breadth (mm)	80.75 (6.70)	78.97 (8.92)	.80
Thumb Length (mm)	61.09 (5.49)	60.01 (4.39)	.77
Thumb Width (mm)	19.65 (1.87)	20.24 (2.58)	.91
Thumb Circumference (mm)	63.2 (5.60)	66.74 (5.88)	**2.17***
Index Length (mm)	71.15 (5.26)	68.54 (5.46)	1.72
Index Width (mm)	15.37 (1.44)	16.31 (1.57)	**2.20***
Index Circumference (mm)	51.14 (4.68)	56 (4.77)	**3.64** **
Thumb Reach (mm)	57.47 (4.67)	58.22 (4.99)	.55
Functional Range of Motion (deg)	120.96 (16.77)	119.16 (20.46)	.34

Note: * indicates statistical significance $p < .05$, ** indicates $p < .01$, and *** indicates $p<.001$, all two-tailed. Also note that listed values for the Jebsen-Taylor were back-transformed.

TABLE 7.3

BETWEEN-GROUPS VOCAL ASSESSMENT MEASURES

	Younger *M* (*SD*)	Older *M* (*SD*)	*t* statistic
Speech Rate (WPS)	2.93 (.39)	2.71 (.37)	**2.07***
Average Fundamental Frequency (Hz)	179.82 (56.93)	154.93 (35.28)	1.86
Frequency Tremor Intensity Index (%)	.29 (1.88)	.64 (2.47)	**3.69****
Amplitude Tremor Intensity Index (%)	4.89 (1.75)	4.53 (1.76)	.42
Pitch Perturbation Quotient (%)	.51 (.04)	.61 (.07)	1.35
Amplitude Perturbation Quotient (%)	.99 (.06)	.94 (.18)	**2.92****

Note: * indicates statistical significance $p < .05$, ** indicates $p < .01$, two-tailed.
Also note that listed values for the Frequency and Amplitude Tremor Intensity, Pitch and Amplitude Perturbation Quotients were all back-transformed.

Discussion

The results from the comparison between younger and older adult novices revealed significant differences between the text input methods on several dependent measures. Most of the observed effects were largely due to the differences between input methods, however, there were a few surprising outcomes between age groups.

Performance.

***Adjusted Words per Minute*.** Results from the mixed factorial suggest that age and input method type impact entry rate. All input methods within each group were found to have significantly different entry rates from each other, the Voice method had the quickest entry rate, followed by Physical Qwerty, Tracing, Onscreen Qwerty, and lastly, Handwriting. These results mimic those found in both studies. In general, older participants tended to be slower entering text compared to the younger group. Due to the many limitations associated with aging and the pervasive literature demonstrating the relationship between aging and slower performance, this was expected. Slower performance by older adults can be caused by the global slowing of neural processes (Salthouse, 1996). Additionally, older adults are error adverse, which means that they tend to favor responding accurately over responding quickly even if it takes longer (Starns & Ratcliff, 2010). However, possibly the most interesting result was the absence of a significant relationship for the Voice input method between age groups. This suggests that younger and older novice users performed similarly when using the Voice input method. This is interesting because previous research on Automatic Speech Recognition systems has favored performance by younger adults (Wilpon & Jacobsen, 1996; Vipperla et al., 2008). Additionally, the two age groups differed on several of the baseline measures. Older participants had a significantly slower speech rate, and increased proportions of frequency tremor and shimmer. This implies that these characteristics are not crucial to the success of Voice output. One explanation for similar entry rates between age groups is that the Voice input method relied very little on manual interaction with the interface. The older adults were less dexterous, which most likely played a role in the discrepancies between age groups for the manual input methods. The Voice input method was mostly hands-free, except during text correction. Even then, it was fairly fast to correct a phrase by pressing and holding the "delete" key to begin again. The pace with which the participants interacted with the interface seems to govern entry rates with this method more than physical or (non-pathological) voice limitations.

Error Rates. Among manual input methods at the character-level, results indicated that Handwriting had the highest error rates, followed by Onscreen Qwerty, Tracing and Physical Qwerty, with the lowest error rates. This trend was also seen in both studies. Differences were not found between age groups for Corrected Error Rate, which indicates that older adults deleted a similar number of characters in their input strings as the younger adults. This was not too surprising, as similar levels of error correction have been observed between age groups in previous research (Rabbitt, 1979). The differences in error rates between groups were seen in the Uncorrected Error Rates, where older adults tended to leave significantly more character errors in their final transcriptions than younger adults. This was surprising, since older adults are more persistent and tend to be more careful than younger adults (Horn, 1989). However, given older adults were slower on the manual input methods, this may in part reflect their persistence to correct mistakes. The mistakes left in the final transcription may have been seen as acceptable, or perhaps the procedure for correcting mistakes was too laborious. At the character-level, for both groups, Handwriting and Onscreen Qwerty had significantly higher Uncorrected Error Rates than Tracing and Physical Qwerty. Probably the most interesting finding was that older and younger adults had similar levels of Uncorrected Error Rates with the Physical Qwerty method, which also happened to be the input method that afforded the least amount of uncorrected errors. This may indicate that the Physical Qwerty method allowed participants to easily make corrections, which may have increased their willingness to do so.

In comparing the results from both studies, the remaining word-level errors within the final transcriptions did tend to be greater for older adults than for younger adults, but only for Handwriting, Onscreen Qwerty, and Tracing. This is interesting because all of these methods are the touch screen manual input modes, and this finding suggests that there is something inherent about the touch screen technology which is more problematic for older adults. It could be that these methods are more complicated or laborious for older adults, and that they were allowed transcribed phrases to contain many more word-level errors using these methods, compared to younger adults. Perhaps even more exciting is that there were no significant differences in Word

Error Rates between younger and older participants with either the Physical Qwerty or the Voice input methods. This suggests that older participants either created roughly the same amount of word-level errors as younger participants with these methods, or that they were more persistent at making corrections to bring down the error rate. It is important to note that word errors are not necessarily equal. For instance, for the purposes of this study a word that contains one substituted character and a word that is completely wrong are both counted as one word error. However, when looking at the Uncorrected Error Rates, one can see that the character-level errors carry the same trend as the word-level errors, including the observation that uncorrected errors with Physical Qwerty are roughly the same for both age groups, and it was observed that the output for Voice was fairly consistent between age groups.

Satisfaction. Comparing satisfaction scores between groups revealed that younger adults tended to be more satisfied with the Onscreen Qwerty and Tracing input methods than older adults. This may be due to the touch screen keypad. Older adults were not as proficient in using the desktop keyboard, and may not have been as familiar with the position of the keys as the younger adults. This may have introduced a level of complexity which made these input methods less appealing. Also, the older age group had more limited dexterity and wider index fingers, which seemed to have large effects on satisfaction scores for Tracing and Handwriting with the older group. Both age groups rated their satisfaction with Voice and Physical Qwerty and dissatisfaction with Handwriting similarly, which was somewhat surprising since Holzinger and colleagues (2007) reported that older adults usually report lower satisfaction levels than younger adults. However, similar ratings for Physical Qwerty, Voice, and Handwriting may reflect levels of satisfaction similar to that of younger adults because these build more upon activities that older adults do on a regular basis (using a tactile keyboard, speaking, and writing). Typing and tracing on a touchscreen may make the use of computer technology more “visible”, which has been shown to be less well received in past research (Coleman et al., 2010).

Workload. Older adults tended to report greater workload for each of the input methods, except for Handwriting, which tended to be rated as requiring more workload with the younger adult group than the older adult group (though this difference was not significant). This was not surprising, as older adults tend to report higher levels of workload for the same task, compared to younger adults (Fox et al., 1995). Handwriting was perceived by both groups as requiring the greatest workload to enter text, then Onscreen Qwerty, then Tracing, then Physical Qwerty, and finally Voice, which was rated as requiring the least workload to operate. These trends were seen in both studies.

When examining each dimension, older adults tended to rate Onscreen Qwerty as more mentally and physically demanding, requiring more effort and frustrating than the younger participants. Indeed, older participants frequently commented that the Onscreen Qwerty method took constant vigilance, that they were often distracted by the pop-up letters on the keyboard and that they were confused about where to press the keys or how much force to use to increase accuracy. They also reported that they felt their fingers would get fatigued due to the way they held their finger. This was reflected in comments for the Tracing input method as well. They tended to feel higher physical workload than younger participants, and they reported that this was because they had to drag their finger across the screen. The older participants were also frustrated that their hands and fingers covered the screen while they were tracing, and confused at times when a trace did not retrieve the correct word. It is interesting that younger participants tended to rate Handwriting as requiring more effort and more frustrating than older participants. Since error rates were lower and entry rates were higher for Handwriting for the younger group, the high ratings may be explained by expectations. For example, perhaps the younger adults expected the Handwriting method to be more successful, and were more disappointed with the technology than older adults. Younger and older adults seemed to feel roughly the same amount of temporal workload ("hurried" or "rushed") across all of the input methods. Older adults' performance ratings were worse than those for younger adults, except for the Voice input method. This trend is also seen in the error rate measures. However, it is surprising that they rated their performance as much worse than younger adults for Physical Qwerty, because while they had significantly slower entry rates for this input method, they did not have much difference in error rates. This may indicate that they were more aware of how long it took to enter text, or the number of mistakes that were made while entering text, which may have altered their opinion on how successfully they were able to use it.

Preference. A comparison between groups for Pre-Test Preference ratings shows that the younger adults favored the Qwerty methods, while older adults favored Voice. The younger adults anticipated that the Qwerty methods would be most familiar to them, and they questioned the reliability of the technology for the other three input methods. The older adults, some of whom were not proficient at keyboarding, tended to prefer Voice as it seemed natural and easy. Both groups felt that Tracing was unfamiliar, and that it would take a lot time to learn. Both groups were concerned about their personal handwriting style and thought that Handwriting would also be too slow.

Post-test preference ratings indicated that both groups similarly rated Physical Qwerty and Voice as the more preferred input methods. This is supported by the similar performance, satisfaction and workload rating differences between age groups for these methods. Younger participants rated the Onscreen Qwerty method as significantly more preferred than the older adult group, and indeed they were more successful and faster with this method than older adults. The Tracing method and Handwriting method were rated similarly by both groups as moderately preferred and least preferred, respectively.

The comparative differences between Pre- and Post-Test Preference for both age groups indicate that both groups felt less favorable about Onscreen Qwerty and Handwriting after using these input methods, but more favorable, or nearly the same, about Physical Qwerty, Tracing and Voice. The biggest change is observed in the more positive view of Voice for younger adults, though there were also large decreases in preference for Onscreen Qwerty and Handwriting for the older participants. This may indicate that expectations of Physical Qwerty, Tracing and Voice were either met or exceeded, and that Onscreen Qwerty and Handwriting performed worse than anticipated.

CHAPTER 8

CONCLUSIONS AND IMPLICATIONS

Study 1 and Study 2 were conducted to evaluate the relationships between five current smartphone text input methods and performance, perceived satisfaction, workload and preference. Additionally, the roles of individual anthropometry, hand function and speech qualities were examined with regard to their relationships with the dependent measures. Study 1 examined these relationships with a group of young adult novices, whereas Study 2 evaluated the same relationships with a group of older adult novices.

Interpretation of Results

The results from Study 1 and Study 2 show that the Voice input method resulted in the fastest entry rate for participants in both age groups, and that the Handwriting input method was the slowest. This is not surprising since the natural rate of speech is much higher than unconstrained handwriting (Cox et al., 2008; Bailey, 1996). However, the differences among Physical and Onscreen Qwerty and Tracing were surprising. For example, since Tracing is a relatively new way to input text, it was not expected to outperform Onscreen Qwerty, which is typically slower and placed the lowest on subjective ratings of these three input methods. It also appears that the lower error rates were obtained for the Physical Qwerty for younger adults, and Voice for older adults. This is probably due to lower error rates during the creation of the input stream, however, it may also suggest less effort or frustration in making corrections, which would increase the likelihood of a more accurate transcribed phrase. Some of the findings contradicted entry rate and error rates found by Castellucci and MacKenzie (2011) and Arif and Stuerzlinger (2009), however, different devices, software, stimuli and sample populations were used. Also, different protocols were in place and different metrics were used to draw inferences.

It was interesting to explore differences between the younger and older age groups, and to find no significant differences between them for the Voice method for entry and error rates, satisfaction, workload, and post-test preference ratings. There were also no differences between groups for Physical Qwerty on uncorrected error rate, word-level error rate, satisfaction, workload, and post-test preference. This implies that, even though the older participants differed to the younger group with regard to dexterity, attitudes, keyboarding entry rate, cognitive processing speed, digit circumference, speech rate, frequency tremor, and shimmer, they were able to achieve a level of performance and perceived opinions about these input methods similar to the younger adult group.

Implications

The results of the these two studies imply that the type of input method used has implications for the user's level of performance, and shapes the user's perceptions of satisfaction, how hard they have to work to achieve their level of performance, and their preference of input method. Additionally, personal characteristics can also impact the user's experience with several of the input methods.

Overall, it appears that novice younger adults have the most positive results with the Physical Qwerty input method, based on performance, subjective ratings, and comments. Participants felt that this method was comfortable, familiar, accurate and easy to control. Though some participants mentioned that the keyboard was too large, there were no significant correlations between hand sizes and error rates or satisfaction for the Physical Qwerty input method. This is interesting because most of the smartphones currently on the market do not have a physical keyboard. In fact, the major manufacturer HTC announced mid-2012 that they would discontinue physical keyboards in their smartphone portfolio (Oryl, 2012). Indeed, including a physical keyboard on a smartphone device increases the cost of production, weight and depth of the form factor. However, phasing out the physical keyboard may not bode well for consumers.

Older novice users had the most positive results with the Voice input method, based on performance, subjective ratings and comments. They reported that Voice was fast, easy, and accurate for entering text. In fact, performance differences between younger and older adults did not differ within the Voice condition, which implies that performance remains relatively constant across generations. Both age groups reported that they would not use Voice in every situation, due to privacy concerns and not wanting to disturb others. For instances where they would prefer to use a manual input method, Physical Qwerty was also a viable option for older adults. Older adults obtained levels of performance that did not differ or were very close on all performance measures, and similar subjective ratings, to the younger age group within the Physical Qwerty condition. However, again it is disconcerting to report that physical keyboards are not as readily available as they once were.

In context of the touch screen input methods, it seems that both groups did slightly better with the Tracing input method over Onscreen Qwerty. Handwriting performed poorly. Tracing seems to be a promising feature that is fairly quick and easy to learn. This might be appealing to users who desire to have a mode of manual input on a slate smartphone. However, as many participants mentioned, the input method may not be very discoverable or intuitive, unless someone shows them what it is and how to operate it. Though this is something that could easily accomplished at the point of purchase. The facilitator provided a quick 60 second tutorial and most participants were able to effectively trace on their first phrase.

A caveat to interpreting these results is that there are most likely situations where the recommended input methods are not the most optimal. The user should have the option to choose from several input methods to accomplish a variety of tasks. It is limiting the user to provide only one form of input. Providing multiple input methods, allowing the smartphone to be multimodal, for text input would be the best course of action.

Limitations

While it appears there are clear recommendations based on the results from Study 1 and Study 2, the task of entering short 4 – 8 word phrases is not representative of all text entry situations these groups are likely to face on a smartphone device. For example, composing emails, writing to-do lists, filling out forms on a website may be more efficient with different input modes. The amount of exposure to these input methods can also be considered a limitation. The time taken to enter twenty phrases with each input method ranged from roughly 10 minutes to 45 minutes. As these input methods were novel to all participants, the attained level of performance and opinions about these input methods formulated in such a short amount of time may not hold across prolonged exposure.

The experimental setting may have also been a limiting factor. People enter text in a variety of environments, and sitting in a quiet room next to a facilitator is not typically one of those. Results may not generalize to multitasking situations, and voice results may not generalize to noisier environments. Also, the experimental setting may have changed the participants' behavior. On occasion, participants indicated that they had formed an opinion of a certain input method after the second or third phrase, and would never choose it for personal use.

Additionally, the studies examined only one type of physical keyboard and one suite of touch screen operated input methods. The Physical Qwerty results may not generalize to the "candybar" form factor (portrait), or physical keyboards which do not have as much tactile feedback (i.e., "clickiness" or crowning of the keys). Other features may be enabled with onscreen methods, such as haptic or audio feedback, which may alter performance and perceptions of the device. Additionally, the introduction of other equipment, such as a stylus, to input text with any of the onscreen input methods may alter performance and perceptions.

Another limitation was the Text Entry Metrics for Android (TEMA) software, which was helpful but did not include a “back” key. At times participants would enter a phrase and express that they wished to go back and make corrections, though they could not. Though this was not a frequent problem, it is possible that this influenced perceptions of the input methods where the “delete” and “enter” keys were close together and more easily confused. Likewise, the navigation buttons on the smartphone itself were brushed up against at times, causing the participant to exit out of TEMA, or for additional menu items to pop up on screen. These actions usually caused confusion for the older participants, and while they were told not to take these actions into account when rating the input methods, there is no way to know whether it was a factor.

Future Research

Future studies should focus on examining these input methods in a more naturalistic environment. For example, it would be beneficial to assess how well the input methods perform in a library or office, on a public transit system, or while a person is in a noisy social environment. Additionally, it would be interesting to evaluate the performance of the input methods while the person is doing some kind of activity, such as walking.

It would also be important for future studies to examine the performance and perception of input methods with different task types. For instance, investigating which input methods better lend themselves to email composition or to filling out forms online may be useful.

Future studies should also examine whether the differences between input methods also exist for keypads with haptic and audio feedback features enabled, and after introducing equipment that may help users interact with the interface, such as a stylus. Additionally, it is important to explore the impact of different input methods with “candybar” style devices, or for smartphones held in portrait mode. Also, as tablets and eReaders become more prolific, it would be interesting to consider whether these findings generalize to larger devices.

Expertise is likely to alter perceptions of the input methods, as well as performance. It would be beneficial to evaluate expert users of each method. A future longitudinal study could assess the amount of time needed to become proficient with each method and how user perceptions of the input method change over time.

Future studies could also explore a wider demographic, including people of different ethnic backgrounds and those with different physical limitations. It was mentioned by several participants that, though they did not particularly like the handwriting input method, that it may be useful for entering characters in another language (i.e., Spanish or German) without having to switch to another keypad. Additionally, the results from Study 1 and Study 2 use samples of participants with no known pathology. It would be useful to examine the accessibility of these input methods among different populations, such as those with speech or movement disorders, or those with visual impairments.

Older participants preferred the Voice input method, but many of the participants initially had a poor mental model of the software and noted that the feature was not very discoverable on the device. As people age, voice recognition may be a viable method for more than text entry of simple phrases. A future study should explore the adoptability of voice recognition by older participants and its impact on quality of life.

REFERENCES

REFERENCES

Administration on Aging. (2010). A profile of older Americans: 2010. Retrieved October 25, 2011, http://www.aoa.gov/aoaroot/aging_statistics/Profile/2010/docs/2010profile.pdf

Adomo, D. E., Martin, B. J., & Brown, S. H. (2007). Age-related differences in upper limb proprioceptive acuity. *Perceptual and Motor Skills, 104*, 1297-1309.

Agnew, P. J., & Maas, F. (1982). Hand function related to age and sex. *Archives of Physical Medicine and Rehabilitation*, *63*(6), 269-271.

Apple. (2007, January 9). Apple reinvents the phone with the iPhone [Press release]. Retrieved March 25, 2013, from http://www.apple.com/pr/library/2007/01/09Apple-Reinvents-the-Phone-with-iPhone.html

Arif, A. S., Lopez, M. H., & Stuerzlinger, W. (2010, June). *Two new mobile touchscreen text entry techniques.* Poster session presented at the annual conference of Graphics Interface, Ottawa, Ontario, Canada.

Arif, A. S., and Stuerzlinger, W. (2009). Analysis of text entry performance metrics. *IEEE Symposium on Human Factors and Ergonomics,* 100-106. IEEE Press, New York.

Aronson, A. E., & Bless, D. M. (2009). *Clinical voice disorders* (4th ed.). New York: Thieme Medical Publishers, Inc.

Austin, D., Jimison, H., Hayes, T., Mattek, N., Kaye, J., & Pavel, M. (2011). Measuring motor speed through typing: A surrogate for the Finger Tapping Test. *Behavior Research methods*, *43*(4), 903-909.

Bailey, R. W. (1996). *Human performance engineering: Designing high quality, professional user interfaces for computer products, applications, and systems* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

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

Barrett, L. L. (2011, January). Health and caregiving among the 50+: Ownership, use and interest in mobile technology. *AARP*. Retrieved October 12, 2011, from http://www.aarp.org/technology/innovations/info-01-2011/health-caregiving-mobile-technology.html

Becker, S. A., & Webbe, F. M. (2006). Designing for older adults users of handheld technology. *Proceedings of the 28th IEEE Engineering in Medicine and Biology Society Annual Conference*, 3297-3300.

Bellsouth, IBM unveil personal communicator phone. (1993, November 8). *Mobile Phone News*. Retrieved March 20, 2013, from http://research.microsoft.com/en-us/um/people/bibuxton/buxtoncollection/a/pdf/press%20release%201993.pdf

Benzeghiba, M., De Mori, R., Deroo, O., Dupont, S., Erbes, T., Jouvet, D., Fissore, L., Laface, P., Mertins, A., Ris, C., Rose, R., Tyagi, V., and Wellekens, C. (2007). Automatic speech recognition and speech variability: A review. *Speech Communication, 49,* 763-786.

[Blackberry 5810 image]. (n.d.). [Photograph]. Retrieved March 25, 2013, from http://dienthoaiblackberry.vn/uploads/images/sanpham/3.jpg

Boatright, J. R., Kiebzak, G. M., O'Neil, D. M., & Peindl, R. D. (1997). Measurement of thumb abduction strength: Normative data and a comparison with grip strength and pinch strength. *The Journal of Hand Surgery, 22*(5), 843-848.

Brooke, J. (1996). SUS: A quick and dirty usability scale. In Jordan, P. W., Thomas, B., Weerdmeester, B. A., and McClelland, I. L. (Eds). *Usability Evaluation in Industry*, 189-194. London: Taylor and Francis.

Brown, T. S., Brown, J. T., & Baack, S. A. (1990). Attitudes of the elderly toward computers. *The Journal of Computer Information Systems, 31,* 28-33.

Byers, J. C., Bittner, A. C., Jr., & Hill, S. C. (1989). Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? In Anil Mital (Ed.), *Advances in industrial ergonomics and safety I* (pp. 481-485). London: Taylor & Francis.

Carmeli, E., Patish, H., & Coleman, R. (2003). The aging hand. *Journal of Gerontology: Medical Sciences, 58A*(2), 146-152.

Castellucci, S. J. & MacKenzie, I. S. (2011). Gathering text entry metrics on Android devices. *Proceedings of the ACM Symposium on Human Factors in Computing Systems (CHI),* Vancouver, BC, Canada, 1507-1512.

Chapuis, O., & Dragicevic, P. (2011). Effects of motor scale, visual scale and quantization on small target acquisition difficulty. *In ACM Transactions on Computer-Human Interaction (CHI), 18*(3), 1-32.

Chaput, S., & Proteau, L. (1996a). Aging and motor control. *The Journals of Gerontology, Series B: Psychological Sciences & Social Sciences, 51*(6), 346-355.

Chaput, S., & Proteau, L. (1996b). Modifications with aging in the role played by vision and proprioception for movement control. *Experimental Aging Research, 22*(1), 1-21.

Charness, N. & Holley, P. (2004). The new media and older adults: Usable and useful? *American Behavioral Scientist, 48*(4), 416-433.

Clarkson, E., Clawson, J., Lyons, K., & Starner, T. (2005). An empirical study of typing rates on mini-QWERTY keyboards. In *Proceedings of the ACM Symposium Computer-Human Interaction, Portland, Oregon*, 573-582.

CNET. (2002, January 30). Palm Treo 180 review [Product review]. Retrieved March 25, 2013, from http://reviews.cnet.com/smartphones/palm-treo-180/4505-6452_7-7629257.html

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillside, NJ: Lawrence Erlbaum Associates, Inc.

Cole, K. J., Rotella, D. L., & Harper, J. G. (1998). Tactile impairments cannot explain the effect of age on a grasp and lift task. *Experimental Brain Research, 121*(3), 263-269.

Coleman, G. W., Gibson, L., Hanson, V. L., Bobrowicz, A., & McKay, A. (2010). Engaging the disengaged: How do we design technology for digitally excluded older adults? In *Proceedings of the Conference on Designing Interactive Systems, Aarhus, Denmark*, 175-178.

Computer Business Review. (1996, March 27). Nokia 9000 communicator sets the standard for intelligent cellular telephones. Retrieved March 25, 2013, from http://www.cbronline.com/news/nokia_9000_communicator_sets_the_standard_for_intelligent_cellular_telephones

comScore. (2013, February 22). *Mobile future in focus 2013*. Retrieved March 30, 2013, from http://www.comscore.com/Insights/Presentations_and_Whitepapers/2013/2013_Mobile_Future_in_Focus

Contreras-Vidal, J. L., Teulings, H. L., & Stelmach, G. E. (1998). Elderly subjects are impaired in spatial coordination in fine motor control. *Acta Psychologica (Amsterdam), 100*(1-2), 25-35.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology, 1*(1), 42-45.

Cox, A. L., Cairns, P. A., Walton, A. & Lee, S. (2008). Tlk or txt? Using voice input for SMS composition. *Personal and Ubiquitous Computing, 12*(8), 567-588.

CTIA. (2011). CTIA semi-annual wireless industry top-line survey results. Retrieved October 20, 2011, from http://files.ctia.org/pdf/CTIA_Survey_MY_2011_Graphics.pdf

Czaja, S. J., & Sharit, J. (1993). Age differences in the performance of computer-based work. *Psychology and Aging, 8*, 59-67.

Domey, R. G., McFarland, R. A., & Chadwick, E. (1960). Dark adaptation as a function of age and time. *Journal of Gerontology, 15*, 267-279.

Duncan, J. & Ferguson, D. (1974). Keyboard operating posture and symptoms in operating. *Ergonomics, 17*, 651-662.

Edison Research. (2012, June). The smartphone consumer 2012. Retrieved March 20, 2013 from http://www.edisonresearch.com/wp-content/uploads/2012/06/The_Smartphone_Consumer_2012_by_Edison_Research.pdf

Fidell, L. S., & Tabachnick, B. G. (2003). Preparatory data analysis. In J. A. Schinka (Series Ed.) and W. F. Velicer (Vol. Ed.), *Handbook of psychology: Research methods in psychology* (Vol. 2, pp. 115-141). New York: Wiley.

Field, A. (2009). *Discovering statistics using SPSS.* London: Sage Publications.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-Mental State": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 189-198.

Fox, S. L., Eggemeier, F. T., & Biers, D. W. (1995). Effects of aging on working memory and workload. In *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting*, 139-142.

Fozard, J. L., Wolf, E., Bell, B., McFarland, R. A. & Padolsky, S. (1977). Visual perception and communication. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the Psychology of Aging* (pp. 497-534). New York: Van Nostrand Reinhold.

Galganski, M. E., Fuglevand, A. J., & Enoka, R. M. (1993). Reduced control of motor output in a human hand muscle of elderly subjects during submaximal contractions. *Journal of Neurophysiology, 69*(6), 2108-2115.

García-Cáceres, R. G., Felknor, S., Cordoba, J. E., & Caballero, J. P. (2012). Hand anthropometry of the Colombian floriculture workers of the Bogota plateau. *International Journal of Industrial Ergonomics, 42*, 183-198.

Gilbert, B. G., Hahn, H. A., & Gilmore, W. E. (1988). Thumbs up: Anthropometry of the first digit. *Human Factors, 30*, 747-750.

Gorham-Rowan, M. M., & Laures-Gore, J. (2006). Acoustic-perceptual correlates of voice quality in elderly men and women. *Journal of Communication Disorders, 39*, 171-184.

Hackel, M. E., Wolfe, G. A., Bang, S. M., and Canfield, J. S. (1992). Changes in hand function in the aging adult as determined by the Jebsen test of hand function. *Physical Therapy, 72*(5), 373-377.

Hart, S. & Staveland, L. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). North Holland: Elsevier.

Hoggan, E., Brewster, S. A., & Johnston, J. (2008). Investigating the effectiveness of tactile feedback for mobile touchscreens. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI), Florence, Italy*, 1573-1582.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65-70.

Holzinger, A., Searle, G., & Nischelwitzer, A. (2007). On some aspects of improving mobile applications for the elderly. In C. Stephanidis (Ed.), *Universal Access in HCI: Part 1* (pp. 923-932). Berlin: Springer-Verlag.

Honjo, I., Isshiki, N. (1978). Laryngoscopic and voice characteristics of aged persons. *Practica Oto-Rhino-Laryngologica, 71*(2), 171-175.

Horn, J. (1989). Models of intelligence. In R. L. Linn (Ed.), *Intelligence: Measurement, theory and public policy* (pp. 29-73). Urbana: University of Illinois Press.

Huang, D. Z., Minifie, F. D., Kasuya, H., & Lin, S. X. (1995). Measures of vocal function during changes in vocal effort level. *Journal of Voice, 9*(4), 429-438.

IBM Corporation. (1994). Simon says “here’s how”! [Simon owner’s manual].

International Data Corporation. (2013, March 4). Smartphones expected to outship features phones for first time in 2013, according to IDC [Press release]. Retrieved March 20, 2013, from http://www.idc.com/getdoc.jsp?containerId=prUS23982813#.UWTPPqIqX-a

[iPhone onscreen keyboard image]. (n. d.). Retrieved April 30, 2013, from http://c185824.r24.cf1.rackcdn.com/3835484-3076032-thumbnail.jpg.

Jay, G. M. (1989). *The influence of direct computer experience on older adults’ computer attitudes, skills, and continued use*. Unpublished doctoral dissertation, The Pennsylvania State University, University Park.

Jay, G. M., & Willis, S. L. (1992). Influence of direct computer experience on older adults’’ attitudes toward computers. *Journal of Gerontology: Psychological Sciences, 47*, 250-257.

Jebsen, R. H., Taylor, N., Trieschmann, R. B., Trotter, M., & Howard, L. (1969). An objective and standardized test of hand function. *Physical Medicine and Rehabilitation*, *50*, 311-319.

Jerram, J. C. K., Kent, B., & Searchfield, G. D. (2010). Communication experiences for the older hearing-impaired: The uptake and usage of modern communication devices. *Health Care and Informatics Review Online, 14*(2), 36-41.

Jin, Z. X., Plocher, T., & Kiff, L. (2007). Touch screen user interfaces for older adults: Button size and spacing. In C. Stephanidis (Ed.), *Universal Access in HCI: Part 1* (pp. 933-941). Berlin: Springer-Verlag.

Joy, S., Kaplan, E., & Fein, D. (2004). Speed and memory in the WAIS-III Digit Symbol-Coding subtest across the adult lifespan. *Archives of Clinical Neuropsychology, 19*, 759-767.

Kalra, L., Jackson, S. H. D., & Swift, C. G. (1993). Assessment of changes in psychomotor performrance of elderly subjects. *British Journal of Clinical Pharmacology, 36*(5), 383-389.

Karat, C.-M., Halverson, C., Horn, D., & Karat, J. (1999). Patterns of entry in large vocabulary continuous speech recognition systems. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI). Pittsburgh, PA,* 568-574.

KayPENTAX. (2007). Multi-dimensional voice program (MDVP) model 5105 [Software Instruction Manual]. Lincoln Park, NJ: PENTAX Medical Company.

KayPENTAX. (2008). Multi-dimensional voice program (MDVP) model 5105 [Computer software]. Lincoln Park, NJ: PENTAX Medical Company.

Kenshalo, D. R. (1979). *Sensory function of the skin.* Plenum Press: New York.

Klemmer, E. T. (1971). Keyboard entry. *Applied Ergonomics*, *2,* 2-6.

Kobayashi, M., Hiyama, A., Miura, T., Asakawa, C., Hirose, M., & Ifukube, T. (2011). Elderly user evaluation of mobie touchscreen interactions. In P. F. Campos, T. C. N. Graham, J. A. Jorge, N. J. Nunes, P. A. Palanque, & M. Winckler (Eds.), *INTERACT 2011. Lecture Notes in Computer Science 6946* (pp. 83-99). Berlin: Springer.

Költringer, T. & Grechenig, T. (2004). Comparing the immediate usability of Graffiti 2 and virtual keyboard. *Extended Abstracts of the ACM Symposium on Human Factors in Computing Systems (CHI). Vienna, Austria,* 1175 – 1178.

Kristensson, P. O. (2007). *Discrete and continuous shape writing for text entry and control*. Linköping, Sweden: LiU-Tryck.

Kristensson, P. O. & Denby, L. C. (2009). Text entry performance of state of the art unconstrained handwriting recognition: A longitudinal user study. *In Proceedings of the ACM Symposium on Human Factors in Computing Systems (CHI). Boston, MA,* 567-570.

Kuwabara, T. (1975). The maturation process of the lens cell: A morphologic study. *Experimental Eye Research, 20,* 427-443.

Lewis, J. R. (1999). Input rates and user preference for three small-screen input methods: Standard keyboard, predictive keyboard, and handwriting. *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting, Houston*, 425-428. Santa Monica, CA: Human Factors and Ergonomics Society.

Lewis, J. R., & Sauro, J. (2009). The factor structure of the System Usability Scale. *In Proceedings of the Human Computer Interaction International Conference (HCII 2009). San Diego, CA*, 94 – 103.

Li, T., & Jo, C. (2005). Parameter changes across different emotions in human speech. *In Proceedings of the 2005 IEEE International conference on Information Acquisition. Hong Kong and Macau, China*, 91-94.

Linville, S. E. (2000). *Vocal aging.* San Diego: Singular.

MacKenzie, I. S., & Soukoreff, R. W. (2002). Text entry for mobile computing: Models and methods, theory and practice. *Human-Computer Interaction, 17*, 147-198.

MacKenzie, I. S. & Soukoreff, R. W. (2003). Phrase sets for evaluating text entry techniques. *Extended Abstracts of the ACM Symposium on Human Factors in Computing Systems (CHI).* 754-755. New York: ACM.

MacKenzie, I. S., & Zhang, S. (1997). The immediate usability of Graffiti. *In Proceedings of Graphics Interface*, 129-137. Toronto: Canadian Information Processing Society.

Madison, J. (2011). *Damn you, autocorrect!* Hyperion: New York.

Matias, E., MacKenzie, I. S., & Buxton, W. (1996). One-handed touch typing on a QWERTY keyboard. *Human Computer Interaction, 11*, 1-27.

Milian, M. (2011, October 4). Apple's Siri voice assistant based on extensive research. *CNN Tech*. Retrieved March 25, 2013, from http://www.cnn.com/2011/10/04/tech/mobile/iphone-siri

Möller, S., Gödde, F., & Wolters, M. (2008). Corpus analysis of spoken smarthome interactions with older users. *Proceedings of the 6th International Language Resources and Evaluation (LREC), Marrakech, Morocco*, 735-740.

Molloy, D. W., Alemayehu, E., & Roberts, R. (1991). Reliability of a standardized Mini-Mental State Examination compared with the traditional Mini-Mental State Examination. *American Journal of Psychiatry, 148*, 102-105.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology, 4*(2), 61-64.

Moroney, W. F., Biers, D. W., Eggemeier, F. T., & Mitchell, J. A. (1992). A comparison of two scoring procedures with the NASA task load index in a simulated flight task. *In Proceedings of the IEEE NAECON 1992 National Aerospace and Electronics Conference,* 734-739. New York: IEEE.

Motorola. (2009). Hello humans: DROID by Motorola arrives next week [Press release]. Retrieved March 25, 2013, from http://mediacenter.motorola.com/content/detail.aspx?releaseid=12058&newsareaid=2

Newell, A. F., & Gregor, P. (2004). HCI and older and disabled people: Applied computing at University of Dundee, Scotland. In A. Dearden & L. Watts (Eds.), *The 18th British HCI Group Annual Conference* (pp. 205-206). Leeds: Leeds Metropolitan University, UK.

Nielsen. (2012a, May 7). America's new mobile majority: A look at smartphone owners in the U.S. *Newswire*. Retrieved March 20, 2013, from http://www.nielsen.com/us/en/newswire/2012/who-owns-smartphones-in-the-us.html

Nielsen. (2012b, July 12). Two thirds of new mobile buyers now opting for smartphones. *Newswire*. Retrieved March 20, 2013, from http://www.nielsen.com/us/en/newswire/2012/two-thirds-of-new-mobile-buyers-now-opting-for-smartphones.html

Nokia. (1996, September 19). Nokia unveils worlds first all-in-one communicator for the Americas [Press release]. Retrieved March 20, 2013, from http://press.nokia.com/1996/09/19/nokia-unveils-world%C2%92s-first-all-in-one-communicator-for-the-americas/

[Nokia Communicator 9000i image]. (n.d.). [Photograph]. Retrieved March 25, 2013, from http://conversations.nokia.com/wp-content/uploads/2012/09/nokia-9000i.jpeg

Nuance Communications. (2011). FlexT9 [Text input method software].

Owsley, C., Sekuler, R., & Siemsen, D. (1983). Contrast sensitivity throughout adulthood. *Vision Research, 23,* 689-699.

Oryl, M. (2012, April 19). HTC moving away from QWERTY keyboards in future smartphones. *Mobileburn*. Retrieved March 20, 2013, from http://www.mobileburn.com/19325/news/htc-moving-away-from-qwerty-keyboards-in-future-smartphones

Pedlow, R., Kasnitz, D., & Shuttleworth, R. (2010). Barriers to the adoption of cell phones for older people with impairments in the USA: Results from an expert review and field study. *Technology and Disability, 22,* 147-158.

Pierce, C. A., Block, R. A., & Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor ANOVA designs. *Educational and Psychological Measurement, 64*(6), 916-924.

Rabbitt, P. (1979). How old and young subjects monitor and control responses for accuracy and speed. *Brittish Journal of Psychology, 70*, 305-311.

Rau, P-L. P., & Hsu J.-W. (2007). Interaction devices and web design for novice older users. *Educational Gerontology, 31*(1), 19-40.

Read, J., MacFarlane, S., & Casey, C. (2001). Measuring the usability of text input methods for children, in A. Blandford, J. Vanderdonckt & P. Gray (Eds.), *People and Computers XV: Interaction without frontiers* (pp. 559-572). Berlin: Springer-Verlag.

Ryan, J. J., Schnackenberg, S. D., & Kreiner, D. S. (2001). Effects of altered instructions on WAIS-III Digit Symbol-Coding subtest. *Psychological Reports, 88*, 987-988.

Salthouse, T. A. (1992). What do adult age differences in the Digit Symbol Substitution Test reflect? *Journal of Gerontology, 47*(3), 121-128.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*(3), 403-428.

Salthouse, T. A., Hancock, H. E., Meinz, E. J., & Hambrick, D. Z. (1996). The influence of visual acuity on age-cognition relations. *Journals of Gerontology: Psychological Sciences, 51B*, 317-330.

Sauro, J. (2011, February 2). Measuring usability with the System Usability Scale (SUS). *Measuring Usability*. Retrieved March 20, 2013, from http://www.measuringusability.com/sus.php

Silfverberg, M. (2007). Historical overview of consumer text entry technologies. In I. S. MacKenzie & K. Tanaka-Ishii (Eds.), *Entry Systems: Mobility, Accessibility, Universality* (pp. 3-26). Amsterdam: Morgan Kaufmann.

Smith, A. (2011a, July 11). 35% of American adults own a smartphone. *Pew Internet*. Retrieved October 17, 2011, from http://www.pewinternet.org/Reports/2011/Smartphones

Smith, A. (2011b, September 19). Americans and text messaging. *Pew Internet*. Retrieved October 17, 2011, from http://www.pewinternet.org/Reports/2011/Smartphones

Smith, A. (2012, March 1). 46% of American adults are smartphone owners. *Pew Internet*. Retrieved March 25, 2013, from http://pewinternet.org/~/media/Files/Reports/2012/Smartphone%20ownership%202012.pdf

Soukoreff, R. W. (2002). *Text entry for mobile systems: Models, measures, and analyses for text entry research.* Unpublished master's thesis, York University, Toronto, Canada.

Soukoreff, R. W., & MacKenzie, I. S. (2001). Measuring errors in text entry tasks: An application f the Levenshtein string distance statistic. *Extended Abstracts of the ACM Conference on Human Factors in Computing Systems (CHI), Seattle*, 319-320. New York: ACM.

Soukoreff, R. W., & MacKenzie, I. S. (2003). Metrics for text entry research: An evaluation of MSD and KSPC, and a new unified error metric. *Proceedings of CHI03*, 113-120. New York: ACM.

Soukoreff, R. W., & MacKenzie, I. S. (2004). Recent developments in text-entry error rate measurement. *Extended Abstracts of the ACM Conference on Human Factors in Computing Systems (CHI), Vienna, Austria*, 1425-1428. New York: ACM.

Starns, J. J., & Ratcliff, R. (2010). The effects of aging on the speed-accuracy compromise: Boundary optimality in the diffusion model. *Psychology and Aging, 25*(2), 377-390.

Stone, R. (2008). Mobile touch interfaces for the elderly. *In Proceedings of IADIS International Conference ICT, Society and Human Beings, Amsterdam, The Netherlands*, 230-234.

Stoyanchev, S. (2012, May). Voice interfaces for smartphones. *IEEE Speech and Language Processing Technical Committee Newsletter*. Retrieved March 30, 2013, from http://www.signalprocessingsociety.org/technical-committees/list/sl-tc/spl-nl/2012-05/voice-interfaces/

Stroud, D. (2012, March 9). Nielsen data about smartphone use in the US by age and income. *20Plus30 Consulting*. Retrieved March 30, 2013, from http://20plus30.blogspot.com /2012/03/nielsen-data-about-smartphone-use-in-us.html

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson Education, Inc.

TechSmith. (2011). Morae Recorder (Version 3.2.1) [Computer software]. Okemos, MI: TechSmith.

Tombaugh, T. N., & McIntyre, N. J. (1992). The Mini-Mental State Examination: A comprehensive review. *Journal of the American Geriatrics Society, 40*(9), 922-935.

Tomporowski, P. D. (2003). Performance and perceptions of workload among young and older adults: Effects of practice during cognitively demanding tasks. *Educational Gerontology, 29*, 447-466.

Torre, P., III, & Barlow, J. A. (2009). Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders, 42*, 324-333.

Trites, R. L. (1989). Lafayette Grooved Pegboard Task [Instruction/Owner’s Manual]. Lafayette, IN: Lafayette Instrument Company.

Tullis, T., & Stetson, J. N. (2004, June). *A comparison of questionnaires for assessing website usability.* Paper presented at the Usability Professionals Association (UPA) 2004 Conference, Minneapolis, MN.

Van Drempt, N., McCluskey, A., & Lannin, N. A. (2011). A review of factors that influence handwriting performance. *Australian Occupational Therapy Journal, 58*, 321-328.

Van Riper, C. (1963). *Speech correction: Principles and methods* (4th ed.) (pp. 484). Englewood Cliffs, NJ: Prentice Hall.

Vercruyssen, M. (1996). Movement control and the speed of behavior. In A. D. Fisk & W. A. Rogers (Eds.), *Handbook of Human Factors and the Older Adult* (pp. 55-86). San Diego: Academic Press.

Vipperla, R., Renals, S., & Frankel, J. (2008). Longitudinal study of ASR performance on ageing voices. *In Proceedings of the 9th Annual Conference of the International Speech Communication Association (INTERSPEECH), Brisbane, Australia,* 2550-2553.

Vipperla, R., Renals, S., & Frankel, J. (2010). Ageing voices: The effect of changes in voice parameters on ASR performance. *EURASIP Journal on Audio Speech, and Music Processing, 2010,* Article 525783. Retrieved October 22, 2011 from http://www.hindawi.com/journals/asmp/2010/525783/

Wacharamanotham, C., Hurtmanns, J., Mertens, A., Kronenbuerger, M., Schlick, C., & Borchers, J. (2011). Evaluating swabbing: A touchscreen input method for elderly users with tremor. *Proceedings of the ACM Symposium on Human Factors in Computing Systems (CHI),* Vancouver, BC, Canada, 623-626.

Walker, N., Philbin, D. A., & Fisk, A. D. (1997). Age-related differences in movement control: Adjusting submovement structure to optimize performance. *Journals of Gerontology: Psychological Sciences, 52B(*1), 40-53.

Ward, D. J., Blackwell, A. F., & MacKay, D. J. C. (2000). Dasher – A data entry interface using continuous gestures and language models. *In Proceedings of the ACM Conference on User Interface and Software Technology (UIST)*. 129-137. New York: ACM.

Weintraub, N., Gilmour-Grill, N., & Weiss, P. L. (2010). Relationship between handwriting and keyboarding performance among fast and slow adult keyboarders. *American Journal of Occupational Therapy, 64*(1), 123-132.

Wilpon, J. G., & Jacobsen, C. N. (1996). A study of speech recognition for children and the elderly. *In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* Atlanta, GA, 349-352.

Wobbrock, J. O., Myers, B. A., & Aung, H. H. (2006). Few-key text entry revisited: Mnemonic gestures on four keys. *ACM Conference on Human Factors in Computing Systems (CHI),* 489-492. New York: ACM.

Wobbrock, J. O. (2007). Measures of text entry performance. In I. S. MacKenzie & K. Tanaka-Ishii (Eds.), *Entry Systems: Mobility, Accessibility, Universality* (pp. 47-74). Amsterdam: Morgan Kaufmann.

Xue, S. A., & Hao, G. J. (2003). Changes in the human vocal tract due to aging and the acoustic correlates of speech production: A pilot study. *Journal of Speech, Language, and Hearing Research, 46*, 689-701.

Yamada, H. (1980). A historical study of typewriters and typing methods: from the position of planning Japanese parallels. *Journal of Information Processing, 2*(4), 175-202.

Zhai, S., & Kristensson, P. O. (2003). Shorthand writing on stylus keyboard. *Proceedings of CHI 2003, ACM Conference on Human Factors in Computing Systems. CHI Letters, 5,* 97-104.

APPENDICES

APPENDIX A

QUESTIONNAIRES

Background Questionnaire

Thank you for your interest in our mobile device studies. These studies will be conducted at Wichita State (WSU) and take approximately one (1) to three (3) hours to complete. Qualified participants will be compensated for their time (e.g., monetary OR class participation points will be awarded, depending on the study).

Please complete the following survey if you are interested in participating. If you qualify, a representative from the Software Usability Research Lab (SURL) at WSU will contact you to schedule a time for participation. Limited availability!

<u>Basic Demographics</u>
What is your gender?
    Male   Female
What is your age? _______
What is the highest level of education you have completed?
    High School
    Some College
    Associate's Degree
    Bachelor's Degree
    Master's Degree
    Doctorate Degree
    Other ________
What is your current occupation? ___________
Are you an employee of Wichita State University?
    Yes   No
How did you arrive at this questionnaire?
    Flyer on Wichita State University campus (Main, West, etc.)
    Flyer on another university campus
    Wichita State University SONA system
    Link sent by friend/acquaintance
    Friend/Word of mouth
    Other ____________

<u>Hands</u>
Are you:
    Left-handed   Right-Handed   Ambidextrous
Please list any past or current finger, hand or wrist problems. (This includes Carpal Tunnel Syndrome, Arthritis, Fractures, etc.) ____________

Please list any movement disorders that you have. (This includes Parkinson's disease, Tourette syndrome, Huntington's, etc.) ____________

Communication

What is your native language? _________

How many years have you spoken English? _________

Have you ever been diagnosed with a communication (e.g., voice or speech) disorder?

Yes No

(If yes to previous question):

What was the disorder? (Please explain.) _______

Has it been resolved through therapy?

Yes No

(If yes to previous question):

How long ago was this disorder resolved through therapy? _______

General Health

Do you have any hearing impairments?

Yes No

Do you have normal/corrected vision?

Yes No

Do you currently have a cold or illness? If so, please describe. ________

Do you currently, or did you ever, smoke tobacco (e.g., cigarettes, pipe, etc.)?

Yes No

(If yes to the previous question):

For how many years have you smoked tobacco? _______

How many times do you smoke tobacco each week? _______

If you quit smoking tobacco, how long has it been? _______

Mobile Device Ownership

Do you currently, or have you ever, owned a tablet device? (e.g., iPad, Motorola Xoom, etc.)

Yes No

If "yes", how long have you owned a tablet device? ________

Do you currently, or have you ever, owned a mobile phone?

Yes No

(If yes to previous question):

Mobile Phone Device Methods

How long have you owned a mobile phone? _____

What is the brand and model of your current mobile phone? (e.g., BlackBerry Bold, Motorola DroidX, LG Rumor, etc.) _________

Does your current mobile phone have a touchscreen?
- Yes
- No

Which method do you MOST frequently use to input text on your mobile phone?
- Physical numberpad
- On-screen keyboard (touchscreen Qwerty)
- Physical mini-Qwerty keyboard
- Other: ______

List the experience that you have with each of these mobile phone input methods:
(1 = None, 5 = Expert)
- Physical Numberpad
- Physical Qwerty Keypad
- On-screen Qwerty Keypad
- Handwriting (e.g., Graffiti, DioPen, FlexT9, etc.)
- Tracing (e.g., ShapeWriter, SWYPE, etc.)
- Stylus

Which orientation do you MOST frequently use when entering text on your mobile phone?
- Portrait (vertical)
- Landscape (horizontal)

Do you use the autocorrect/predictive text feature on your mobile phone?
- Yes
- No
- Not sure

What do you LIKE about the design of your mobile phone's keyboard? __________
What do you DISLIKE about the design of your mobile phone's keyboard? ___________
What do you find to be the BIGGEST problem(s) in using your mobile phone? (Check all which apply.)
- Not being able to find an option I am looking for
- Not being able to understand terms used in the menus
- Not being able to understand icons used in menus
- Display is difficult to read/see
- Keypad is difficult to use
- Touchscreen is difficult to use
- It costs too much
- Other _______

<u>VR Software – Computers</u>

Have you ever used Voice Recognition Software on a computer (e.g., Dragon Naturally Speaking, MacSpeech, Windows Speech Recognition, etc.)?
- Yes No
- (If yes to the previous question):
- How much experience do you have with the Voice Recognition software feature on computers? (1 = Novice, 5 = Expert)

For which tasks do you tend to use the Voice Recognition software feature? ______
What are your impressions of the accuracy of the Voice Recognition software? ____
(If no to the previous question):
Why have you not used Voice Recognition Software on computers in the past? _____

VR Software – Mobile

Have you ever used Voice Recognition Software on a computer (e.g., Genius button, VLingo, Google Speak, Siri, etc.)?

Yes No

(If yes to the previous question):
How much experience do you have with the Voice Recognition software feature on mobile phones? (1 = Novice, 5 = Expert)
For which tasks do you tend to use the Voice Recognition software feature? ______
What are your impressions of the accuracy of the Voice Recognition software? ____
(If no to the previous question):
Why have you not used Voice Recognition Software on mobile phones in the past? __

Participation

Would you be interested in being contacted for future studies regarding mobile devices (e.g., mobile phones, tablets, etc.?)

Yes
No

Please indicate if you have participated in a previous study through the Wichita State University Software Usability Research Lab (SURL). (Select all which apply.)

Cell phone
Tablet PC
Multi-Monitor Computing
Laptop security
None
Other ______

(If yes to the above participation question in the previous section):

Contact Info

Please provide some contact information so that we can contact you for participation in future studies.

Do you have a valid US Social Security Number?

Yes
No

What is your first and last name? _____
Please list the e-mail address that you check most frequently. ______
Please list your phone number. ______

Thank you for participating in this survey!

Attitudes Towards Computers Questionnaire

This questionnaire was developed by Jay (1989), and consists of 35 questions that assess seven dimensions of computer-related attitudes: comfort (comfort or familiarity with computers), efficacy (self-perceptions of computer competence), gender equality (belief that both men and women find computers important), interest (the desire to learn and use computers), control (belief that people are in control of computers), dehumanization (the degree to which computers are seen as dehumanizing), and utility (the usefulness of computers) (Jay & Willis, 1992).

Participants in this study were given the ATCQ prior to the performance portion of the study. Afterward, the scales were constructed by reverse-coding appropriate items (denoted in the following questionnaire) and summing the items for each scale. The comfort dimension consisted of items 1, 13, 18, 31 and 34; the efficacy dimension consisted of items 10, 22, 24, 29 and 35; the gender equality dimension consisted of items 2, 4, 11, 16 and 19; the control dimension consisted of items 3, 23, 26, 28 and 33; Dehumanization consisted of 6, 7, 8, 12, 21, and 33 (reverse coded); Interest consisted of items 5, 9, 15, 17 and 30, and utility consisted of items 14, 20, 25, 27, 28 and 32. A reliability analysis was conducted for the items on each dimension, and Cronbach's alphas were similar to reliability indices reported in the literature for older adults (Jay & Willis, 1992). Once the items were summed for each dimension, the total scores were normalized, and a reliability analysis was conducted among the seven dimensions. Only four dimensions were found to reliably reflect computer attitudes for this sample: comfort, efficacy, interest and utility (Cronbach's $\alpha = .76$). The other three dimensions: gender equality, control and dehumanization, all appeared to be assessing other types of attitudes, and were left out of the analysis.

The questionnaire is as follows:

The below items are rated along a 5-point Likert response scale (1 = strongly agree, 2 = agree, 3 = neither agree or disagree, 4 = disagree, and 5 = disagree strongly). Reversed items are denoted by (R).

1. I feel comfortable with computers. (R)
2. Using computers is more important for men than for women.
3. Computers will never replace the need for working human beings. (R)
4. More women than men have the ability to become computer scientists.
5. Learning about computers is a worthwhile and necessary subject. (R)
6. Computers turn people into just another number. (R)
7. The use of computers is lowering our standard of living. (R)
8. Computers control too much of our world today. (R)
9. Reading or hearing about computers would be (is) boring.
10. I know that if I worked hard to learn about computers, I could do well. (R)
11. Using computers is more enjoyable for men than it is for women.
12. Computers are making the jobs done by humans less important. (R)
13. Computers make me nervous.
14. Life will be (is) harder with computers.
15. I don’t care to know more about computers.
16. Working with computers is more for women than for men.

17. Computers would be (are) fun to use. (R)

18. I don't feel confident about my ability to use a computer.

19. Women can do just as well as men in learning about computers. (R)

20. Everyone could get along just fine without computers.

21. Computers are dehumanizing. (R)

22. Computers are *not* too complicated for me to understand. (R)

23. Our world will never be completely run by computers. (R)

24. I think I am the kind of person who would learn to use a computer well. (R)

25. It is *not* necessary for people to know about computers in today's society.

26. People are smarter than computers. (R)

27. Computers are too fast.

28. People will always be in control of computers. (R)

29. I think I am capable of learning to use a computer. (R)

30. Learning about computers is a waste of time.

31. Computers are confusing.

32. Computers make the work done by people more difficult.

33. Soon our lives will be controlled by computers. (Reversed for "Dehumanization" only)

34. Computers make me feel dumb.

35. Given a little time and training, I know I could learn to use a computer. (R)

System Usability Scale

The System Usability Scale (Brooke, 1996) was developed in 1986 by John Brooke, who was at the time working for the Digital Equipment Corporation. The point was to provide a quick instrument for evaluating usability, and it has applied to many different types of systems (including websites, mobile devices, software, etc.). The SUS is a 10-item survey, with each item measured on a 1 – 5 Likert-type scale (1 = strongly disagree, 5 = strongly agree). Once the data is obtained, 1 is subtracted from the odd-numbered data, and 5 is subtracted from the even numbered scores, which in effect reverse-scores these items. The recoded data is summed, and then multiplied by a constant of 2.5.

The range of satisfaction scores is from 0 to 100, and the higher score indicates higher levels of usability and perceived satisfaction with the system being tested. This instrument is commonly used to measure satisfaction in usability studies, and has been found to be reliable (Tullis & Stetson, 2004) and valid. Sixty-eight is the reported average score for this instrument (Sauro, 2011). To aid interpretation, a “grading” system has been suggested, which assigns grades “A” through “F” to percentiles of the overall distribution of SUS scores from a large collection of studies using the measure. SUS scores above 80.3 were considered to have achieved an “A” and scores below 51, an “F” (Sauro, 2011). Additionally, it has been found that the SUS is not a uni-dimensional measure, and can interpret both usability and learnability of a system (Lewis & Sauro, 2009). Items four and ten reflect the learnability scale.

This instrument has been slightly modified, changing the wording from “system” to “input method”, and “cumbersome” to “awkward” on question 8, recommended by Lewis and Sauro (2009).

APPENDIX A (continued)

The modified SUS used in this study:

1. I think that I would like to use this input method frequently

1	2	3	4	5

2. I found the input method’s navigation (e.g., up, down, left, right) unnecessarily complex.

1	2	3	4	5

3. I thought the input method was easy to use

1	2	3	4	5

4. I think that I would need the support of a technical person to be able to use this input method.

1	2	3	4	5

5. I found the selection functionality (e.g., tap, double tap, press) in this input method was well integrated.

1	2	3	4	5

6. I thought there was too much inconsistency in this input method.

1	2	3	4	5

7. I would imagine that most people would learn to use this input method very quickly

1	2	3	4	5

8. I found the input method very awkward to use

1	2	3	4	5

9. I felt very confident using the Input method

1	2	3	4	5

10. I needed to learn a lot of things before I could get going with this input method

1	2	3	4	5

NASA Task Load Index

The NASA Task Load Index (NASA-TLX), developed by Hart and Staveland (1988), is a subjective rating method for assessing six dimensions of workload: mental, physical, temporal, performance, effort and frustration. Each dimension is assessed by 1 item, rated on a 1 – 21 scale. Traditionally, the instrument was developed to include weights derived from paired-comparisons between workload dimensions. However, this portion of the evaluation adds considerable time. Research has shown that the correlation between the original NASA-TLX method and a "raw" variation, consisting only of the initial rating scales, was high (r = .94 to .98) (Moroney et al., 1992; Byers et al., 1989). Thus, this study used the raw data without asking participants to do the paired-comparisons portion. Additionally, the individual dimensions were found to have adequate reliability for each input method (see Table A1), so the dimensions for each input method were summed together to create composites before running analyses on workload ratings between text input methods.

TABLE A1

RELIABILITY OF WORKLOAD COMPOSITE SCALES

	Physical Qwerty	Onscreen Qwerty	Tracing	Handwriting	Voice
Younger Adult Group	.851	.771	.891	.831	.713
Older Adult Group	.812	.837	.889	.827	.741

Note: Cronbach's alphas are for unstandardized scores.

APPENDIX A (continued)

Mental Demand
How mentally demanding was the task?

1	2	3	4	5	6	7	8	9	10	**11**	12	13	14	15	16	17	18	19	20	**21**

Very Low Very High

Physical Demand
How physically demanding was the task?

1	2	3	4	5	6	7	8	9	10	**11**	12	13	14	15	16	17	18	19	20	**21**

Very Low Very High

Temporal Demand
How hurried or rushed was the pace of the task?

1	2	3	4	5	6	7	8	9	10	**11**	12	13	14	15	16	17	18	19	20	**21**

Very Low Very High

Performance
How successful were you in accomplishing what you were asked to do?

1	2	3	4	5	6	7	8	9	10	**11**	12	13	14	15	16	17	18	19	20	**21**

Perfect Failure

Effort
How hard did you have to work to accomplish your level of performance?

1	2	3	4	5	6	7	8	9	10	**11**	12	13	14	15	16	17	18	19	20	**21**

Very Low Very High

Frustration
How insecure, discouraged, irritated, stressed, and annoyed were you?

1	2	3	4	5	6	7	8	9	10	**11**	12	13	14	15	16	17	18	19	20	**21**

Very Low Very High

APPENDIX B

OPEN ENDED COMMENTS ADDRESSING INITIAL IMPRESSIONS OF THE INPUT METHODS

Prior to the text entry tasks, participants were given a short introduction to each input method, which mostly drew upon actual advertising or marketing statements for the input devices tested. They were asked to rate the input methods on a 0 – 50 scale (see Figure B1) based on their initial impressions of desirability and value of each input method for entering text on a smartphone device. Participants were not allowed to see the input methods at this time, but instead were given laminated notecards with a descriptive word (i.e., “Handwriting”, “Voice”) printed on one side. They were asked to place the cards along the scale, and to give comments about why they assigned the input methods each rating. Tables B1 through B5 give summaries of the users’ initial impressions for each of the input methods.

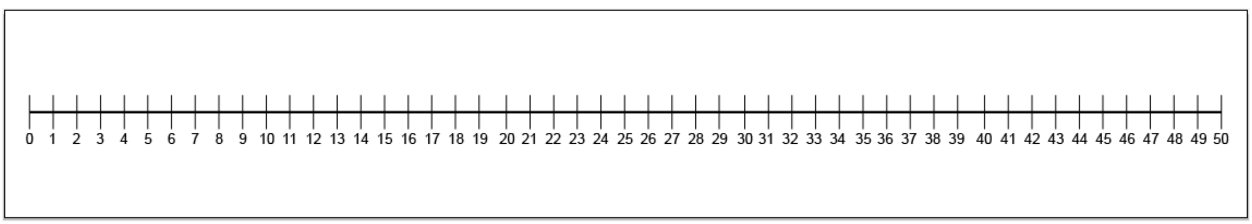


Figure B1. The 0 – 50 point rating scale, used for preference ratings.

Pre-Test Comment Summary

Physical Qwerty**.** The younger participants expressed that they initially thought the Physical Qwerty input method was familiar to them, regarding the layout of the keyboard and the mode of interaction (physical key press). Many of the participants reported that they had a physical numeric keypad on their phones, and felt that they were comfortable interacting with their phone in this way, which would transfer to the smartphone. They felt this input method would give them more control over the output, compared to others, and that their computer keyboarding skills might transfer to this method.

Initially, older adults felt that the Physical Qwerty was comfortable and familiar to them, because of their exposure to typewriter and computer Qwerty keyboards. Touch typists seemed to feel that this method would be very accurate, fast and convenient. They reported that they liked having the tactile responsiveness of a push button, but were somewhat hesitant about the size of the keys on a small device, as well as how much force would be needed to depress the keys. After entering twenty phrases into the smartphone using the Physical Qwerty keypad, participants reported that they liked the haptic and audio “click” feedback when the buttons were depressed. They enjoyed the feeling that they had more control over the input method, because they had to be intentional with each key press. The familiar Qwerty layout was helpful in finding the keys, and it seemed to be simple to use. Some liked the size of the keys, and some thought they were too small. Some additional dislikes included the feeling that the input method was slow, that the “click” of the keypad was annoying, and that there was too much key resistance. Others reported that they were disappointed that they were not able to transfer their keyboarding skills to this input method. One suggested that the keyboard materials seemed “cheap”.

***Onscreen Qwerty*.** Younger participants’ initial impressions of the Onscreen Qwerty method indicated that they thought this input method would require less hand strain than its physical counterpart, as there were no buttons to depress. They liked that it had the familiar Qwerty layout and anticipated that they would be able to transfer their keyboarding skills to this method. Some commented that they thought that the use of touch screen technology make this method “cool” and “new”, and some perceived that it must be a trustworthy product if it was being released commercially. There were some concerns about the touch screen, mainly in its lack of haptic feedback and that it would be difficult to orient fingers to the correct location without being able to feel keys or any type of “home row” markers. There were also concerns about the small size of the keypad and how this would affect accuracy.

Although the onscreen Qwerty had the same layout as its physical counterpart, older adults seemed to be uncertain about interacting with the touch screen technology, unless they had prior positive experiences with other similar technologies (e.g., Apple’s iPad). The main concerns addressed the lack of tactile feedback and being limited in entering text by only using one finger (many of them had this impression, even though nothing was mentioned beforehand regarding the number of fingers they could use to input text with this method). Some felt it might be faster, as there were no physical buttons to depress, however, they also felt that it would be less accurate and that they would easily press more than one button simultaneously. A couple of participants were concerned that their occasional problems with hand steadiness would create less accurate results with this method.

Tracing. Pre-test impressions of the Tracing input method by the younger adults overwhelmingly revolved around uncertainty about the technology and unfamiliarity with gesturing as a way to enter text. Participants thought that this input method would be less accurate, since the trace includes letters that are not in the intended word. They also felt that it would be more labor intense to keep a finger constantly on the screen, and also for the amount of corrections they anticipated they would have to make. Some participants thought that it would be more difficult for people who were not familiar with the Qwerty layout, and that the method itself would take time to learn. Some felt that it was an interesting way to interact with the phone, and some indicated that they trusted the adequacy of this input method simply because it was commercially available.

Initially, older participants thought the tracing input method was very unfamiliar and unclear. Many participants reported that they did not understand how technology could “know” which keys you intended to include in your word, nor did they trust that the technology would perform well. This uncertainty seemed to influence their perceptions of decreased accuracy and speed with this method. Other concerns were whether this method would work well with unsteady hands, and knowing how much pressure to use. Some felt that this method would have easy, smooth and fast interactive qualities.

Handwriting. Younger adults’ initial impressions of the Handwriting input method were mixed. Participants who felt that they had legible handwriting assumed that the software would work well. Those who had poor handwriting expected that it would not recognize their input. Some felt that it was familiar to them, as people write every day, but others felt that the technology used to interpret their handwriting was foreign and introduced uncertainty and insecurity about the input method. Many had concerns about their personal writing style and whether it would be interpreted correctly by the software, and some questioned whether they would have to change their handwriting to increase accuracy. A few thought it would be fast, but most thought it would be much slower, in that they would have to “draw” every character with their finger. Some did not feel there was any utility in this input method.

Many of the initial impressions of the handwriting method revolved around the older participants' personal impressions of their own handwriting style. If the participant felt that they had legible handwriting, they felt that this input method should be easy and were confident that it would perform well. If they had poor handwriting, they were not confident that the software could recognize their writing. Most felt that this input method would be familiar, straightforward, and appreciated that they would not have to use a keyboard to enter text. Many felt that this method would be slower than the others, as handwriting is generally slow. They were unsure about the technology behind this input method and thought that the concept was surreal. A couple of the left-handed participants expressed concern about how they might have to change their writing style so their hand would not drag across the device.

Voice. Younger participants' pre-test impressions of the Voice input method indicated they perceived it to be fast, easy, and hands-free, which would allow them to better multitask. They also thought that this method would be beneficial to those with disabilities. Some concerns included the effect of background noise, accents and natural voice variation on the output accuracy. They also wondered how the software handled homophones. Many expressed their doubts about the inadequacies of voice technology. There were also social concerns in that they did not want to look "weird" talking to their phone and the lack of privacy when composing a message in public.

Initially, older participants felt that the Voice input method would be fast and convenient. They liked that this method would be hands-free, which would be great for multitasking and also for people who have hand problems. Those who thought that they naturally spoke clearly had positive impressions of this method and figured it would work well. Others felt that they did not enunciate well, and that this method would reflect it by being less accurate. Some initial concerns included limitations on where it could be used, such as noisy environments or in public places where they may have privacy concerns. Those which had positive experiences with computer dictation software in the past felt that this would be a comfortable method to use, and thought that they would be able to easily train the software to recognize their voice. Some participants did not have a clear idea of what this method was, and thought it was similar to an automated voice system.

APPENDIX B (continued)

TABLE B1

PARTICIPANTS' INITIAL IMPRESSIONS OF PHYSICAL QWERTY

Younger Adults	**Older Adults**
I am most familiar with typing on a keyboard, and I am a fast typer.	I am a good touch typist, so the layout is familiar.
You can type fairly fast while you do something else (while using a desktop keyboard), so you can look at something else while you are typing, whereas you have to look at some of the other methods while using them.	I do not do a lot of typing, so the action of typing on a small keyboard may not be easy. Keyboards can be so sensitive. I learned to type on a typewriter and had to build up my finger muscles to use it, and I am not accustomed to the force needed to use this method.
There is more control with a physical keyboard because the phone will not have to interpret what I input.	It might be faster because you could type using both hands.
I like having individual letters (keys) to push.	I've been typing for 50 years, so it's in my habit.
I am comfortable typing at work (desktop keyboard) and like having the feeling (tactile quality) of my fingers entering words.	I use my computer keyboard, and I know that it is easy to correct errors using that when I type.
The Physical Qwerty has a classic layout which is something we all learn to use.	There is more room for error.
I like the physical keyboard because you can tell where you are at based on the keys that you are hitting. You are able to easily reorient yourself.	I am familiar with entering text on my flip phone. I like having physical keys.
I would prefer this over other methods because I'm used to it. I use a keyboard at work.	Having a physical keyboard on a small device concerns me. I have large fingers.
It seems like the most precise way to interact with the phone, and that there would not be many mistakes. There are separate keys and you have more control over and can feel the keys you press.	I am not a very fast typist, and even though you get immediate feedback it might not always be the letter I intended to push. But at least you know you pressed it.
I am familiar with a computer keyboard, so it would take very little time to transition to a phone that used a physical Qwerty for text entry. It is a sure feeling that you are getting the right letters and words.	I have done this type of input with machines the longest, and once you've done something a long time, you feel more comfortable with it.
If the keys are raised up (on the Physical Qwerty) there might be more of a chance that I could accidentally hit the wrong key.	To use this, someone would have to know the layout of the keyboard and how much pressure you need in order to get a response.
I text a lot on a regular phone (T9) so being able to text on a small keyboard would be nice; more familiar.	I'm a hunt and peck typist, so I have to look at the keyboard.
We are introduced to keyboards at a young age, so I am comfortable with this. It would be faster, too.	We all type now and have keyboarding skills, so it would be convenient and fast.

APPENDIX B (continued)

TABLE B2

PARTICIPANTS' INITIAL IMPRESSIONS OF ONSCREEN QWERTY

Younger Adults	**Older Adults**
It might have less hand strain because there is less resistance to pressing a key, and it sticks to the familiarity of the Qwerty layout.	I'm not familiar with this method, but it sounds like it would be comfortable to use. I would prefer to use one finger instead of multiple fingers, though.
I do not think it will be as accurate because touch screens are not very accurate, and they are small on smartphones. To make the keyboard big enough to be convenient, you would have to make a huge phone.	I have seen touchscreens before in places, but I do not feel comfortable with them. I do not feel they are very precise. You cannot feel any physical space between keys.
The keyboard would obscure the rest of the screen.	Sometimes my hands are unsteady so I could accidentally type the wrong thing.
It would take more time because I think you can only select one letter at a time.	Seems more appropriate to use one finger or a fingernail because I think the keys would be smaller.
It is easier than pushing a button, but you might push more than one key at the same time. They keys would have to be tiny to fit a Qwerty on a phone.	There would not be any resistance. The surface is the same for all of the keys, so it is not as physically satisfying. I want to know I've poked something.
I have an iPad and am familiar with the keyboard, but I miss using a regular keyboard.	I would be afraid of being "heavy handed"; it could lead to misspelling if it is overly touch sensitive.
The layout is familiar, but you have to locate each key independently. You cannot rely on muscle memory because it is not at a scale that is familiar.	I see other people using their smartphones and they appear to be able to use this method quickly.
I like touchscreens. They are more compact and you do not have to make as large of movements.	This would be similar to the physical keyboard, but slower because you would have to use a single finger.
It should be easier to use than pressing physical keys.	Once someone shows me how, it should be easy.
It is cooler than using physical keys, because those have been around for awhile. It is a new technology.	Sometimes touchscreens are not very sensitive, but it depends on the instrument.
You should be able to use the same skills from typing on a keyboard, but I'm not sure about interacting with a touchscreen. How do you know whether a button has been pushed? There is no way to find the home row.	I have used this method on an iPad and I am comfortable and familiar with it.
It seems like it would be more precise because most people are proficient with keyboards, plus, it is new technology. You also do not have to press a hard key, which would make it more comfortable to use.	I think it is more gratifying to push a physical button. However, I think it might be quicker because you are just hitting the screen, not pushing a button.
The buttons would be small on a touchscreen and my fingers are large. I would have to backspace a lot!	It seems you would have to be more precise because your finger could hit another key very easily.
I view this method as the same as Physical Qwerty.	Tiring and tedious to ensure you hit the correct letter.

APPENDIX B (continued)

TABLE B3

PARTICIPANTS' INITIAL IMPRESSIONS OF TRACING

Younger Adults	**Older Adults**
It might be of some value to interact with the keyboard in this way, but it is unfamiliar and confusing. I am afraid I would have to edit a lot.	It might pick up letters in between those you intend while you trace from one side of the keyboard to the next to create your word. It might be more error prone.
It seems interesting, but how does the phone know which letters you are trying to get to?	It sounds easier to glide your finger than to lift it up to press the next key.
It does not sound easy to use, but it be better once you are familiar with it. It seems like it would flow smoothly and go fairly fast.	I have seen people demonstrate this method before, but I do not think it would be faster than the physical or onscreen Qwerty.
You rely on the phone's interpretation of your input.	Is this like cursive writing on a keyboard?
It would be less accurate because you trace over letters you do not want to include in your word.	I cannot imagine doing this. It would probably require a lot of editing.
I wouldn't be comfortable using this method.	It is unclear on what to expect from this method.
It is easier to create a word by putting it into a string of characters in comparison to entering one letter at a time with other methods.	Moving your finger in an uninterrupted pattern and not having to lift it should make it faster than some of the other methods.
If you do not know the Qwerty layout, then you might get lost on the keyboard.	Using gestures to communicate is foreign to me.
It is more difficult since not many people have used it.	It does not sound very accurate.
It would be easy since you use the traditional keyboard layout, but it might be more difficult since you cannot lift your finger.	I am having trouble being convinced that this technology would be good enough to select only the keys you intend to select.
I saw a commercial on TV advertising this method and it looks like it would be fast and easy to do.	It seems like it would be too slow because you would have to find each letter in an intentional search.
It would probably take awhile to get used to making gestures instead of typing. My friends use it sometimes. Some like it and some do not.	It seems to be a smoother way to input. You would not have to worry about placing your fingers on the right key as much as the Qwerty methods.
Seems like it would be very labor intensive.	I do not understand how this technology can work.
Sliding your finger is less time consuming than hitting each individual letter, so it might be quick.	This method of interaction does not seem "normal".
It seems very disconnected from any kind of writing or typing I have done. Unappealing and unfamiliar.	I'm not familiar with it. It would be hard to keep your hand pressure the same for the whole word.
It should be better than typing because the phone is offering this technology, I would assume it is better and more precise.	Sometimes my hands shake, so it might be less accurate for me to use.

APPENDIX B (continued)

TABLE B4

PARTICIPANTS' INITIAL IMPRESSIONS OF HANDWRITING

Younger Adults	**Older Adults**
The software probably will not recognize my handwriting, so I will have to edit a lot.	I know how to write, and it is generally legible, so this method would probably work well.
It seems like it would take more time than just typing.	I would be surprised if it recognized my handwriting!
This is more familiar to me than the typing methods, so it would be faster. But not as fast as Voice.	If I was in a situation where I would want to take notes, I might be more comfortable using this method.
You have to rely too much on what the phone thinks you are trying to write.	We are all familiar with writing, so it should be straightforward.
I am not familiar with this technology, but it sounds like it would be slow. Also, I am worried about how the oil in my fingers would effect the speed and accuracy of entering text.	It seems like it would be slower and more cumbersome than typing, but it might be more accurate since it is more intentional.
I have bad handwriting, but maybe the phone's autocorrect can help with the interpretation.	If I cannot tell what I've written, how can I expect a machine to figure it out?
It takes a lot of effort because you have to "draw" each character. It is more time consuming than writing with a pencil because you are constrained by the space you have on the screen and that you are using your finger to input text.	I have done handwriting on a touchscreen when I sign my name for credit card purchases at the store. So far, I have noticed that the output does not look like my signature.
It would take more time because you have to physically make each stroke for each letter.	How can a phone recognize what you write? The concept is magical to me.
It should work well, because I am familiar with writing. I do it all of the time!	I write well, so I am confident this will work.
It would be a lot of work. It is a small screen, and I write large and sloppy. I think I would have to change the way that I write to make it accurate.	I am left-handed, so my hand rests on what I've already written. That might make it less accurate on a touch screen.
I make my letters differently (I put dashes through my z's and 7's) so it might not recognize my style.	This would be nice because I would not have to pay attention to locating letters on a keyboard.
It seems like it would be a clumsy method because when you normally write, you have something in your hand. This is the least familiar of anything I have ever tried.	This would be less efficient than the other methods because of the small screen size, large finger size, and that you have to wait for the phone to convert your writing to text output.
If I want to hand write something, I will use a sheet of paper.	Less restricting because you can use more gross motor skills.
I like the way my handwriting looks, but I think that it would not appear very similar on the phone.	I do not dot my i's and I write inconsistently, so the phone would have to learn my writing style.

APPENDIX B (continued)

TABLE B5

PARTICIPANTS' INITIAL IMPRESSIONS OF VOICE

Younger Adults	**Older Adults**
This would be great for driving situations.	I can talk faster than I can write or type.
I would like this if it worked, but "phones being phones" I cannot expect the accuracy to be great.	I can understand how to use this method so it seems to be the simplest and most familiar.
This would be the fastest and easiest.	My daughter has this and she says it works well.
All of the extraneous background noise would affect the output. Plus, what does it do with accents?	If it really does work well, then you can speak into the phone while you are doing other things. It does not require dexterity, only vocabulary.
I do not like seeing people talk to their phones.	I do not enunciate very well.
I have a lot of variability in how I speak, so I do not know how it would work if I say things differently every time.	Since it does not involve physical manipulation, I cannot make any mistakes.
If the software is good enough, it would be easy to get my thoughts out quickly without having to type it.	It seems convenient to say something and have text appear.
This is a cool idea, but the technology probably is not good enough. I have used commands on my phone ("call mom") and it does not do very well. It is annoying having to repeat the same command.	It would be easy to use, but I am self-conscious. It puts me off to hear other people's business, so I do not want to do the same. I might use it at home.
It would be easy to use because it is hands-free, so you can use it when you are busy. It would be less distracting.	My voice is variable depending on the environment, the weather, and even my mood. So it might not be very reliable for me.
It probably is not as accurate as the Qwerty methods, but it should be quick and hands-free, which I like.	I am lazy by nature, and this seems like an easy method to use.
It would be convenient to use when your hands are dirty. Or for people with visual disabilities.	You are limited in the environmental situations where this can be used, especially with background noise.
I have had too many automated conversations with my insurance company and having to repeat myself. Any time there is background noise, the system does not get what I am trying to say.	I want to get away from having to use my hands to interact with machines. But my experience with automated systems has been awful. If you can train the phone to your voice, this might be good.
It might be convenient, but sometimes I do not want other people to hear what I am saying (privacy).	I like to talk, so this would be a good method for me!
I am not sure how it handles words that sound like other words.	I have done some dictating to my computer and I feel pretty comfortable with it.
If I am going to speak, I would rather call the person. I do not trust technology to translate my voice and it seems like an unnatural way to enter text.	This would be great for people who have had hand problems because you would not have to select anything.

APPENDIX C

OPEN ENDED COMMENTS ADDRESSING LIKES, DISLIKES, AND RECOMMENDED IMPROVEMENTS TO THE INPUT METHODS

After entering 20 phrases with each input method, participants were asked to give feedback regarding what they liked, disliked and anything they would change about it. Participant comments for each input method are give in Tables C1 through C15.

Likes/Dislikes/Recommendations Comment Summary

***Physical Qwerty*.** After entering twenty phrases with the Physical Qwerty input method, younger participants reported that they liked the familiarity of the layout, the simplicity of it, and characteristics of the keyboard, including keyboard size, key size and spacing between keys. They also felt that they were in control of the input, and that it was accurate and forgiving when they made mistakes. Participants felt that the keyboard was accessible, as it was not necessary to navigate through a GUI to access it. However, participants also reported that they did not like the key size (too small) or the keyboard size (too large or too small). Some felt that it was slow and that there was too much key resistance.

Younger participants suggested decreasing the resistance of the keys, changing the size of the keyboard (larger or smaller), and to increase the crowning of the keys to make them easier to detect. They also suggested to change the icons printed on the “backspace” and “return” keys so they would be more identifiable, and to change the location of the space bar, as it was difficult for some participants to reach.

Older participants reported that they liked the haptic and audio “click” feedback when the buttons were depressed. They enjoyed the feeling that they had more control over the input method, because they had to be intentional with each key press. The familiar Qwerty layout was helpful in finding the keys, and it seemed to be simple to use. Some liked the size of the keys, and some thought they were too small. Some additional dislikes included the feeling that the input method was slow, that the “click” of the keypad was annoying, and that there was too much key resistance. Others reported that they were disappointed that they were not able to transfer their keyboarding skills to this input method. One suggested that the keyboard materials seemed “cheap”.

The older adults suggested that there should be less key resistance, and to increase the size of the keys and the spacing between them. The accessibility of the space bar was mentioned, and a recommendation was to either increase the size of it or to put one on each side of the keyboard to make it easier to reach. Additionally, some thought placing the “enter” and “delete” keys farther apart would help with accuracy.

Onscreen Qwerty. Younger participants reported they liked that it was quiet and that there was no key resistance. They also stated that they thought the method was “cool” and “modern”, and that they liked the pop-up letter graphics, suggested words and autocorrect features. Some also felt that the key size and spacing was adequate. However, participants also felt that the pop-up graphics and pop-up symbol menus were annoying and distracting, that the lag time for pressing a key and receiving output was too long, and that it was inaccurate. Autocorrect was confusing to some participants because they were unclear whether the mistake was their fault or a software glitch. Participants were confused about the use of an underline in the output, because they tended to relate that to misspellings (such as in a word processor). Some reported that it took too much concentration to operate, and that they had to spend more time attending to the keyboard than to the output. Others expressed that they had a difficult time with the placement of the “shift” key, since they would select that instead of “A”. There were also complaints about the space bar sensitivity, which tended to not be sensitive enough. Participants also reported that they thought the mode of entering text by tapping was too repetitive. Some felt that the key size was too small.

The younger adults recommended adding some audio or haptic feedback, decreasing the sensitivity of the keys but increasing the sensitivity of the space bar, and make the keys larger. They expressed they would like to have a faster processing time for key presses, and an easy way to disable pop-up graphics and additional input modes (e.g., Tracing). They thought that the “backspace” and “enter” keys were too close in proximity, as they would accidentally press one in lieu of the other. Some participants suggested to split the keyboard in half to make the center keys easier to reach, or to make the keys dynamically expand in size so that it is easier to hit the right key.

Older participants reported that they liked the Qwerty layout of the keys, the suggested words that appeared on the screen, that it was easy to correct and fairly fast. They did not like the graphic “pop up” of the letter after it was pushed, the autocorrect feature, the low accuracy, and the ambiguity regarding the amount of pressure they needed to use to get a response. Some felt that the key area was not equally sensitive and that the “hot spot” was right in the middle of the key. The placement of the “enter”, “delete” and “shift” keys was problematic for many, and the space bar created many space insertions or omissions. Some also felt that they had to have constant vigilance and awkward hand positions while operating this input method, which caused them to attend to the keyboard more than the output and made it tiresome. Again, ambiguity on how hard to press the screen was an issue with this audience.

The older adults recommended increasing the size of the keyboard, keys and spacing between the keys, making the space bar responsiveness equal to that of the other keys (some felt it was not as responsive compared to the other keys, and some felt it was too responsive). Some participants felt that accuracy could be improved by introducing a stylus.

Tracing. Once younger participants had finished entering twenty phrases using the Tracing input method, they indicated that they were surprised at the high level of accuracy and speed that they were able to enter text. They had a positive impression of the autocorrect/autocomplete feature and the automatic spacing between words. Some reported that it was very learnable, in that they had memorized gestures for some 3- and 4-letter words they had frequently traced through the trials for this method. Some also said that they thought this was "fancy and cool" and that the gestures felt "fluid". However, participants also reported that they did not like the fact that their hands were obscuring the screen as they traced, making it difficult to locate letters. They also mentioned that it took much more concentration to trace longer words, and that the green trace, which followed their finger while tracing, was messy, annoying and further obscured the keyboard. Some indicated that they thought it was slow, because they were limited to using only one finger to input text, and that constantly using the same finger could lead to hand cramps after awhile. Some even said that their finger started sticking to the screen if it was a little bit sweaty, making the trace "skip", and create and error. They also mentioned they thought the software had a limited vocabulary.

Younger adults suggested increasing the screen and key size, introducing a stylus, and increasing the accuracy of the autocorrect/autocomplete feature. They also thought that the green trace should fade at a different rate (some wanted it to fade faster, some slower). Additionally, some felt that it would be nice to have a tracing tutorial built into the application.

After the older participants had entered twenty phrases using the tracing input method, they reported that they were surprised at how fast and easy it was. They reported that the gestures they used were easy to learn, and that it was forgiving. They liked how words would complete themselves at times, and also liked the green trace that showed them where their finger had been. Some even thought it was fun. However, participants also reported that they disliked having to drag their finger over a the screen; that it was too much movement. Additionally, since the finger remained on the screen, it obscured the keyboard making it more difficult to locate the next character in the phrase. This, in turn, would create other problems because if a participant lifted his/her finger to find the next letter, the software would insert the word closest word match which would usually cause the participant to redo the word. When a participant hesitated too long on the first letter of a word, additional symbol keys would appear in a pop-up menu over the keyboard, which was confusing to most. The space bar would “stick”, and continue spacing if the participant accidentally selected the space bar instead of the intended key at the beginning of a trace. If a trace was not recognized by the software, an audio “beep” was heard, which confused some participants. One participant reported that she thought it meant that she was running out of time. Participants were confused in cases where they made an acceptable and the software retrieved a completely different word. Some felt that they would have needed an instruction manual to be aware of this method and how to use it.

Older adults suggested larger keys, introducing a stylus, and changing the key arrangement from Qwerty to ABC, since it is not similar to typing. They also thought that the “enter” and “delete” keys were too close together.

Handwriting. Younger adults reported that they liked not having to search a keyboard layout to enter text, or to push buttons. They also felt that it was a familiar way to input text and that it was fun to use. They liked having the option to write letters on top of each other, rather than being constrained to write from left to right, and they felt that it would be useful for entering special characters without having to switch to and hunt through different keypads, as is done with the Onscreen Qwerty. They also mentioned that this might be a method well suited for those with poor dexterity or those who were not technologically savvy. However, they also reported that they thought it was awkward and unnatural to use, due to problems they typically do not encounter with natural handwriting. They had to enter a space after each word, change their writing style to increase the accuracy, and occasionally they would run out of space while writing, which caused them to write letters on top of each other or switch to the left side of the screen while in the middle of a word. Some said they felt like they had to press the screen hard to get it to respond. Others said that they had problems with their finger getting sweaty and sticking to the screen. They felt that the feedback, given by a green trace which followed their finger while making a letter, was inconsistent and misleading. The trace would appear to be the “correct” letter, but the software would interpret it as a different character. Left-handed participants also felt that it was awkward to interact with this method, since they were concerned about their hand obscuring what had been “written” and that it might drag across the touch screen. Many participants were confused that the output contained upper and lowercase letters, when they were only aware of entering lowercase characters.

Younger adults recommended the introduction of a stylus, increasing the accuracy of the software, increasing processing speed, and making it more forgiving and adaptable to personal writing styles. They also thought it would be helpful to add an automatic spacing feature, where the software would recognize physical spaces between words on the screen, and to improve autocorrect to fix the insertion of capital letters appearing in the middle of a word.

The older participants liked that their finger size did not seem to effect accuracy, that it was a quiet way to enter text, that they did not have to search through letters on a keyboard, that they did not have to keep their eyes on the input field while they were entering text, and that they did not have to learn a new skill to use it. They also appreciated the novelty of it and that they did not have to be “tech savvy” to use it. Dislikes of this method included the unnatural and unintuitive aspects of the interface, including having to gesture or hit a space bar to insert a space, running out of room while writing, writing letters on top of each other, and uncertainty on how much pressure to use. Additionally, participants felt that there was a lot of discrepancy from the green trace that followed their finger to the letters that appeared in the text field. They thought that the software had registered their writing correctly, and were surprised when the output did not match the input. They also were confused that capital letters would appear in the middle of phrases when they felt they did not make any capital letters. The location of the delete key next to the space bar was problematic, and many had a problem with the sensitivity of the space bar. Some thought this was fatiguing and difficult to use.

The older participants expressed the wish for better accuracy, and suggested introducing automatic spacing between words, and to decrease processing time. Participants wanted the interface to be more forgiving and accepting of their own personal writing style. Others suggested that a manual should be included to instruct them on how to write their letters.

Voice. Younger participants thought it was fast, straightforward and accurate. They said that it was efficient and would be something they could use while multitasking. They appreciated that they achieved good results without having to change their speech. They also commented that the technology was much better than expected and that voice technology would be beneficial for those who are poor spellers. However, they also expressed that the software launch button was too small, unnoticeable, and that having a soft key to launch the application would take away from the hands-free benefit that was appealing to many. They also felt that the processing time took too long, that the endings of the words were less accurate, and that substitution of homophones and similar words (i.e., “hey” for “hello) was an issue. Some participants reported they felt they had to slow their rate of speech, or enunciate more clearly than they would if they were speaking naturally. Many also expressed situational and social concerns, regarding background noise, looking “weird” talking to the phone, and privacy of their messages. They also felt that erasing the entire message and starting over would be the easiest method for correcting errors.

The younger adults suggested improving the accessibility of the software launch button by making it larger, more noticeable, or to make it a hardware or voice-controlled feature. Others also suggested changing the icon on the button to something more representative of “voice” or “speech” recognition, instead of a flame. They also thought that receiving dynamic output while speaking or having a time feature which displayed the amount of time you had been talking would be useful. Some thought that being able to use this method without network access would be helpful, and that allowing to train it to their voice would increase its accuracy.

Older participants appreciated its ease of use and speed, and that they did not have to access a keyboard to use it. Many were surprised at its accuracy, and were glad that they did not have to change their speaking style to successfully interact with it. They liked that it was hands-free and accessible, and that it had a sort processing time. However, participants also felt that it took some trial and error to discover how the software “wanted” them to speak. Some reported that they felt they had to enunciate more clearly, and that word endings and homophones were less accurately processed. Some reported problems with the interactivity, saying that they forgot to press the button before they started speaking, and some felt that there were too many steps required to enter a message. They also mentioned concerns with the small size of the software launch button and how closely it was located to other buttons, making it easy to instead press the mode shift button or the comma button. Many said that they would not want to use this in public because of privacy concerns and not wanting to disturb others.

The older adults recommended enlarging the size of the software launch button, changing its location, making it more noticeable, changing the icon on the button to something more representative of “voice”, and possibly making it a hardware feature. They thought that it might be helpful to streamline the process of entering text into fewer steps. Other suggestions were to have a customizable vocabulary (such as slang or acronyms) and including a “pick list” for homophones.

TABLE C1

PARTICIPANTS’ “LIKES” FOR THE PHYSICAL QWERTY

Younger Adults	**Older Adults**
I felt in control. I did not have to guess whether the software was going to get it right or not. I was able to directly select the letters I wanted.	I knew where to find the keys. You just had control where you placed your finger, not rely on the sensitivity of software or a touch screen.
I can feel the button and knew when I pushed it. I knew when my finger was in the middle of the key.	I was surprised! I thought that I would have more trouble since the keyboard was small.
The buttons were easy to hit, so you could not accidentally brush up against something and hit the wrong letter. It seemed to be accurate.	I liked the familiar layout and the sound (clicking) of the keypress. It was similar to using a typewriter.
If I made a mistake, I knew it was because it was something I did, not the software’s fault.	I liked that you had to enter text intentionally with more than just a tap of your finger.
I felt very confident using this. I liked the familiarity of the layout, the accuracy and that it was easy to correct mistakes.	I liked being able to see what I was doing without my hand obscuring the keyboard. I can see why people use their thumbs, because it would be easier than taking your finger to punch each individual letter.
It was straightforward and simple. It did not require navigating through a GUI to select the input method.	I knew exactly where to touch, having physical buttons to depress.
I liked the key spacing, key size and that they were raised in the middle. They were easy to find, and I could tell which button I was hitting.	I liked the simplicity in the operation of this keyboard.
It is more accurate because you can click a solid button rather than rely on the phone recognizing your touch.	I could use my fingernails with this method!
	I liked the size and resistance of the keys.

APPENDIX C (continued)

TABLE C2

PARTICIPANTS' "DISLIKES" FOR THE PHYSICAL QWERTY

Younger Adults	**Older Adults**
After awhile, it hurts your thumbs and wrists.	The keys seemed small for my finger size, so I hit some of the wrong keys.
The keyboard was too big for my hand size. I felt like I had a hard time reaching the center keys.	I felt confined because of the small keyboard size.
You have to have both hands and watch the keyboard.	The keyboard is too small, which make it slow to select the keys without error.
It felt like the buttons were shifted to the left, so I felt like I was reaching with my right hand a lot.	I did not like the "clickiness" of the keypad. The sound was annoying, like cheap plastic.
The buttons were difficult to push, but this might have helped increase accuracy.	When you learn to type, you use all of your fingers. Now I feel like I have to forget my keyboarding skills!
It was uncomfortable to use and felt like it was slow. I am not used to typing on a keyboard this small, and it felt like it took more force to operate this compared to a regular (desktop) keyboard.	I had to press harder than I normally would to select the keys. I can see how a person could get fatigued using this method. Especially older adults with arthritis.
The keypad is too small and the keys are too close together. The keys were also small.	The size of the keys compared to the size of my finger!

TABLE C3

PARTICIPANTS' RECOMMENDED CHANGES FOR THE PHYSICAL QWERTY

Younger Adults	**Older Adults**
Decrease the resistance of the keys.	Decrease the resistance of the keys.
Make the keys and keyboard larger.	The keys are too small and close together.
Make the keypad smaller, more compact so I can use it one-handed.	It is impressive that you can get a full keyboard on a device this small, but increase the size of the keys.
Make the buttons more raised so I can feel them.	Increase the quality of materials used for the keypad.
The symbols on the backspace and return keys were too similar, so I got them mixed up at times.	Put two space bars on the phone or increase the width of the space bar.
The location of the space bar was too far down.	Move the enter and delete keys farther apart.

APPENDIX C (continued)

TABLE C4

PARTICIPANTS' "LIKES" FOR THE ONSCREEN QWERTY

Younger Adults	Older Adults
It was easy to touch and there was no audio clicking. It annoys me to hear people typing on their phone.	I had more control over this method than the tracing method.
It seems like it has the potential to be faster than the physical keyboard because I can move my fingers around the keyboard more quickly.	I can type using this method the same way that I can type with the physical Qwerty (computer and phone).
The autocorrect feature seemed to get the right words.	It seems similar to my iPad.
I liked that it was quiet while I was typing.	Easier to do double letters on this method versus the tracing method (double tap versus a circle gesture).
It is cool and modern; high tech.	I like the idea of the suggested words on top.
Straightforward to use and easy to access.	Easy to make corrections.
I liked the spacing between and the size of keys.	It seemed to be fairly fast.
I like how the letters pop up after I select them.	
I could go fast because the screen was responsive.	
I like that it shows options for word spelling.	

TABLE C5

PARTICIPANTS' "DISLIKES" FOR THE ONSCREEN QWERTY

Younger Adults	Older Adults
The device could not keep up with my typing speed. It seemed like the faster I typed, the more inaccurate the words were. I could not feel whether I made a mistake when I was typing, because I could not feel the keys, so I could not tell whether the mistakes were my fault.	I liked the pop up letters, but it was also a little distracting. I wanted to pay attention to what I typed, but I kept being distracted by the letter I just pressed.
Every time I hit a key, it was not the right key.	I would like it better if it didn't have autocorrect.
There was a lag time between pressing the key and getting feedback on the screen.	It took me awhile to figure out how much pressure was necessary to hit a key.
The pop up letters. It was distracting. It made me focus more on where the letters were and not what I had written.	It seemed like the hotspots on the keys were right in the center of the key. You had to be very accurate.

APPENDIX C (continued)

TABLE C5 (continued)

Younger Adults	**Older Adults**
It seemed to be very repetitive, which might lead to cumulative injuries if you use it a lot.	The keyboard was too small for my fingers, and it was too sensitive.
Buttons on the screen are small and close together and it is easy to hit the wrong button.	The space bar was not registering when I pressed it.
The size of the screen is tiny.	It seemed that I had to go very slow to be precise.
I was confused about the underlined words because usually you see this when something is misspelled in your word processor.	I felt very clumsy and awkward with this. I kept pressing the return key and having spacing issues. The shift key was too close to the "A" key. I am familiar with typing, but this keyboard made me feel that I did not know what I was doing.
The space bar is not very sensitive, it seemed to stick.	Everything was too small and close together.
Other functions would pop up on the screen while I was trying to type.	The enter and delete keys are too close together.
You have to find every single key because you do not have the tactile feedback, so you have to visually locate the key you want to press. No physical buttons and orientation aids (like home row markers).	The keyboard was imprecise. I thought I was pressing an "n" and I got an "m". When I thought I touched the space bar, either it would not move at all or it would move too much.
I could not tell if I skipped a letter when I was typing because I could not feel anything.	I spent too much time watching the keyboard rather than what I had actually hit.
This method would deter me from writing long messages.	My finger was over extended while I was typing, and it got tiring after awhile.
You have to look at the output to know whether you have hit the correct keys.	It is like seeing a fish in a pond. You think you know where he is, but he is really in another location. I felt that way with the keys on this keyboard. I thought I pushed the right key, but really I hit something else.
I kept hitting the shift key instead of the "A". My thumbs kept getting in the way.	This method took constant vigilance to use.
The accuracy is bad. I was tapping letters next to the ones I intended to hit. Too sensitive.	The frustrating part was figuring out how much force to use. When I realized I just needed to barely touch the keys, it became more accurate.
It seemed like the phone completely changed the word I was typing at times. Is this a software problem?	

APPENDIX C (continued)

TABLE C6

PARTICIPANTS' RECOMMENDED CHANGES FOR THE ONSCREEN QWERTY

Younger Adults	Older Adults
Add some kind of audio or haptic feedback.	Bigger keys and a larger keyboard.
Decrease the sensitivity of the keys, but increase the sensitivity of the space bar.	An instruction manual on how to use it.
Split the keyboard in half so it is easier to reach with both thumbs. I had a problem reaching letters in the center of the keyboard.	Slow the response from after you touch the key to when the letter pops up on the screen so you can see which one you pressed more easily.
Make the keys larger.	Have built in phrases that I use frequently (quick text).
Move the backspace and enter keys so that they are father apart.	Make the space bar more sensitive. It seems like the sensitivity is different for the space bar versus the other letters, so I do not know what to do.
I had a problem with finger oils, so if there is some material that would help alleviate this problem…	Increase the size of the smartphone.
Have an easy way to disable additional modes on the keyboard (i.e., voice recognition, handwriting, tracing)	I would rather use a stylus instead of my finger.
Disable key pop ups. They are too distracting.	Make it go away.
Make it to where the buttons could dynamically be expanded in size or the spacing between keys could be increased so that it is easier to hit the right button.	
Reduce lag time from keypress to when it is displayed.	

APPENDIX C (continued)

TABLE C7

PARTICIPANTS' "LIKES" FOR TRACING

Younger Adults	**Older Adults**
It was very accurate. I liked how quickly the software "caught on" to what I was trying to spell.	It was easy to trace from one letter to the next. I liked the line indicating where I had been.
Even if I did not trace over the right letter, it knew what I was trying to say.	Words completed themselves if I lifted my finger.
The automatic spacing was a plus! I did not have to stop what I was doing to press the space bar.	It is nice to not have to lift your finger to search for the next key.
Relatively easy to use once I got used to it.	The gestures were very easy to learn.
Much faster than I thought it would be initially.	It is quiet.
Everything! I liked the accuracy and it was easy to find the letters.	There is a nice fluidity to the movement, like cursive handwriting and doodling.
This might be good for people with disabilities or hand problems.	I started to learn patterns for words that I used frequently, rather than having to hunt for individual letters within the word. Very convenient.
It was a very fluid way to interact.	It was fun to use!
Very quick and easy to use.	It seemed faster than both of the Qwerty methods.
I was able to remember the gestures for the common 3- or 4-letter words by the end of the task.	It was very forgiving. Sometimes my finger was in the wrong location, but it was still accurate.
It was surprisingly accurate. Even when you screw up. It is smarter than I am!	Even if you did more than 2 loops to get a double letter, it would give you the right word.
It is kind of like connecting the dots. Really easy to use and really easy to learn.	It was much quicker and easier to learn than the onscreen Qwerty.
Using gestures looks fancy and cool.	I was intentionally trying to mess up a couple of words, and it still got it right.

APPENDIX C (continued)

TABLE C8

PARTICIPANTS' "DISLIKES" FOR TRACING

Younger Adults	**Older Adults**
I wish I could have moved my finger to see where the letters were. My finger and hand were in the way.	I did not like the gesture for double letters. It was difficult to keep track of how many times I "looped".
Tracing larger words required more thought.	I did not like dragging my finger.
This method felt like learning another language by mapping gestures to words. It was unsettling.	If you did not have someone show you how to use it, you would not know how to do it.
It takes a lot of concentration.	I felt like my finger was always in the way.
The amount of time I had to keep my finger on the screen was too long. I started to get a hand cramp, so I switched to a different finger.	Stressful. I felt tense while I was using it because I could not do the task. I kept hitting the return key instead of the delete key.
I would not be able to use this one while I am driving.	It is not clear how to get to other types of characters or punctuation.
Even though you traced over the right letters, sometimes it would give you a different word.	I was not very confident. I kept wondering what words were going to be retrieved. I was constantly dealing with uncertainty.
Sometimes the green line was covering the letters you were looking for.	Tracing longer words is cumbersome. You lose track of where you are at in the word while you are tracing.
The green line was annoying and made the screen look messy.	The phone kept entering the wrong words.
It is slow having to drag my finger from one side of the screen to the other.	The number and function keys are accessed differently, and I was confused when they would pop up on the screen.
Sometimes you get lost when you are trying to think about the next letter to type, and you forget what you already typed.	I did not like that my hand obscured the screen. If I did not know where a letter was, I would lift up my finger and then I would have to start all over again.
It took awhile to get used to not having to press the space bar, which is normal behavior.	When a word popped up that I did not trace, I was really confused on what I did wrong.
You can only use one hand, so it is slower.	The space bar kept getting stuck!
My finger kept sticking to the screen when I was tracing.	I thought the "beep" noise meant I was running out of time.
Limited vocabulary within the software.	Having to remember where the letters are located.

APPENDIX C (continued)

TABLE C9

PARTICIPANTS' RECOMMENDED CHANGES FOR TRACING

Younger Adults	**Older Adults**
Make the screen size larger and the keys bigger.	Make the keys larger.
The software (autocorrect) could be more accurate.	A stylus would help me use this method better.
I would like to use a stylus instead of my finger.	The delete and enter keys are too close together.
Have a mini-tutorial of tips while using this method for people who are not familiar with it, like a button or a question mark icon or something.	Make the keys larger so that your fingers do not mask the keys, which would increase the visibility of letters you are looking for.
It would be nice if the trace line would fade faster.	This is very different from typing, so maybe a different arrangement of letters, like "ABC" rather than Qwerty would be helpful.
I would like it if the green line stayed on the screen longer. Especially for novice users.	Maybe you could have a "trace" area below and an image of the keyboard above it, so you could do the tracing, but watch which letters you select elsewhere. That way your hand wouldn't be obscuring the letters.

APPENDIX C (continued)

TABLE C10

PARTICIPANTS' "LIKES" FOR HANDWRITING

Younger Adults	**Older Adults**
There are no keys to search for or push.	I liked that my finger size did not matter!
You did not have to look at the screen while you were using it.	You do not have to be technology savvy to use this.
It seemed like it went fairly quick.	It is quiet.
It might be helpful when writing letters that have accent marks, because then you would not have to switch to another keyboard.	Seemed magical that you can write on a screen and that it would recognize it, and that it was not dependent on my penmenship.
Familiar, and it seemed pretty straightforward.	It feels like playing a game.
Everyone knows how to write, and it was not difficult.	I did not have to search the keyboard for letters.
I liked that I could write letters on top of each other.	Aesthetically pleasing. It is entertaining and gives you an artistic visual.
I liked the green trace following my finger, which helped me know that the software registered my finger.	I felt more in control of the interface. I did not always have to look at the screen to enter text.
This might be easier to use for people that do not have good dexterity, or for people who are not tech savvy.	I already know how to write, so I did not have to learn a new skill to input text.
It was fun to use my finger!	I liked the novelty of it.

APPENDIX C (continued)

TABLE C11

PARTICIPANTS' "DISLIKES" FOR HANDWRITING

Younger Adults	**Older Adults**
It was very awkward. You have your own style of writing and it did not recognize a lot of my letters.	It was unintuitive to write letters on top of each other, but I had to do this at times because I did not have enough space to finish the word.
It did not feel natural when I ran out of room on the screen.	Sometimes I would have a word or a letter in there and the software would delete it.
It did not recognize my letters. Sometimes it inserted uppercase letters when I wrote in lowercase.	Normally I dot my j's and i's at the end of a word, but I could not do that here. It would insert punctuation.
Very time consuming to use.	My finger did not drag across the screen smoothly.
I felt that, as a left-handed person, I had to change the way I wrote due to the touch screen.	Noticeably slower to recognize letters; a lot of lag time while waiting for words to process.
It was not natural to stop and hit the space bar.	I was uncertain how much pressure I should use.
I had to write letters one at a time to make sure they were being interpreted correctly. I did not feel comfortable writing the entire word at once.	There was a problem if you wrote above or below the text entry space. It was hard to keep track of where you were writing.
I felt like I had to press the screen hard.	The space bar is not sensitive enough.
I had to remember to write each letter separately. Normally I tie a lot of them together.	Frustrating. The trace showed me what I wrote, but then some other letters would pop up on the screen. The feedback was inconsistent with what I expected.
The green trace would show what I wrote, and it looked like what I intended, but then the word would appear on the screen and be a different word.	It did not feel natural because usually when I write, I use a pen, and if I make a mistake, I can write over it to correct it. Here I cannot do that. I kept feeling like I wanted to erase what I had just done.
I had to write more deliberately to increase accuracy	This was fatiguing to use.
I felt very disconnected from the writing process because I had to make letters individually.	I could not get my finger to fit in the space to write very well because my fingers are large.
I felt like I had to exaggerate characteristics of the letters to increase accuracy.	Having to hit the space key was unintuitive. You should just be able to leave space between words.
I felt my finger start to stick to the screen.	This was very difficult to do.
The backspace button is too close to the space bar.	I had to write in uppercase so it would recognize it.
I would not have figured it out on my own!	I could not figure out why I kept getting capital letters.

APPENDIX C (continued)

TABLE C12

PARTICIPANTS' RECOMMENDED CHANGES FOR HANDWRITING

Younger Adults	**Older Adults**
I want it to recognize my words faster.	It needs to have better accuracy.
Accuracy.	It needs to accept cursive writing.
Autocorrect should catch instances when capital letters are inserted in the middle of a word.	A stylus might help the user be more precise, but it would be another piece of equipment to lug around.
A stylus might help.	Automatic spacing between words.
Add guidelines, like we have when we first learn to write. That way you know where to put the letter.	Improve the feedback of the trace. The line reminds me of what you would expect from an Etch-A-Sketch.
Automatic spacing would be better.	It should be more forgiving.
The software should be more forgiving of our own personal style of writing.	Instructions on how to make your letters and how to operate the software.
Maybe have some initial setup program where each person writes the alphabet. The phone might be able to learn your handwriting.	Everything needs to be better. It seems disarmingly simple, but is completely unpredictable.
The software should accept ligatures in words.	Decrease processing time. The feedback should be immediate instead of waiting word by word.
Is it necessary to translate into text? Maybe this could just be a handwriting program which allows you to send notes to people as an image rather than as text.	It is beyond help.
The software should recognize a physical space between words as a space rather than having to use the space bar or some other gesture.	

APPENDIX C (continued)

TABLE C13

PARTICIPANTS' "LIKES" FOR VOICE

Younger Adults	**Older Adults**
Very fast!	Very easy to use and fast.
Easy to use and pretty accurate.	Being free from the keyboard is wonderful!
This could be a very efficient method, which is desirable. Great for multitasking.	Once you get accustomed to what you need to push, and when, it is easy to use.
I would feel safer sending an email while I was driving using this method than anything hands-on.	All you have to do is speak distinctly and it picked it right up. I did not have to change the way I speak.
The speech software works so much better than it used to, this is a viable input method now.	For those who are handicapped, this could be a new arena of communication. This is better than typing.
This is a straightforward way to interact, just one button press and say what you want to say.	I just spoke and the phrase popped up. I did not have to do anything mechanical to it.
I thought I would have to talk like a robot, but I did not. I could talk at my regular pace and it could keep up. It is so quick and had fast processing time.	Once you master this, it is the simplest and easiest to use. This is the fastest of all of the input methods, after all, you can say a phrase faster than you can type.
This is great for people who are not good at spelling!	Very short processing time.
It even hyphenated an adjective-adverb.	Much better than I expected!
I would not have thought to use this feature on my own, but now that I am aware of it and was pleasantly surprised with the results, I would use it.	I liked that this was hands-free; accessible.
I did not have to speak any louder than normal to get accurate results.	I liked the graphic that showed that it was processing.

APPENDIX C (continued)

TABLE C14

PARTICIPANTS' "DISLIKES" FOR VOICE

Younger Adults	**Older Adults**
It seems difficult to fix words in the middle of the message. It would be easier to just erase the whole thing and start over.	It is very sensitive, because when I would exhale at the end of the phrase it would tack some other word on to the end of the phrase.
I had difficulty selecting the software launch button. Sometimes I hit a different button.	You have to figure out how clearly you have to speak.
There were a few words that I said that I could not get it to recognize, especially homonyms.	The launch button was too close to other buttons. I kept pressing the comma key.
I would not like to speak to my phone in public. No privacy.	I expected the words to come up on the screen dynamically while I was speaking.
It seems the endings of words were wrong at times.	The button was too small. I had to use my pinky.
I had a hard time finding the button to launch the application. It was small and kind of hidden. And if you want to access it hands-free, that kind of defeats the purpose.	It seemed like a lot of steps to this method: hitting the entry field, the flame icon, the done button, and then waiting for it to process. It was kind of complicated.
Sometimes it will substitute words, like "hey" for "hello". But those are typical conversation starters, so maybe that is where the phone gets it from.	You had to be more conscious about what you were saying. I would not want to use this in public, where I could distract others.
I felt like I had to slow down my speech and enunciate more to get it to be accurate.	It seemed like I had to enunciate more clearly for the end of words, and I always had to check the accuracy of the output before I sent it.
The processing time took awhile.	Sometimes I forgot to press the "record" button.
You would have to play around with it to see where the best environments are to use it and how loud you have to speak, and how you have to hold the phone.	This would be problematic as a standalone method. It doesn't handle homophones well. For example, instead of "irregular" it put "your regular".

APPENDIX C (continued)

TABLE C15

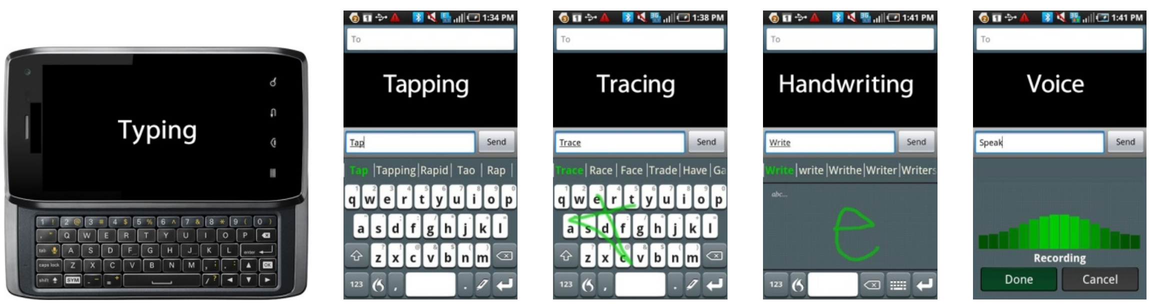
PARTICIPANTS' SUGGESTED IMPROVEMENTS FOR VOICE

Younger Adults	**Older Adults**
Make it easier to activate the software, like using a physical button, rather than having to use a key on the touchscreen, which defeats the purpose.	If there was a way to streamline the process so there were not so many steps, that would help.
Initiate the software with a voice command.	A stylus might help with hitting the launch button.
Get immediate feedback of what you are saying on the screen as you are talking.	It would be neat if it accustomed to my vocabulary (slang, acronyms, etc.) that I use at my job.
It would be nice to do this without network access.	Accuracy of the output could be improved.
Maybe a stronger microphone could pick up word endings more accurately.	Having an easier way to edit different areas of the message. It seems easier just to erase the whole thing.
Setting it up to train to a person's voice might make it more accurate.	It would be nice if you did not have to push any buttons at all, that way it would be hands-free.
Change the flame icon to something more intuitive; something that makes me think of "voice recognition".	I would like an audio cue of when to start talking.
I would like to have a bigger button to launch the application.	Having a "pick list" for homophones might be nice.
The location of the enter and delete keys are too close together.	I would rather have a physical button or a textured surface on the touch screen to tell me where to push without looking at it.
It might be nice to see a graphic showing how long you had been talking (in seconds).	Enlarge the keys.
Make the launch button more noticeable.	
The way it handles numbers. It does not put spacing between each number, but you might want it to.	

APPENDIX D

OPEN ENDED COMMENTS ADDRESSING IMPRESSIONS
OF THE INPUT METHODS POST-TEST

After they had experienced entering 20 phrases on all of the input methods, participants were asked to give their final impressions and preference of the input methods on a 0 – 50 scale (see Figure B1). Their ratings were based on the desirability and value of each input method for entering text on a smartphone device. Participants were allowed to use the input methods at this time if they felt they needed to refresh their memory, and they were given laminated cards with a screen shot of each of the input methods printed on one side (see Figures D1).



Figures D1. The screen shots used to rate preference, post-test.

They were asked to place the cards along the scale, and to give comments about why they assigned the input methods each rating. Tables D1 through D5 give summaries of the users' initial impressions for each of the input methods.

Post-Test Comments Summary

Physical Qwerty. Final comments gathered post-test indicated that younger participants generally liked the Physical Qwerty keyboard. They appreciated the familiarity of the layout and the comfort of the keyboard. They also liked that they could use two thumbs to input text, which made them feel more efficient. The tactile quality of the keys was useful for increasing accuracy, and also was expressed for being valuable in instances of multitasking. They felt that the keyboard was forgiving, but that it could take effort to operate due to the resistance of the keys. Some thought that it was slow for entering text.

Post-test impressions reported by older adults were, overall, positive. The familiarity of the keyboard layout and the physical keys added to the comfort of use for this input method. Participants felt that they were slower with this input method, but that it had the advantage of accuracy, which was a worthwhile trade-off. Some participants felt that they did not do well with this method, and most of this was attributed to the small key size.

Onscreen Qwerty. The younger adults' final impressions of Onscreen Qwerty were generally negative. While participants liked the decreased physical demands of this method and some thought it was fast, others reported they were disappointed that this input method failed to meet their expectations. They felt they were inaccurate using this method, due to small keys, a lack of tactile feedback, and the sensitivity of the screen. They said that the autocorrect feature was not helpful and that it took a long time to fix mistakes.

Older adults also had many negative comments to report for this input method. Participants felt this method was slow, and took a lot of energy and concentration. They felt that their finger size was large enough to impede their accuracy with this input method. Most felt the keys were too small, and that interacting with it using a touch screen made it more difficult to determine where to press. A few participants, however, did feel that this method went smoothly and that it was easy to use.

Tracing. Younger participants' impressions of the Tracing input method post-test were generally positive. They indicated that it was not as physically demanding as the Physical

Qwerty, that it was fast comfortable, and easy to use. They also thought that it was accurate and that the software was sophisticated, especially with the implementation of automatic spacing. However, some also mentioned that they did not enjoy holding their finger down on the screen for an extended period of time, that it took concentration to operate and that the gestures were not entirely intuitive. They disliked that some gestures were not recognized by the software.

For the older adults, final impressions of tracing were mixed. Some felt it was very accurate, easy, and quick. They liked the novelty of this method and thought that it was fun to use. Some felt that they also did not have to be as precise with this method. Others felt that it was frustrating, stressful and that it required too much movement.

Handwriting. Younger adults' final impressions of the Handwriting input method were negative. They reported that it was not very accurate, cumbersome, slow, and not practical. Participants felt that they had to change their writing style and that they had to be very deliberate about creating each letter form. They did disliked that the software would randomly insert capital letters into words, when they felt that they were making all of their characters in lowercase. Some even reported that they would stop texting if this was the only method available.

Older participants liked that they did not have to look at the screen while they were entering text and that it was entertaining to use. However, most felt that this input method was frustrating, clumsy, slow, tiring and that they had to adjust their handwriting style to accommodate the software.

Voice. Younger participants generally had positive final comments for the Voice input method, but they expressed they would only use it in certain situations. They felt that it was fast, easy, simple, reliable and convenient. However, they reported they would only use it in situations where they were trying to multitask (e.g., driving), when they were in a quiet environment, and not around other people. They also said that it would be useful for composing longer messages, but that it would take more time to use this method, compared to the others, to compose shorter messages. They felt that although this method was relatively hands-free and eyes-free, the usefulness diminished slightly because they would have to proofread the message.

During the post-test comments, older participants expressed that the Voice input method was quick and effortless. They liked the accuracy of the input method, that it was hands-free, there was no keyboard to manipulate. They thought this method would be especially helpful for composing longer messages. Some expressed privacy concerns and that it could not be used as a standalone method.

APPENDIX D (continued)

TABLE D1

PARTICIPANTS' IMPRESSIONS OF PHYSICAL QWERTY POST-TEST

Younger Adults	**Older Adults**
I am most familiar with this layout, so it was the easiest to pick up.	The key size made this more laborious.
I can use both thumbs, so it seems to go faster and I feel like I am getting more text entry accomplished.	I am familiar with typing on a keyboard, but the small size was awkward to use.
I am most comfortable with this method, and it worked the best. It was easiest to fix mistakes.	I am not a big fan of the clicking of the keys, but it is accurate.
The Physical Qwerty was the most accurate.	I am familiar with the layout, so it was comfortable.
I like the physical keyboard. I feel like I have more control and that there is less software interpretation.	I liked having physical keys so I knew where to put my finger.
Even if the Onscreen Qwerty was more fun, what is the point if you are not very accurate? This was more accurate, and you could feel the keys.	I felt it was accurate. It may have been a little slower to use, but because I did not have to make as many corrections, it was worth it.
I felt more confident using this input method, though it seemed like it was more time consuming and took more effort to push the buttons.	This input method was in the middle of the road for me. It is accurate, but it is old hat. It is more reliable than Onscreen Qwerty and Handwriting.
It is less comfortable to type with my thumbs. I might not have enough dexterity. I felt like I was fumbling to hit the letters.	It is very much like a keyboard. It has resistance and you can easily see the keyboard because my hands are not obscuring the keys.
I value that I could divert my attention to other things and still know where the keys were located. This would be great for multitasking.	It was frustrating and stressful to use. I made many terrible mistakes.
I am kind of sick of typing. I have a numeric keypad now and I do not like physical keys.	I like having physical keys over a touch screen.
I can type without having to look at the screen, but I would prefer to have larger keys.	
I liked the tactile feedback. I could tell if I accidentally pushed two buttons at once.	
It was a little slower than I expected it to be.	
I like the raised keys. After awhile, I learned where the keys were located. You could learn to use this method without looking faster than the others.	

APPENDIX D (continued)

TABLE D2

PARTICIPANTS’ IMPRESSIONS OF ONSCREEN QWERTY POST-TEST

Younger Adults	**Older Adults**
It does not have the physical demands of the physical Qwerty keyboard.	The key size made this method more laborious and frustrating.
I hated it. I thought I would like it the best, but the buttons were too small and mistakes took forever to fix. Autocorrect was also bad.	It was annoying because my fat fingers were touching too many keys at once.
It was fairly fast and easy to use, but not as accurate as Voice or Physical Qwerty.	It was slow and I made too many mistakes.
I do not like the onscreen keyboard.	I had to use one finger, which made me feel restricted.
The keyboard seems to be more packed with keys than Physical Qwerty.	My fingers would hit unintended keys, which was frustrating.
Pretty straightforward, but I kept accidentally pushing buttons.	I was able to work with this method easily.
I liked using the touch screen, and liked that it had suggested text so I knew how to spell the word. The keyboard was comfortable to use.	It was hard getting my finger on the right key. The keys are too small. It reminds me of looking at a fish in the pond; he isn’t where you think he is.
This is the quickest way for me to input text.	This was less accurate than I thought it would be.
I like the familiarity of the layout, but at times I was not sure whether the phone registered my touch and the pop up letters were distracting.	It took a lot of energy and concentration, and I had a lot of errors. I would not use this one.
I did not like this one. It was difficult to get a sense of where my fingers were and which buttons I was pressing. I had to do a lot of corrections. Texting is supposed to make life easier, but this was more labor intense. It took focus away from my sentiment and drew attention to the device, which I did not appreciate.	It seemed to go smooth, even though you had to tap out each letter. I thought it was smoother than tracing.
I made a lot of mistakes, but I would get better with practice.	I struggled with how hard I needed to tap to get it to respond.
I thought I would like it more because of the familiarity of the layout, but not having tactile feedback and knowing which button I pressed made me go slower.	

APPENDIX D (continued)

TABLE D3

PARTICIPANTS' IMPRESSIONS OF TRACING POST-TEST

Younger Adults	Older Adults
I preferred this the most because it does not have the physical demands of a keyboard, and it is more accurate than typing.	I thought it was very accurate and almost as quick as the Voice input method.
It was nice not having to space between words, but it was difficult to correct an error when you made one. If you picked up your finger too soon, it was an error.	I could not see the keys because I had to keep my finger on it, or else it would make another word. The keyboard was too small.
I did not think it was very accurate and it took concentration to ensure you trace over the right letters.	For some reason, it seemed really easy. Maybe it just was not as tiring. It was very novel.
It did a good job interpreting what I intended and it allows for a very fast pace of entering text, if you know your way around the Qwerty layout.	It was frustrating and stressful to use. I could not see the letters. I would rather just call the person.
Even though this was more accurate than the Onscreen Qwerty, it took longer.	It was easy and quick and I like the automatic spacing. I just wish my hand did not cover the keyboard.
It is annoying when gestures used to create a word are not recognized by the software.	It was fun to use and I liked the green line.
It is much easier than I thought it would be.	This was far more accurate than I thought it would be.
This was very quick and does not require much effort.	It is faster and less effort than Handwriting.
I did not like that I had to constantly have my finger down, and that I had to hold the phone to use it.	Enjoyable to use! I did not have to aim as much, just had to slide my finger which was kind of lazy.
It was really slow and took a lot of thinking power. Spelling a word fast is easy, but having to remember how to spell things more slowly is difficult.	There were many opportunities for errors, but it seems easy to learn. You do not need the coordination of all of your fingers to use it, just one. The only problem is when your finger is in the way and you cannot see the screen.
I like how it compensated for my errors. It is easy, comfortable to use and fast.	This method required too much movement.
This was really accurate, even when I was intentionally sloppy. It is hard to mess up with this.	Awkward to go across the screen and back and forth and up and down keeping your finger on the screen.
I thought it was awesome. I am impressed at how easy it is to use. It seemed to be sophisticated at picking up what you were trying to say.	Fairly easy and did not seem to take too much energy.
It is easy to do, but the gestures I had to learn to use were not completely intuitive.	
Very accurate and quick. I would definitely use this!	

APPENDIX D (continued)

TABLE D4

PARTICIPANTS' IMPRESSIONS OF HANDWRITING POST-TEST

Younger Adults	**Older Adults**
It was not very accurate and cumbersome to use. It was new and fun, but not very practical.	I liked how easy it was to use. I just had to adjust the way I write to get it to recognize my letters.
It would capitalize things when it should not, and it did not recognize some of my letters.	It was easy to use and I did not have to look at the phone to enter text.
It was not accurate and it was the slowest.	It made me feel clumsy.
It was very difficult to get the software to interpret your writing correctly. Also, you would probably have to shift to another method to make dashes and underscores and things, so if that is the case, why even use this method?	I had difficulty. I had to try over and over to get the correct letters, and in the process I had to learn what the computer wanted, which was not normally what I would write. I got better, but I did not like it.
I did not like it. It took too much time.	It was not too bad.
It is bothersome that it does not recognize my writing.	It was frustrating. I would rather call.
This was fairly easy to use, but I had to be more deliberate. Usually I tie some of my letters together, and I could not do that with this. I had to change the way I write.	This was a real struggle to get through. It picked up a lot of errors, and after awhile I just decided to live with them instead of going through the cumbersome process of correcting them.
This was a waste of time. I had to hold my hand in an awkward position to use this because my nail was getting in the way. Also, people type because it is supposed to be faster than handwriting and this was even slower than regular handwriting.	Being left-handed, I had to drag my hand over the machine, so it was more difficult to use.
I am not keen on changing my writing style to match what the software expects.	I did not like it because it was tiring for my hand. It was awkward to form my letters slowly.
It would be alright for certain situations.	Once I learned what it wanted, it was ok.
It is a cool feature, but it is slow and not very useful.	It liked it. It was entertaining to use and the green trace was pretty. I liked seeing how I made my letters.
It is slow, inaccurate and it almost upset me a bit. It is so against what you are used to (writing left to write, top to bottom). Here you have to break a word and start over on the other side of the screen, write letters on top of each other. You write with your finger in a little space. The feedback (green trace) that I got made me feel like I was writing like a Kindergartner, so I felt bad about how I was making my letters. I would stop texting if this was the only way to text.	

APPENDIX D (continued)

TABLE D5

PARTICIPANTS' IMPRESSIONS OF VOICE POST-TEST

Younger Adults	**Older Adults**
I would not use this very much because I do not have a need for it. If I am going to talk, I will just call.	I am surprised it picked up my voice because I am soft spoken. It was very quick in terms of feedback.
It is easy to use, but I am not sure how it handles accents.	I could get used to doing this easily and in some circumstances, like long messages, it would be faster.
This is the easiest method to use, and you can use it while you do other things.	This was fantastic because you can just speak and it was very accurate. Much less effort and no keyboard!
The software was very accurate, but it seems like it would be awkward to do in public places. I do not want to dictate to the phone like it is my secretary and I do not want other people to know what I am writing. I would also feel weird using it alone.	I liked it, but could not use it as a standalone method because there are problems with homophones and limitations to the dictionary. Otherwise, very fast.
If you are looking for something simple it is a good option.	This was the easiest to use. We speak all of the time. The only deterrent is to not interfere with other people
This was easy to use, hands-free and semi-reliable. All you have to do is push a button and speak into the phone. It did work very well, but I like to have more control over the input. I had to proofread the message afterward.	I like it because it was quick and simple, if you pronounce your words correctly. However, making corrections could be easier and I do not always want people to hear what I say.
It would be very convenient to use. Not a lot of buttons to press.	I liked the hands-free aspect. Not tiring at all.
This would be great to have while I am driving. The processing time was a little bit long.	It recognized everything I said. I have a very positive opinion on its capabilities.
I would use this for longer messages, because for quicker messages (like "ok") it would be faster just to type it than to hit the launch icon, speak it, and wait for it to process. Plus, you are limited on environments where you can use it. You can not use it with a lot of background noise, and not in a place with a lot of people . They would think you were weird.	
I did not like it at all. I felt it was distracting to hold it up to my mouth and having to wait to see the feedback	
I liked it. It was accurate and convenient, but I would not use it because of privacy concerns.	
If I had it I would use it. But only when I was by myself. It was simple and easy to use, but you still had to proofread it.	

APPENDIX E

MOTOR AND HAND FUNCTION ASSESSMENT MEASURES

Finger Tapping Test

The Finger Tapping Test is a common method for assessing motor speed. Poor performance on this task may indicate the beginnings of cognitive impairment and slowing due to age (Austin et al., 2011). Typically, participants would be asked to depress the lever on a piece of specialized equipment with their index finger as much as possible within a 10 second timeframe. Usually, there are 5 trials, or as many trials as required (but not to exceed 10 trials) to ensure tap count consistency to within 5 taps per trial. The index finger of each hand is tested separately. There are many variants of this technique in the literature; some of which use a computer keyboard rather than a lever. For practical reasons, this is the technique used in this study.

Participants depressed the space bar as quickly as possible with their right index finger during 5 10-second durations, taking a 30-second rest between trials, and a 60-second rest after the third trial. This will be repeated with their left index finger. The spaces for each trial were typed into individual cells in Microsoft Excel. Time was measured using a stopwatch.

Baseline Keyboarding

The purpose of gathering a baseline measure of keyboarding was to assess the participant's familiarity with the Qwerty layout. Typically, baseline typing speed is assessed using a paragraph selection or a series of short phrases, such as pangrams. Pangrams are phrases which contain every word in the alphabet. The examples used in this study were: "the quick brown fox jumped over a lazy dog", "pack my box with five dozen liquor jugs", and “sympathizing would fix quaker objectives”. Participants were asked to type three pangrams into a Microsoft Word document, quickly and accurately, without capitalization or punctuation. The entry of each phrase was timed with a stopwatch, and the average entry rates were computed across the three trials.

Hand Function

The Jebsen-Taylor test for hand function includes a variety of measures, and one is the writing test. Participants are asked to write a phrase in cursive on a sheet of paper. Each phrase is comprised of 24 characters, and is at a third-grade level of difficulty (Hackel et al.,1992). The three Jebsen-Taylor sentences used in this study are: “The old man seemed to be tired”, “Fish take air out of the water”, and “John saw the red truck coming”. Normative data (see Table E1) reflects Jebsen-Taylor sentences copied with the dominant hand. Evaluating hand function is important, and it was expected to be correlated with the entry rate for the Handwriting text input method. It has also been found to correlate with baseline keyboarding speed (Weintraub et al., 2010). It is also helpful because it reflects the participants’ dexterity. Typically, women write faster in all age groups. Handwriting speed tends to decrease after the age of 26 in women and after the age of 36 in men. Studies have shown that the slowest handwriting is typically produced by people over the age of 65, however, young men aged 16 – 25 have been shown to write more slowly than older men. There is usually a trade-off between speed and decreasing legibility. Normative data does exist for handwriting speed, however, it has been noted that this data is 20 years old or older and may not accurately reflect any changes in the population which could be due to increases in technology use (van Drempt et al., 2011).

APPENDIX E (continued)

TABLE E1

AGE AND GENDER NORMS

FOR COPYING JEBSEN-TAYLOR SENTENCES (SEC)

Age group	Gender	Time (SD)
16 – 25[1]	Male	12.8 (6.1)
	Female	9.6 (2.0)
26 – 35[1]	Male	10.1 (2.4)
	Female	10.1 (2.6)
56 – 65[1]	Male	15.1 (8.8)
	Female	11.3 (3.1)
60 – 69[2]	Male	15.3 (5.4)
	Female	12.4 (2.8)
70 - 79[2]	Male	17.2 (4.6)
	Female	13.4 (2.6)
80 – 89[2]	Male	20.5 (4.7)
	Female	18.2 (4.8)

Note: Values obtained from studies published in Van Drempt et al., 2011.
[1]Results from Agnew and Maas, 1982 (n = 382)

[2]Results from Hackel et al., 1992 (n = 121)

Finger Sensitivity

Sensation levels for the palmar surface of the distal phalanges were assessed using the Semmes-Weinstein monofilament hand mini-kit, which included filaments for testing 50mg, 200mg, 2g, 4g, and 300g of mean force. These pressure thresholds are color-coded as green, blue, purple, red and orange, respectively. This test evaluated sensation of fingers associated with the median (thumb and index) and ulnar (pinky) nerves. Each quadrant of these fingers (see Figure E1) was tested. Testing started with the lightest filament (green), held perpendicular and touching an area of the finger for 1.5 second intervals, up to three times. If a person did not respond that they felt it, the next filament was applied. The green filament is considered "normal", while blue indicates diminished light touch, purple is interpreted as diminished protective sensation, red implies loss of protective sensation and orange indicates the loss of all sensation except deep pressure. At the "blue" level, the use of the hand is close to normal, but at the "purple" level, there is some difficulty in object manipulation. The sensation levels at each quadrant denoted by Figure E1 are displayed for participants in both age groups, provided in Tables E2 and E3.

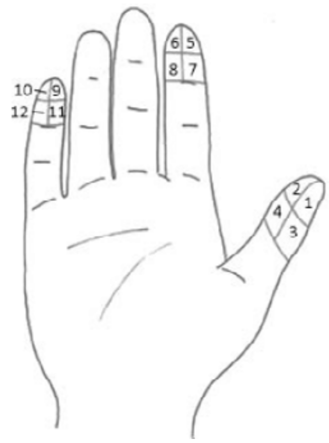


Figure E1. The diagram for recording twelve areas of sensitivity on the right hand.

TABLE E2

YOUNG ADULT FINGER SENSITIVITY FOR EACH DIGIT

	1	2	3	4	5	6	7	8	9	10	11	12
Green	23	22	24	24	24	24	24	24	24	23	24	23
Blue	2	3	1	1	1	1	1	1	1	2	1	2

Note: Values reflect frequencies of participants at each sensitivity level by digit quadrant.

TABLE E3

YOUNG ADULT FINGER SENSITIVITY FOR EACH DIGIT

	1	2	3	4	5	6	7	8	9	10	11	12
Green	9	9	10	10	13	15	15	17	17	18	18	18
Blue	14	14	14	13	11	9	10	8	8	7	7	7
Purple	2	2	1	1	1	1	0	0	0	0	0	0

Note: Values reflect frequencies of participants at each sensitivity level by digit quadrant.

Functional Range of Motion

Functional thumb reach and range of motion was assessed using an apparatus similar to that suggested by Gilbert and colleagues (1988), see Figure E2. The measurement apparatus is a box with dimensions 6 x 4 x 0.5 inches. The depth of the apparatus approximates the depth of the device under evaluation, however, since smartphones typically have small form factors, the height and width are larger to accommodate the full range of reach and angular displacement measurements for the thumbs of both hands. The purpose for including this measurement was to gather maximum and most comfortable angles of thumb rotation around the metacarpophalangeal joint for study participants. The intent was to help explain any potential fatigue and performance deficits for participants with lowered ability for angular displacement.

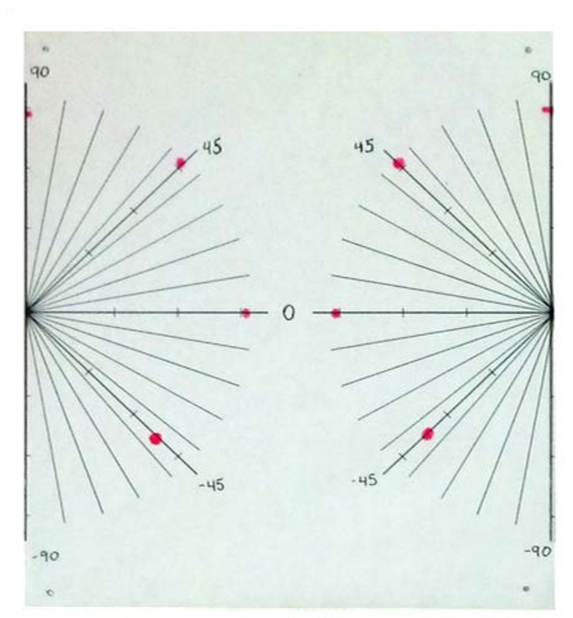


Figure E2. The board used to assess range of motion and thumb reach.

APPENDIX F

ANTHROPOMETRIC ASSESSMENT MEASURES

Since text entry on smartphones is mostly a manual task, it was hypothesized that anthropometry would influence participants' error rates and perceived satisfaction. Hand and digit measurements were acquired using a set of 12" digital calipers, and digit circumference was gathered using a flexible measuring tape. All measurements were in millimeters. Analyses were conducted on averages calculated between hands for each participant. Measurements were taken using the research from García-Cáceres and colleagues (2012) as a guide.

Hand Measurements

Overall hand length and breadth were gathered. The hand was in a neutral posture, with extended fingers, and the palm up. The hand length was taken from the tip of the middle finger to the wrist crease. Breadth was taken across the palm, not including the thumb.

Digit Measurements

Likewise, for digit measurements the hand was in a neutral posture with extended, abducted fingers. Length measurements were taken palm up, while width and circumference measurements were taken palm down. Thumb length was measured from the fingertip to the crease of the metacarpophalangeal joint along the middle line of the thumb. Thumb width (or joint thickness) was measured across the crease of the interphalangeal joint. Thumb joint circumference was taken using a flexible measuring tape around the circumference of the interphalangeal joint. The index finger length was measured the same way.

APPENDIX G

SPEECH AND VOICE ASSESSMENT MEASURES

Speaking Rate

The Grandfather Passage was developed by Charles Van Riper (1963), and was created to contain all American English phonemes at least once in the passage. The word sequences were intended to be somewhat difficult for people who have speech disorders, making this standard passage an informal diagnostic tool. The passage is as follows:

> "You wished to know all about my grandfather. Well, he is nearly ninety-three years old. He dresses himself in an ancient black frock coat, usually minus several buttons; yet he still thinks as swiftly as ever. A long, flowing beard clings to his chin, giving those who observe him a pronounced feeling of the utmost respect. When he speaks his voice is just a bit cracked and quivers a trifle. Twice each day he plays skillfully and with zest upon our small organ. Except in the winter when the ooze or snow or ice prevents, he slowly takes a short walk in the open air each day. We have often urged him to walk more and smoke less, but he always answers, "Banana Oil!" Grandfather likes to be modern in his language." (p. 484)

Participants were asked to read this passage using their normal rate and intensity of speech. Time was kept with a stopwatch. As there are 132 words in this passage, speaking rate was figured by dividing 132 by the amount of time taken to read the paragraph, in seconds. Thus, speech rate was figured in Words per Second (WPS).

Voice Qualities

Since it was hypothesized that various voice qualities would impact performance with the Voice input method, the Multi-Dimensional Voice Program (MDVP) (KayPENTAX, 2008) was used to gather selected measures of pitch, jitter, shimmer, and breathiness. Participants were asked to hold a flat, sustained “ah” utterance for four seconds, while holding a microphone roughly six inches away. This was repeated twice, and the measures were averaged for each participant.

Average fundamental frequency was the measure used to describe participants’ pitch. This was of interest because fundamental frequency has been shown to change with age. It increases with males and declines with females (Honjo & Isshiki, 1978), likely due to gender differences in the aging of vocal structures (Aronson & Bless, 2009). Both the fundamental frequency and the amplitude of a signal will vary, and the magnitude of these changes is expressed through frequency and amplitude tremor intensity. Pitch perturbation (cyle-to-cycle variation in signal frequency) is a measure for jitter, and amplidue perturbation (cycle-to-cycle variation in signal amplitude) is a measure for shimmer, both of which evaluate the perceived hoarseness or breathiness of a voice. The presence of perturbation is an indication of some irregularity in the vibration of vocal folds (Aronson & Bless, 2009), but could be related to neuromuscular activity, or characteristics of the vocal fold. Voice turbulence is also a breathiness parameter, which is related to incomplete or loose adduction of vocal folds (KayPENTAX, 2007).

APPENDIX H

COGNITIVE ASSESSMENT MEASURES

Standardized Mini-Mental State Examination

The original Mini-Mental State Exam (Folstein et al., 1975) consists of 11 questions, and was created to assess cognitive aspects of mental function for older adults. The first section of the test asks participants to vocalize their responses to orientation, memory and attention questions. The second section covers naming, ability to follow commands, writing a sentence and copying complex shapes. Maximum MMSE total score is 30, 21 points are possible in the first section. Typically, a cutoff score of 23 is recommended for "normal" cognitive function (Tombaugh & McIntyre, 1992). The test requires 5-10 minutes to administer, but is not timed.

The MMSE is reported to be significantly correlated with WAIS Verbal IQ (r = 0.776) and Performance IQ (r = 0.660). Additionally, test-retest reliability was high after 24 hours (r = 0.887) and 28 days apart (r = 0.98), and scores were not significantly different from each other (Folstein et al., 1975). Some effects have been found for age and education.

This test became wildly successful, however, the instructions on administering the test were somewhat vague. Molloy and colleagues proposed a standardized form of the MMSE, adding explicit instructions and time limits. Interestingly, the intrarater variability was reduced by 86% and interrater variability was reduced by 76% when compared to the MMSE. Intraclass correlation for the SMMSE was 0.90, versus 0.69 of the MMSE. Additionally, mean assessment duration was reduced by nearly 3 minutes (Molloy et al., 1991).

Symbol-Digit Substitution

The Symbol-Digit Substitution (also known as Digit Symbol Substitution) Test requires participants to use a code table (see Figure H1) to draw a symbol in a box underneath each box with a corresponding digit. There are 9 digit-symbol pairs (for numbers 1 through 9). There are 7 numbered practice boxes, followed by either 90 test boxes (shown in the 1955 version) or 133 test boxes (in the 1997 version) (Ryan et al., 2001). The participant's score is determined by counting the total number of responses (*tsds*) and the number of correct responses (*csds*). The number of incorrect responses is calculated by:

$$psds = \frac{(tsds - csds)}{tsds} \tag{H1}$$

Items which are completed out of sequence are not counted as correct, and the test is restricted to 90 seconds in length for tests with 90 boxes (Kalra et al., 1993) and 120 seconds for the test version with 133 boxes (Ryan et al., 2001). It has been included as a subtest in the Wechsler Adult Intelligence Scale (WAIS) because it is highly correlated with full-scale intelligence (r = .74). It is also correlated with age (r = -.54). The number of correct scores on the 90-second version of this test decline from a median of 65 for people in their 20s to 45 for people in their 70s. However, this decline due to aging does not result from a decline in handwriting speed. Instead, it appears as though older adults' slower responses are caused by slower cognitive processing speeds (Salthouse, 1992), which appear to account for 50% of the instrument's variance. Memory only accounts for roughly 7% of the variance (Joy et al., 2004).

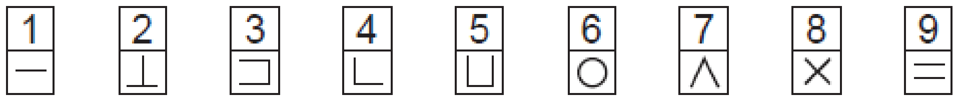


Figure H1. The SDST coding key, which displays each number and corresponding symbol.

Test-retest reliability is high for elderly participants (*r* = 0.82); intraindividual coefficient of variation was low, when comparing initial test scores to scores obtained 4 weeks later (*mean CV* = 4.7%) (Kalra et al., 1993). This study used a recording sheet with 133 boxes, thus, participants to were told to copy symbols for 120 seconds. Analyses were conducted using the number of correct responses.

APPENDIX I

SMARTPHONE DEVICE, INPUT METHODS AND SETTINGS

The smartphone device used in this study was the Motorola Droid 4, its keyboard was used for the Physical Qwerty evaluation. The software was the FlexT9 suite, which included the Onscreen Qwerty, Tracing, Handwriting and Voice input methods. At the time of data collection, which ran from March through August 2012, the Droid 4 had one of the best reviewed physical keyboards on the market. The FlexT9 software was chosen for the other input methods to maintain consistency of look and feel and the autocorrect algorithms across input methods. Images of the input methods are shown in Table I1.

TABLE I1

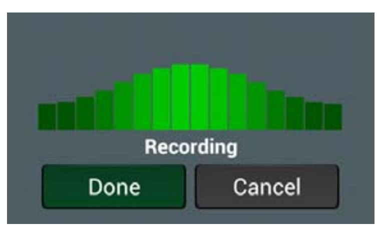
SMARTPHONE INPUT METHODS

Physical Qwerty

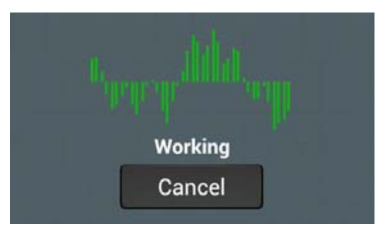
Onscreen Qwerty

Tracing

Handwriting



Voice (prompt stage)



Voice (processing stage)

Keyboard Dimensions

The dimensions of the Physical Qwerty keyboard are approximately 3 cm in height, and 10 cm in width, not including the bezel (which adds an additional 7.5 cm in height, including the screen, and 2.5 cm in width). The space bar is approximately 32 x 5 mm, and character keys are 7 x 5 mm. The numeric, shift, caps lock, tab and enter keys are different dimensions, which are not listed here as they were not used during the text entry task. Spacing between keys is approximately 1 mm, horizontally and vertically. The dimensions of the Onscreen Qwerty are approximately 28 x 89 mm, not including the bezel (which adds an additional 3.75 cm in height, including the text field, and 3.5 cm in width). The space bar is approximately 26 x 6 mm and the character keys are 8 x 6 mm. Spacing between keys is less than 1 mm horizontally and approximately 1 mm vertically. The keybord and key size and spacing are the same for the Tracing and Handwriting input methods. The launch button for the Voice input method is the same size as the character keys for the Onscreen Qwerty.

Keyboard Settings

TABLE I2

INPUT METHOD SOFTWARE SETTINGS

Input Method	**Enabled**	**Disabled**
Physical Qwerty	Auto-replace	Auto-capitalization Auto-punctuate
FlexT9	English Auto-correction (low) Enable trace input Auto-accept words with a space Detect End-of-Speech	Vibrate on keypress Sound on keypress Auto-capitalization Auto-punctuation Next word prediction Personalization Contribute usage data

APPENDIX J

PHRASE SET STIMULI

my watch fell in the water
prevailing wind from the east
never too rich and never too thin
breathing is difficult
I can see the rings on Saturn
physics and chemistry are hard
my bank account is overdrawn
elections bring out the best
we are having spaghetti
time to go shopping
a problem with the engine
elephants are afraid of mice
my favorite place to visit
three two one zero blast off
my favorite subject is psychology
circumstances are unacceptable
watch out for low flying objects
if at first you do not succeed
please provide your date of birth
we run the risk of failure
prayer in schools offends some
he is just like everyone else
great disturbance in the force
love means many things
you must be getting old
the world is a stage
can I skate with sister today
neither a borrower nor a lender be
one heck of a question
question that must be answered
beware the ides of March
double double toil and trouble
the power of denial
I agree with you
do not say anything
play it again Sam
the force is with you
you are not a jedi yet
an offer you cannot refuse
are you talking to me
yes you are very smart
all work and no play
hair gel is very greasy
Valium in the economy size
the facts get in the way
the dreamers of dreams
did you have a good time
space is a high priority
you are a wonderful example
do not squander your time
do not drink too much
take a coffee break
popularity is desired by all
the music is better than it sounds
starlight and dewdrop
where did I leave my glasses
on the way to the cottage
a lot of chlorine in the water
do not drink the water
my car always breaks in the winter
santa claus got stuck
public transit is much faster
zero in on the facts
make up a few more phrases
my fingers are very cold
rain rain go away
bad for the environment
universities are too expensive
the price of gas is high
the winner of the race
we drive on parkways
we park in driveways
go out for some pizza and beer
effort is what it will take
where can my little dog be
if you were not so stupid
not quite so smart as you think
do you like to go camping
this person is a disaster
the imagination of the nation
universally understood to be wrong
listen to five hours of opera
an occasional taste of chocolate
victims deserve more redress
the protesters blocked all traffic
the acceptance speech was boring
work hard to reach the summit
a little encouragement is needed
stiff penalty for staying out late
the pen is mightier than the sword
exceed the maximum speed limit
in sharp contrast to your words
this leather jacket is too warm
consequences of a wrong turn
this mission statement is baloney
you will loose your voice
every apple from every tree
are you sure you want this
the fourth edition was better
this system of taxation
beautiful paintings in the gallery
a yard is almost as a meter
we missed your birthday
coalition governments never work
destruction of the rain forest
I like to play tennis
acutely aware of her good looks
you want to eat your cake
machinery is too complicated
a glance in the right direction
electric cars need big fuel cells
the plug does not fit the socket
drugs should be avoided
the most beautiful sunset
we dine out on the weekends
get aboard the ship is leaving
the water was monitored daily
he watched in astonishment
a big scratch on the tabletop
salesmen must make their monthly
quota
saving that child was an heroic effort
granite is the hardest of all rocks
bring the offenders to justice
every Saturday he folds the laundry
careless driving results in a fine
microscopes make small things look big
a coupon for a free sample
fine but only in moderation
a subject one can really enjoy
important for political parties
that sticker needs to be validated
the fire raged for an entire month
one never takes too many precautions
we have enough witnesses
labour unions know how to organize
people blow their horn a lot
a correction had to be published
I like baroque and classical music
the proprietor was unavailable
be discreet about your meeting
meet tomorrow in the lavatory
suburbs are sprawling up everywhere
shivering is one way to keep warm
dolphins leap high out of the water
try to enjoy your maternity leave
the ventilation system is broken
dinosaurs have been extinct for ages
an inefficient way to heat a house
the bus was very crowded
an injustice is committed every day
the coronation was very exciting
look in the syllabus for the course
rectangular objects have four sides
prescription drugs require a note
the insulation is not working
nothing finer than discovering a treasure
our life expectancy has increased
the cream rises to the top
the high waves will swamp us
the treasurer must balance her books
completely sold out of that
the location of the crime
the chancellor was very boring
the accident scene is a shrine for fans

APPENDIX J (continued)

the living is easy
fish are jumping
the cotton is high
drove my chevy to the levee
but the levee was dry
I took the rover from the shop
movie about a nutty professor
come and see our new car
coming up with killer sound bites
I am going to a music lesson
the opposing team is over there
soon we will return from the city
I am wearing a tie and a jacket
the quick brown fox jumped
all together in one big pile
wear a crown with many jewels
there will be some fog tonight
I am allergic to bees and peanuts
he is still on our team
the dow jones index has risen
my preferred treat is chocolate
the king sends you to the tower
we are subjects and must obey
mom made her a turtleneck
goldilocks and the three bears
we went grocery shopping
the assignment is due today
what you see is what you get
for your information only
a quarter of a century
the store will close at ten
head shoulders knees and toes
vanilla flavored ice cream
frequently asked questions
round robin scheduling
information super highway
my favorite web browser
the laser printer is jammed
all good boys deserve fudge
the second largest country
call for more details
just in time for the party
have a good weekend
video camera with a zoom lens
what a monkey sees a monkey will do
that is very unfortunate
the back yard of our house
this is a very good idea
reading week is just about here
our fax number has changed
thank you for your help
no exchange without a bill
the early bird gets the worm
buckle up for safety
this is too much to handle
protect your environment
world population is growing
the library is closed today
I just cannot figure this out
please follow the guidelines
an airport is a very busy place
mystery of the lost lagoon
is there any indication of this
the chamber makes important decisions
this phenomenon will never occur
obligations must be met first
valid until the end of the year
file all complaints in writing
tickets are very expensive
a picture is worth many words
this camera takes nice photographs
it looks like a shack
the dog buried the bone
the daring young man
this equation is too complicated
express delivery is very fast
I will put on my glasses
a touchdown in the last minute
the treasury department is broke
a good response to the question
well connected with people
the bathroom is good for reading
the generation gap gets wider
chemical spill took forever
prepare for the exam in advance
interesting observation was made
bank transaction was not registered
your etiquette needs some work
we better investigate this
stability of the nation
house with new electrical panel
our silver anniversary is coming
the presidential suite is very busy
the punishment should fit the crime
sharp cheese keeps the mind sharp
the registration period is over
you have my sympathy
the objective of the exercise
historic meeting without a result
very reluctant to enter
good at addition and subtraction
six daughters and seven sons
a thoroughly disgusting thing to say
sign the withdrawal slip
relations are very strained
the minimum amount of time
a very traditional way to dress
the aspirations of a nation
medieval times were very hard
a security force of eight thousand
there are winners and losers
the voters turfed him out
pay off a mortgage for a house
the collapse of the Roman empire
did you see that spectacular explosion
keep receipts for all your expenses
a tumor is OK provided it is benign
please take a bath this month
rent is paid at the beginning of the month
for murder you get a long prison sentence
a much higher risk of getting cancer
quit while you are ahead
knee bone is connected to the thigh bone
safe to walk the streets in the evening
luckily my wallet was found
one hour is allotted for questions
so you think you deserve a raise
they watched the entire movie
good jobs for those with education
jumping right out of the water
the trains are always late
sit at the front of the bus
do you prefer a window seat
the food at this restaurant
Canada has ten provinces
the elevator door appears to be stuck
raindrops keep falling on my head
spill coffee on the carpet
an excellent way to communicate
with each step forward
faster than a speeding bullet
wishful thinking is fine
nothing wrong with his style
arguing with the boss is futile
taking the train is usually faster
what goes up must come down
be persistent to win a strike
presidents drive expensive cars
the stock exchange dipped
why do you ask silly questions
that is a very nasty cut
what to do when the oil runs dry
learn to walk before you run
insurance is important for bad drivers
traveling to conferences is fun
do you get nervous when you speak
pumping helps if the roads are slippery
parking tickets can be challenged
apartments are too expensive
find a nearby parking spot
gun powder must be handled with care
just what the doctor ordered
a rattle snake is very poisonous
weeping willows are found near water
I cannot believe I ate the whole thing
the biggest hamburger I have ever seen
gamblers eventually loose their shirts
exercise is good for the mind
irregular verbs are the hardest to learn
they might find your comment offensive
tell a lie and your nose will grow
an enlarged nose suggests you are a liar

Mary had a little lamb
teaching services will help
we accept personal checks
this is a non profit organization
user friendly interface
healthy food is good for you
hands on experience with a job
this watch is too expensive
the postal service is very slow
communicate through email
the capital of our nation
travel at the speed of light
I do not fully agree with you
gas bills are sent monthly
earth quakes are predictable
life is but a dream
take it to the recycling depot
sent this by registered mail
fall is my favorite season
a fox is a very smart animal
the kids are very excited
parking lot is full of trucks
my bike has a flat tire
do not walk too quickly
a duck quacks to ask for food
limited warranty of two years
the four seasons will come
the sun rises in the east
it is very windy today
do not worry about this
dashing through the snow
want to join us for lunch
stay away from strangers
accompanied by an adult
see you later alligator
make my day you sucker
I can play much better now
she wears too much makeup
my bare face in the wind
batman wears a cape
I hate baking pies
lydia wants to go home
win first prize in the contest
freud wrote of the ego
I do not care if you do that
always cover all the bases
nobody cares anymore
can we play cards tonight
get rid of that immediately
I watched blazing saddles
the sum of the parts
they love to yap about nothing
peek out the window
be home before midnight
he played a pimp in that movie
I skimmed through your proposal
he was wearing a sweatshirt
no more war no more bloodshed
the assault took six months
get your priorities in order
traveling requires a lot of fuel
longer than a football field
a good joke deserves a good laugh
the union will go on strike
never mix religion and politics
interactions between men and women
where did you get such a silly idea
it should be sunny tomorrow
a psychiatrist will help you
you should visit to a doctor
you must make an appointment
the fax machine is broken
players must know all the rules
a dog is the best friend of a man
would you like to come to my house
February has an extra day
do not feel too bad about it
this library has many books
construction makes traveling difficult
he called seven times
that is a very odd question
a feeling of complete exasperation
we must redouble our efforts
no kissing in the library
that agreement is rife with problems
vote according to your conscience
my favourite sport is racketball
sad to hear that news
the gun discharged by accident
one of the poorest nations
the algorithm is too complicated
your presentation was inspiring
that land is owned by the government
burglars never leave their business card
the fire blazed all weekend
if diplomacy does not work
please keep this confidential
the rationale behind the decision
the cat has a pleasant temperament
our housekeeper does a thorough job
her majesty visited our country
handicapped persons need consideration
these barracks are big enough
sing the gospel and the blues
he underwent triple bypass surgery
the ropes of a new organization
peering through a small hole
rapidly running short on words
it is difficult to concentrate
seasoned golfers love the game
he cooled off after she left
my dog sheds his hair
join us on the patio
these cookies are so amazing
I can still feel your presence
the dog will bite you
lie detector tests never work
do not lie in court or else
most judges are very honest
only an idiot would lie in court
important news always seems to be late
please try to be home before midnight
if you come home late the doors are locked
dormitory doors are locked at midnight
staying up all night is a bad idea
you are a capitalist pig
motivational seminars make me sick
questioning the wisdom of the courts
rejection letters are discouraging
the first time he tried to swim
that referendum asked a silly question
a steep learning curve in riding a unicycle
a good stimulus deserves a good response
everybody loses in custody battles
put garbage in an abandoned mine
employee recruitment takes a lot of effort
experience is hard to come by
everyone wants to win the lottery
the picket line gives me the chills
give me one spoonful of coffee
two or three cups of coffee
just like it says on the can good
companies announce a merger
what a lovely red jacket
do you like to shop on Sunday
I spilled coffee on the carpet
the largest of the five oceans
shall we play a round of cards
toss the ball around
I will meet you at noon
I want to hold your hand
the children are playing
superman never wore a mask
I listen to the tape everyday
he is shouting loudly
correct your diction immediately
a most ridiculous thing
where did you get that tie
do a good deed to someone
quick there is someone knocking
flashing red light means stop
sprawling subdivisions are bad
olympic athletes use drugs
my mother makes good cookies

APPENDIX K

CONSENT FORMS FOR YOUNGER AND OLDER ADULTS

WICHITA STATE UNIVERSITY

Consent Form

Purpose: You are invited to participate in a study investigating various smartphone text input methods. We hope to learn how these input methods impact smartphone usability.

Participant Selection: You were selected as a possible participant in this study because you are 18 years of age or older, and you fit the criteria of the population we are interested in studying. You will be one of approximately 50 participants recruited from Wichita State University and the Wichita community.

Explanation of Procedures: If you decide to participate, you will be asked to complete a series of tasks on a smartphone. Your participation in this study will take approximately 3 hours.

Discomfort/Risks: There is no foreseeable discomfort or risks associated with participation in this experiment.

Benefits: This will provide a better understanding of the design of smartphone text input methods. Recommendations from the data collected in this study may help increase the future usability of these text input methods.

Confidentiality: Any information obtained in this study in which you can be identified will remain confidential and will be disclosed only with your permission.

Refusal/Withdrawal: Participation in this study is entirely voluntary. This experiment will be digitally recorded. There is no expected risk or discomfort. You will be permitted to take a break should you so desire. Information collected in this study will be analyzed and submitted solely to the corporation sponsoring this research. You will be compensated either class credit or $30 for your participation. Your decision whether or not to participate will not affect your future relations with Wichita State University. If you decide to participate, you may withdraw from the study at any time without affecting your status with Wichita State University.

Contact: If you have questions during the study, please ask the experimenter. If you have additional questions about this research you may contact Dr. Barbara Chaparro at 316-978-3683, or Amanda Smith at 620-664-1337. If you have questions pertaining to your rights as a research subject, or about research-related injury, you can contact the Office of Research Administration at Wichita State University, Wichita, KS 67260-0007,
telephone (316) 978-3285.

You will be offered a copy of this consent form to keep.

You are making a decision whether or not to participate. Your signature indicates that you have read the information provided above and have voluntarily decided to participate.

______________________________________ ______________________

Signature of Participant Date

______________________________________ ______________________

Signature of Investigator Date

WICHITA STATE UNIVERSITY

Consent Form

<u>Purpose</u>: You are invited to participate in a study investigating various smartphone text input methods. We hope to learn how these input methods impact smartphone usability.

<u>Participant Selection</u>: You were selected as a possible participant in this study because you are 18 years of age or older, and you fit the criteria of the population we are interested in studying. You will be one of approximately 50 participants recruited from Wichita State University and the Wichita community.

<u>Explanation of Procedures</u>: If you decide to participate, you will be asked to complete a series of tasks on a smartphone. Your participation in this study will take approximately 3 hours.

<u>Discomfort/Risks</u>: There is no foreseeable discomfort or risks associated with participation in this experiment.

<u>Benefits</u>: This will provide a better understanding of the design of smartphone text input methods. Recommendations from the data collected in this study may help increase the future usability of these text input methods.

<u>Confidentiality</u>: Any information obtained in this study in which you can be identified will remain confidential and will be disclosed only with your permission.

<u>Refusal/Withdrawal</u>: Participation in this study is entirely voluntary. This experiment will be digitally recorded. There is no expected risk or discomfort. You will be permitted to take a break should you so desire. Information collected in this study will be analyzed and submitted solely to the corporation sponsoring this research. You will be compensated with either $30 cash or gift card for your participation. Your decision whether or not to participate will not affect your future relations with Wichita State University. If you decide to participate, you may withdraw from the study at any time without affecting your status with Wichita State University.

<u>Contact</u>: If you have questions during the study, please ask the experimenter. If you have additional questions about this research you may contact Dr. Barbara Chaparro at 316-978-3683, or Amanda Smith at 620-664-1337. If you have questions pertaining to your rights as a research subject, or about research-related injury, you can contact the Office of Research Administration at Wichita State University, Wichita, KS 67260-0007, telephone (316) 978-3285.

APPENDIX K (continued)

You will be offered a copy of this consent form to keep.

You are making a decision whether or not to participate. Your signature indicates that you have read the information provided above and have voluntarily decided to participate.

____

Signature of Participant Date

____

Signature of Investigator Date

APPENDIX L

FACILITATOR PROTOCOL

[Participant signs consent form]

INTRODUCTION

Thank you for participating in this study. We appreciate your time and effort!

There are two main parts of this study. The first part assesses several of your individual characteristics, including memory, handwriting, vision, manual dexterity, and hand measurements. This portion of the study will take a little less than an hour to complete.

During the second part of the study, you will be evaluating five different input methods on a smartphone device. We expect that you will need between 1 ½ to 2 ½ hours to work through this part of the study.

You will be given breaks periodically, however, if you need to take additional breaks, let me know.

Do you have any questions before we begin?

BEGIN INDIVIDUAL CHARACTERISTIC ASSESSMENTS (at long table)

- Near Visual Acuity
- Standardized Mini-Mental (***if P ID# 26 – 50***)
- Digit Symbol – Coding
- Write Alphabet
- Jebsen-Taylor sentences
- Hand sensitivity (monofilaments)
- Grooved Pegboard Test
- Finger Tapping Test
- Anthropometric measurements
- Speech Qualities (*done only directly before the Voice Recognition condition*)

MOVE TO COMPUTER AT PARTICIPANT STATION

- Keyboarding Speed (pangrams)
- Attitudes Toward Computers Questionnaire
- Read Grandfather Passage (*done only directly before the Voice Recognition condition*)

Near Visual Acuity (Binocular)

Materials:

- Pocket eye chart (for participant)
- Sheet of paper to record participant's acuity

Facilitator:

[Hand the eye chart to the participant.]

"We're going to test your near vision with this eye chart. Note the string attached to the chart. I'm going to ask you to put the string to the tip of your nose, holding the chart at that distance from your face. "

"Please read the first letter of each row, starting with the top row.

Continue with the middle groups of letters. There are two lines where there is no middle row. Use the groups of letters on the left side for those two rows, and then continue on in the middle."

[Have the participant read the first letter of each row until they make an error. When they make an error, move back one line and ask them to read the whole line.

The last line where they get at least 3 of the 5 correct is considered to be their acuity.]

Scoring (for the middle row):

Row	Acuity	Letters
1	20/400	**C O H Z V**
2	20/320	**S Z N D C**
3	20/250	**V K C N R**
4	20/200	**K C R H N**
5	20/160	**O S D V Z K D V C**
6	20/125	**(left) N O Z C D (right) H V O R K**
7	20/100	**(left) R D N S K (right) R H S O N**
8	20/80	**C D H N R**
9	20/60	**H N K C D**
10	20/50	**N D V K O**
11	20/40	**D H O S Z**
12	20/30	**V R N D O**
13	20/25	**C Z H K S**
14	20/20	**O R Z S K**
15	20/16	**S C N D Z**

Semmes-Weinnstein Monofilament Instructions

Materials:

- Hand diagram worksheet
- Colored pencils (green, blue, purple, red and orange)
- Monofilaments (2.83, 3.61, 4.31, 4.56 and 6.65)
- Towel to support the hand

Facilitator:

"Now we are going to test the sensation levels of your hands. To do this, we will touch these plastic fibers to your fingers and ask whether you can feel them. To keep your hand stable, I'm going to ask you to please rest your hand, palm up, on this towel. Now let me demonstrate."

[*Have the participant rest their hand (palm up) on a towel to stabilize it. Demonstrate on the palm of their hand so they understand the process. Use the purple or red monofilament so that they can feel it. Take the monofilament and press it at a 90 degree angle against the skin until it bows. Hold it in place for 1.5 seconds and then remove. Instruct the participant to say "yes" or "touch" when they feel it.*]

"Now I am going to test your fingers in the same way. You will not be allowed to watch this test, so I am going to ask you to also either keep your eyes closed or to look away while I apply the fibers. Remember to tell me if you feel something. There may be durations of time where I am not applying the fibers."

[*Make sure their eyes are closed or that they are looking away from the testing area.*

Test the thumb, index finger (median nerve function) and pinky finger (ulnar nerve function).

For each monofilament, make sure to apply it at a 90 degree angle until it bows (in a "C" shape) and hold for 1.5 seconds before releasing. Apply them somewhat randomly. Apply the 2.83 and 3.61 monofilaments 3 times to the same area. For a positive result, the participant must be able to feel the monofilament 2 of the 3 times it is applied. Apply the 4.31, 4.56 and 6.65 monofilaments only once; a positive result is obtained if they feel the monofilament the one time it is applied.

Begin with the 2.83 monofilament. If the participant responds to the stimulus in all sites, they have normal cutaneous sensation and the examination is complete. If they do not respond, retest with the next largest monofilament and repeat the process.

When a participant indicates a response, record the result using a colored pencil that corresponds to the color of the monofilament. If they are unable to respond to ANY of the monofilaments, fill in the area with orange hash marks.]

Standardized Mini-Mental State Exam

Materials Needed: wrist watch, pencil, blank sheet of paper cut in half, and the page with intersecting polygons printed on it.

Facilitator Directions:

1. Before the questionnaire is administered, try to get the person to sit down facing you. Assess the person's ability to hear and understand very simple conversation (e.g., What is your name?). If the person uses hearing or visual aids (e.g., glasses), make sure they are wearing them.
2. Introduce yourself and try to get the person's confidence. Before you begin, get the person's permission to ask questions (e.g., "Would it be alright to ask you some questions about your memory?").
3. Ask each question a maximum of 3 times. If the participant does not respond, score 0.
4. If the person answers incorrectly, score 0. Accept this answer and do not ask the question again, hint, or provide any physical slues such as head shaking.
5. If the person answers: "What did you say?", just repeat the same directions a maximum of 3 times. Do not explain or engage in conversation.
6. If the person interrupts (e.g., "What is this for?"), reply: "I will explain in a few minutes, when we are finished. Now if we could proceed please… we are almost finished."

Digit Symbol Instructions

Materials:

- Response sheet
- Pencil with eraser
- Stopwatch

Facilitator:

[*Point to the key above the test and sample items.*]

“Look at these boxes. Notice that each has a number in the upper part and a special mark in the lower part. Each number has its own mark.”

[*Point to the number 1 and its mark in the key, then to the number 2 and its mark. Next, point to the 7 squares marked as “sample items”.*]

“Now look down here where the squares have number in the top part but the squares at the bottom are empty. In each of the empty squares, put the mark that should go there. Like this.”

[*Point to the first sample item, then point back to the key to show its corresponding mark*.]

“Here is a 2; the 2 has this mark. So I put it in this empty square like this.”

[*Write the symbol. Point to the second sample item*.]

“Here is a 1; the 1 has this mark (point to the second sample item, then to the mark below the 1 in the key), so I put it in this square.”

[*Write the symbol. After marking the first 2 sample items say…*]

“Now you fill in the squares up to the heavy black line.”

[*If the participant makes an error on any of the sample items, correct the error immediately and review use of the key. Continue to provide help if needed. Don’t move onto the actual test until the participant clearly understands the task. When they complete a sample item correctly, offer encouragement like “yes” or “right”. When all sample items are complete, say…*]

“Now you know how to do them. When I tell you to start, you do the rest of them.”

[*Point to the first square to the right of the heavy line and say…*]

"Begin here and fill in as many squares as you can, one after the other without skipping any. Keep working until I tell you to stop. Work as quickly as you can without making any mistakes."

[*Sweep across the first row with your finger and say…*]

"When you finish this line, go on to this one."

[*Point to the first square in the second row*.]

"Start."

[*Begin timing.*

If participants ask what they should do if they make a mistake, encourage them to continue to work as fast as they can. It's ok for them to make corrections if they do them spontaneously, they shouldn't do many corrections that would impede their performance.

After 120 seconds, say **"Stop"**.]

Scoring the SDS Test

- A symbol is "correct" if it is clearly identifiable as the keyed symbol, even if it is drawn imperfectly or if it is a spontaneous correction of an incorrect symbol.
- **Each** correctly drawn symbol completed within 120 seconds is worth **1 point**.
- Responses to sample items are NOT included.
- Items completed out of sequence are NOT given credit.
- Use the Digit Symbol Scoring Template (yellow template) to check responses.
- MAXIMUM score of 133.
- Report the following information on the participant's physical form, **as well as** online in the study's Google Form under the Symbol Digit Substitution section:
    - Total number of responses
    - Number of correct responses

Jebsen-Taylor Writing Task Instructions

Materials:

- Alphabet (upper/lowercase) worksheet
- Jebsen-Taylor worksheet (on a clipboard)
- Jebsen-Taylor phrase notecards (3)
- stopwatch
- pen

Facilitator:

"The next two tasks involve handwriting. First, I would like for you to write the alphabet in uppercase on the top of this worksheet, and in lowercase on the bottom of this worksheet. Write the letters as you normally would."

[*Hand them the alphabet worksheet and a pencil.*

Note which hand is dominant as they do this task.]

"Now we are going to do another task. Take the pen in your non-dominant hand and arrange everything so that it is comfortable for you. On the other side of this card is a sentence. When I turn over the card and say "go", write the sentence as quickly and clearly as you can using your non-dominant hand. Write (cursive), do not print. Do you understand?"

[*Turn over one notecard.]*

"Go."

[*Time how quickly they can do this, by starting the stopwatch from when the pen touches the paper on the first word, to when the pen is lifted up on the last word.]*

"Now we are going to repeat this task with another sentence, only this time you will use your dominant hand. Are you ready?"

[*Turn over one notecard.]*

"Go."

[*Time how quickly they can do this, by starting the stopwatch from when the pen touches the paper on the first word, to when the pen is lifted up on the last word.]*

Grooved Pegboard Test Instructions

Materials:

- Grooved Pegboard
- stopwatch
- Paper to mark participants' times

Facilitator:

[The pegboard is placed in mid-line with the participant so that the board is at the edge of the table and peg tray is immediately above the board.]

"This is a pegboard and these are the pegs."

[Point to each and then pick up a peg.]

"All the pegs are the same. They have a groove, that is, a round side and a square side and so do the holes in the board. What you must do is match the groove of the peg with the groove of the board and put these pegs into the holes like this."

[Demonstrate by filling the top row. Then remove the pegs, putting them back into the tray.]

"When I say go, begin here and put the pegs into the board as fast as you can, using only your (dominant) hand. Fill the top row completely from this side to this side. Do not skip any; fill each row the same way you filled the top row. Any questions? Ready, as fast as you can, go."

[Begin timing when the participant starts the task until the last peg is put in OR at 3 minutes.

For the right hand trial, you demonstrate that the pegs are placed from left to right.

They are placed from right to left for the left hand trial.

The dominant hand is administered first, followed by the non-dominant hand trial.]

Additional instructions:

- You may tell them to speed up during the trial if necessary.
- You may need to point out the first hole of a new row.
- They are not allowed to pick up more than one peg at a time.
- Only one hand may be used at a time. They cannot use the other hand to help turn a peg around. You may tell the participant to keep the other hand in their lap.

- Anything that might affect performance should be noted (e.g., sore finger, bandages, etc.)
- If a peg is dropped on the floor, do not make any attempt to pick it up during the trial. Instead, a correctly placed peg (1st or 2nd peg) can be taken out and used.

Scoring:

- If the participant did not finish the trial (at the 3 minute mark), the trial is flagged as “A” which indicates an incomplete test.
- The first score is the number of seconds to complete the test, separately, for each hand.
- A second score is also taken, comprised of the number of “drops” made during each trial. This is considered to be any unintentional drop of a peg from the time that a participant attempts to pick up a peg until it is placed correctly in the hole. If more than one peg is picked up from the tray and the participant intentionally discards all but one of the pegs, it is not considered a drop. If a peg is intentionally laid down on the side of the tray or table to manipulate the peg, this is not considered a drop. If a peg is turned with the hand not being tested, this is noted. If this occurs more than once, the score is given a “D” flag for a nonstandard assessment.
- The third score is the number of pegs correctly placed in the holes for each trial.

For each hand, these three scores are summed (total time, number of drops, and total number of pegs correctly placed in the board) to get the complete score.

Finger Tapping Test (modified) Instructions

Materials:

- Finger Tapping Test excel document
- computer and keyboard
- stopwatch

Facilitator:

[*Seat the participant in front of the computer with the FTT document displayed on the screen. Make sure that the excel tab is selected for the appropriate participant (e.g., "P25").]*

"This task assesses your manual motor speed. You will be asked to tap the space bar as rapidly as possible, first using the index finger of your dominant hand. After I say "go", you will tap the space bar for 10 seconds, until I say "stop". We will do this 5 times, and you will be given 30 seconds of rest after each trial, except for the third trial where you will receive 60 seconds of rest. Then we will repeat this procedure with your non-dominant hand. I ask that you move ONLY your index finger, not your entire hand or arm. Are you ready? Go."

[Begin taking time when you say "go". Say "stop" when the stopwatch hits 10 seconds. Then let them rest for 30 seconds.

Repeat this procedure for the second, fourth and fifth trials. They receive 60 seconds rest after the third trial.

Note the variation in range for finger tapping (max and min values). If this is greater than 5 taps, you must do additional trials, omitting the highest/lowest value so that the range is 5 or less.

Save the FTT document after each participant.]

Scoring:

- The mean values will be reported for each hand per participant.
- If the participant does not obtain a range of taps (of 5 or less) within 10 trials for one hand, the five scores *closest* in range are used.

Keyboarding Instructions

Materials:

- "Pangram" word document
- computer and keyboard
- stopwatch

Facilitator:

[*Seat the participant in front of the computer with the Pangram document displayed on the screen.]*

"The next task will assess your keyboarding speed. The word document on the screen has 3 phrases typed in it. Your task will be to first place your hands in a comfortable position on the keyboard, then read the phrase aloud. When you are ready to type the phrase, say "start" so that we can start keeping time with a stopwatch. Do not worry about any capitalization or punctuation, but type the phrase quickly without making mistakes. You will type three phrases. Are you ready?"

[Begin keeping time when they hit their first key, and stop the stopwatch when they are finished with the phrase. Record the time in seconds. Repeat for each phrase.

Save the Pangram document for each participant as the participant's ID number (e.g., "P25").]

Scoring:

- Record the time (in seconds) for each phrase.
- WPM is calculated as: $[(|T| - 1) / S] * 12$
  T = the length of the transcribed string (including spaces); S = the seconds taken to transcribe it
- Since spelling/grammar is turned off, participants MAY leave errors within their phrases. These will need to be counted and tabulated in terms of:
  - Insertion errors (letters that are inserted)
  - Omission errors (missing letters)
  - Word spacing errors (missing or extra spaces)
  - Substitution errors (erroneous letters substituted for intended letters)

VOICE QUALITIES

Materials:

- Laptop with MDVP software program; Window “A” is the active window
- Microphone (switched “on”)

Facilitator:

[*Seat the participant in front of the laptop.]*

“Now we are going to assess some characteristics of your speech. To do this, you will need to hold this microphone 6” away from your mouth and produce an “ah” sound for about 4 seconds. Try to keep this at the same pitch and intensity that is typical to your speech. First, I will demonstrate.”

[Articulate an “ah” sound and it the “record” icon in the software. After 4 seconds, hit “stop”. Hand the microphone to the participant.]

“Now it is your turn. We will do this a total of 3 times. Ready? Go.”

[After each articulation, save the recording for the proper gender and with the appropriate participant ID.]

Scoring:

Calculate average values for each participant, based on appropriate parameters.

Speech Rate

Materials:

- Grandfather Passage
- Morae Recorder software
- stopwatch

Facilitator:

[Make sure that the video camera is recording and that Morae Recorder is on and recording!!!

Hand the copy of the Grandfather Passage to the participant.]

"Now we are going to assess your speaking rate. To do this, I am going to ask that you please read this paragraph at your normal speaking rate and volume. Start reading when I say "go". I will be keeping time with a stopwatch. Ready? Go."

[Mark the start/stop times for this task in Morae Recorder.

Begin taking time when you say "go". When they say "language", stop the stopwatch. Record the time it took them to read the passage in seconds.]

Scoring:

- There are 132 words in the passage. To figure speaking rate (WPM) calculate: 132 / S, where S is the number of seconds taken to articulate the entire passage.
- Record this value.

PERFORMANCE INSTRUCTIONS

Today you will be helping us evaluate five smartphone input methods. When I say "smartphone", I am referring to a handheld mobile device that allows you to take pictures, send e-mails, browse the internet, play music and provides GPS navigation, in addition to placing calls.

First, I will ask you to rate your preference of each method on a 0 – 50 point scale, based on your initial impressions.

Next, I will demonstrate how one particular input method works. You will then be asked to enter phrases on the phone using that input method. A phrase will appear on the phone. You will be asked to read the phrase aloud, this is so I know that you read the phrase correctly and so that I can retrieve the same phrase to display on the monitor in front of you. You will enter a total of 20 phrases for each input method; the first 5 are practice.

After entering phrases, you will rate the input method in terms of how satisfied you were with it, what you liked or disliked about it and your perceived level of workload while using it.

We will repeat this set of procedures 4 more times with the other 4 input methods. At the end, you will again rate your preference with each method on the 0 – 50 point scale. We expect that this portion of the study will take 1 ½ - 2 ½ hours.

As you participate in today's evaluation, please keep in mind:

We are testing the input methods, not you. We just want to gain a better understanding of the strengths and weaknesses of each input method so that we can ultimately improve their design.

Do you have any questions before we begin?

PREFERENCE RATINGS (PRE-task)

[UNCOVER THE 0 – 50 POINT SCALE
PROVIDE ONE-WORD DESCRIPTION OF EACH METHOD FOR RANKING, *IN ORDER*]

I am going to briefly describe all 5 of the input methods ***in the order that you will be using them****. For each method, I will hand you a card with a one-word description of that input method printed on it.*

[READ DESCRIPTIONS LISTED BELOW IN THE ORDER THEY ARE RECEIVED.]

I know you do not have any experience with these methods, but please rate each input method based on your ***initial impressions*** *of how desirable or valuable you perceive it to be* ***for entering text on a smartphone****.*

Place each of these words along this 0 – 50 scale, with ***0 being "least preferred"*** *and* ***50 being "most preferred"****. You are not allowed to have any ties.*

Why?

DESCRIPTIONS OF EACH METHOD

Note that when I say "Qwerty keyboard", I'm referring to the keyboard layout typical of what we use to type on computers and typewriters. A "mini" version of this is commonly used on smartphones.

- **(A) Typing (Physical Qwerty):** use your fingers to type on a mini-Qwerty keyboard.
- **(B) Tapping (Onscreen Qwerty):** use your fingers to tap on a touchscreen mini-Qwerty keyboard.
- **(C) Tracing:** create words by gliding your finger from one letter to the next on a touchscreen mini-Qwerty keyboard.
- **(D) Handwriting:** use your finger to draw what you want to type. The handwriting recognition will turn it into text.
- **(E) Voice Recognition:** Speak into the phone and the phone's software will convert it into text.

PHRASE ENTRY

The input method we are getting ready to use is called ______. Before we use this input method, I'll briefly describe how to use it and demonstrate it for you.

[Read appropriate description and demonstrate "hello world".]

FOR MANUAL METHODS:

- *Do not worry about capitalization or punctuation.*
- *Input text as quickly and accurately as you can.*
- *You are allowed to make corrections, but you are not forced to correct every error.*
- *Pretend that you are sending this to a friend.*
- *Do not use any shorthand (e.g., typing "u" instead of "you")*

FOR VOICE RECOGNITION:

[MAKE SURE TO READ THE GRANDFATHER PASSAGE AND TAKE VOICE QUALITY ASSESSMENTS.]

[CHECK SOUND LEVEL OF ROOM and RECORD DB in NOTEBOOK]

- *Hold the phone at a comfortable distance and speak as you normally would.*
- *You are allowed to make corrections, but you are not forced to correct every error.*
- *Pretend that you are sending this to a friend.*

DESCRIPTIONS OF EACH METHOD

(A) Typing (Physical QWERTY)

1. Tap individual letters to make a word.
2. Tap the space bar in between each word.
3. Once you are satisfied, hit the "enter" key.
   To make corrections: hit the "delete" key.

(B) Tapping (On-screen QWERTY)

1. Tap individual letters to make a word.
2. Tap the space bar in between each word.
3. Once you are satisfied, hit the "enter" key.
   To make corrections: hit the "delete" key.

(C) Tracing

1. Glide your finger from one letter to the next to create words.
2. The green line appearing across the keyboard shows your movements, so you are aware that the trace is following your finger movements.
3. Trace each word individually.
4. If a word has two of the same letter (e.g., “hello”), make a loop on that letter.
5. You do not need to enter spaces between words. Lift your finger between the end of one word and start of the next word. The phone will insert spaces for you.
6. Once you are satisfied, hit the “enter” key.
   To make corrections: hit the “delete” key.

(D) Handwriting

1. Print each letter with your finger in the writing area.
2. You can write your characters as a word from left to right, or with letters written on top of each other.
3. To insert a space, hit the space bar.
4. Once you are satisfied, hit the “enter” key.
   To make corrections: hit the delete key.

(E) Voice Recognition

1. Tap the Flame icon to launch speech input
2. Voice recognition will begin automatically after tapping
3. Speak the phrase
4. Tap the screen when you have completed the phrase
5. Once you are satisfied, hit the “enter” key.
   To make corrections: hit the “delete” key.

SATISFACTION & WORKLOAD

[*After gathering performance data with each input method, 20 phrases, bring up the SUS & NASA questionnaires on the monitor in front of the participant*.]

Now I want you to fill out a couple of questionnaires. The first assesses your satisfaction.

Additionally, I will ask about your likes, dislikes and your ideas of how to improve this input method.

The next survey will assess your perceived levels of workload while using this input method to enter phrases.

PREFERENCE RATINGS (POST-task)

[PROVIDE SCREEN SHOT FOR EACH INPUT METHOD, *IN THE ORDER RECEIVED BY THE PARTICIPANT*]

Now that you have used all 5 input applications, I am going to ask you to rate them. ***OVERALL*** *how desirable or valuable do you perceive these input methods to be based on your* ***experience with using them to enter text on a smartphone****? Use the 0 – 50 scale, with 0 being "least preferred" and 50 being "most preferred"?*

Why?

THANK YOU!!!

[Give payment.]

Turn off camera.

Turn off Morae.

Record responses.

APPENDIX M

OUTLIERS RECODED FOR EACH CONDITION

Distributions for each condition for each measure were examined graphically (histograms) and statistically (skewness, kurtosis, tendency and variance). Values that were a standardized value of ± 2.58 were examined and recoded to ± 1 the next most extreme score, as suggested by Tabachnick and Fidell (2007).

Reasons behind the occurrence of outliers were also taken into account. Detailed notes were kept from the study sessions, as well as video. The outliers were not the same people across the dependent measures, otherwise the deletion of individual cases would have been considered. Generally, extreme scores may have been related to unusually good or poor performance, and the way that the participant interacted with the smartphone (method of holding the device and entering text). The extreme error rates may have been a result of decreased motivation or fatigue, even though participants regularly took breaks. Extreme ratings may simply be due to the fact that it was difficult to “fairly” rate the first input method, not having had anything to compare it to (see Tables M1 and M2).

APPENDIX M (continued)

TABLE M1

YOUNGER ADULT OUTLIERS

Measure	Physical Qwerty	Onscreen Qwerty	Tracing	Handwriting	Voice
Uncorrected Error Rate	3.63 to 2.08%	-	-	-	-
Word Error Rate	6.43 to 5.7%	12.09 to 10.68% 11.07 to 9.68%	-	11.68 to 10.29%	10.97 to 8.67%
Satisfaction	60 to 70.5 62.5 to 71.5	-	-	-	-
Workload (mental)	18 to 15	21 to 18	-	-	16 to 9
Workload (physical)	-	-	-	3 to 7	21 to 7
Workload (temporal)	19 to 14 19 to 14	-	-	-	-
Workload (performance)	19 to 8	-	19 to 14	-	-
Workload (effort)	-	-	-	-	17 to 8
Workload (frustration)	17 to 9 13 to 8	-	20 to 16	3 to 5	-
Post-test preference	25 to 30 25 to 30	-	-	-	0 to 9

TABLE M2

OLDER ADULT OUTLIERS

Measure	Physical Qwerty	Onscreen Qwerty	Tracing	Handwriting	Voice
Uncorrected Error Rate	5.97 to 4%	16.47 to 8.78%	15.42 to 7.22%	11.96 to 9.68%	-
Word Error Rate	28.55 to 18.2%	48.55 to 24.2%	33.85 to 14.7%	36.36 to 24.6%	-
Satisfaction	-	-	-	-	20 to 63 37.5 to 64
Workload (mental)	-	-	-	-	18 to 15 16 to 14
Workload (physical)	18 to 13	-	-	3 to 6	20 to 13 19 to 12
Workload (temporal)	-	-	-	-	13 to 11 12 to 11
Workload (performance)	-	-	-	-	13 to 9
Workload (frustration)	16 to 14 15 to 13	-	-	-	20 to 8 19 to 7
Post-test preference	-	-	-	-	7 to 23 10 to 24

APPENDIX N

TABLE N1

SUMMARY OF PARTICIPANT DEMOGRAPHICS

Participant Demographic	**Younger Adult Frequency**	**Older Adult Frequency**
Age	*M* = 24.4, *SD* = 5.6	*M* = 68.8, *SD* = 7.4
Gender	Male = 8, Female = 17	Male = 9, Female = 16
Current occupation	Student = 16, Retail, Unemployed = 2, Athletic Trainer, Secretary, Advocate Care Services, Vet Tech, Fire Fighter = 1	Retired = 15, Administrator = 3, Professor = 2, Librarian, Master Electrician, EHS Director, Social Worker, High School Teacher = 1
Highest level of completed education	High School = 2, Some College = 15, Associate's Degree = 2, Bachelor's Degree = 4, Master's Degree = 2	High School = 2, Some College = 5, Associate's Degree = 1, Bachelor's Degree = 4, Master's Degree = 11, Doctorate Degree = 2
Handedness	Left = 2, Right = 23	Left = 4, Right = 21
Hearing impairments	No = 24, Yes = 1	No = 22, Yes = 3
Speech disorder	No = 24, Yes = 1 (resolved)	No = 24, Yes = 1 (resolved)
History of smoking tobacco	No = 18, Yes = 7 (4 currently)	No = 15, Yes = 10 (None currently)
Past/Current finger, hand or wrist problems	Wrist sprain as a child = 1, Left wrist fracture (past) = 1, Left wrist injury (past) = 1, Ring finger on right hand is ¼ inch shorter than left hand = 1	Mild arthritis = 7, Carpal tunnel surgery (past) = 2, Fractures (past) = 1, Wrist ganglion removal (past) = 1
Currently /Ever owned a mobile phone	Yes = 22, No = 3	Yes = 24, No = 1

APPENDIX O

EXAMPLES OF COMMON ENTRY TECHNIQUES

Younger Adults

Most of the younger adults used the Physical Qwerty with both thumbs, held in both hands with their hands, wrists or forearms supported on the tabletop (see Figure O1). The same was true for the Onscreen Qwerty (see Figure O2). However, the nature interaction is different with Tracing and Handwriting, since only one digit can enter text. For these two text input methods, younger adults generally used a single index finger, holding the device in one hand, and using the tabletop for support of their hands, wrists or forearms (see Figures O3 and O4).

Figure O1. Most common technique used by younger adults to enter text on a Physical Qwerty.

Figure O2. Most common technique used by younger adults to enter text on Onscreen Qwerty.

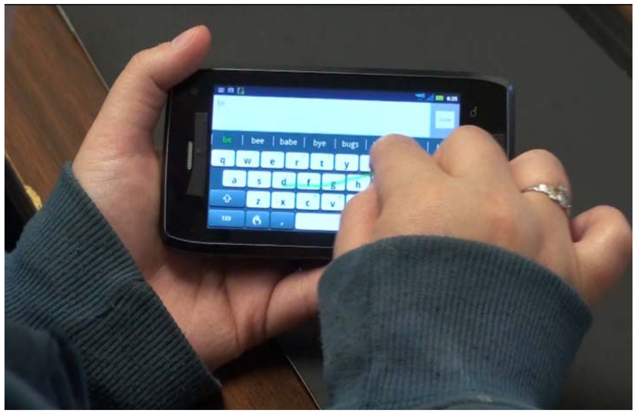


Figure O3. Most common technique used by younger adults to enter text with Tracing.

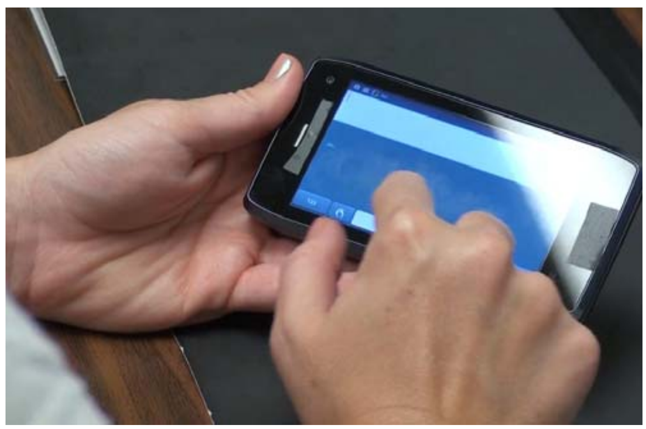


Figure O4. Most common technique used by younger adults to enter text with Handwriting.

Older Adults

Most of the older adults also used both thumbs to enter text with the Physical Qwerty, holding the device with both hands, supported on the tabletop (see Figure O5). However, for the Onscreen Qwerty, Tracing and Handwriting, they typically used an index finger, stabilizing the phone on the tabletop in a type of pinch grip (see Figures O6 through O8). The older adults were more variable in their techniques for holding the device and also for entering text.

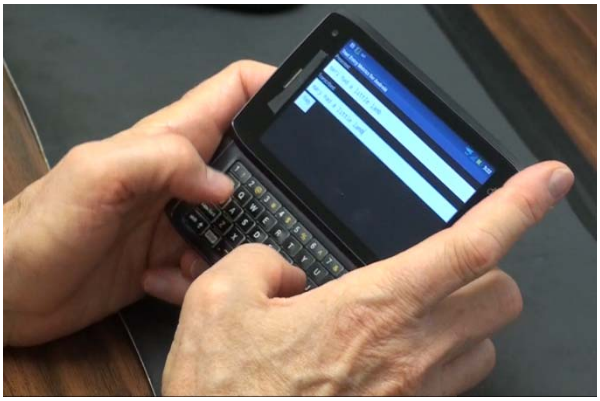


Figure O5. Most common technique used by older adults to enter text on the Physical Qwerty.

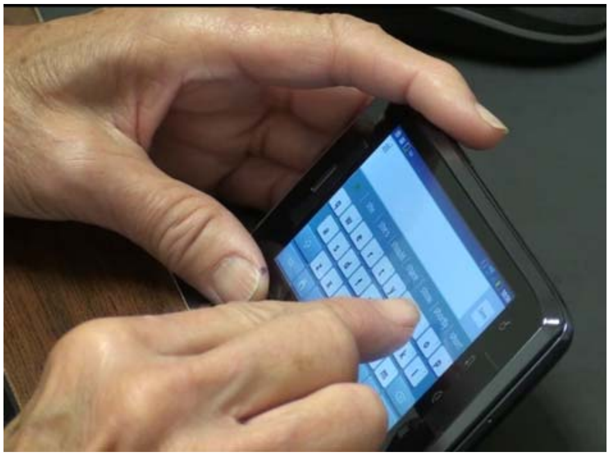


Figure O6. Most common technique used by older adults to enter text on the Onscreen Qwerty.

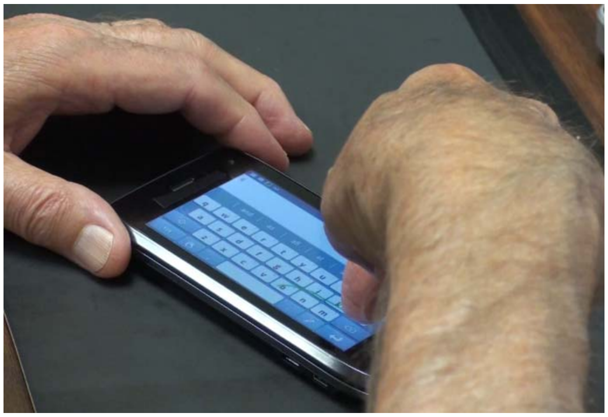


Figure O7. Most common technique used by older adults to enter text with Tracing.

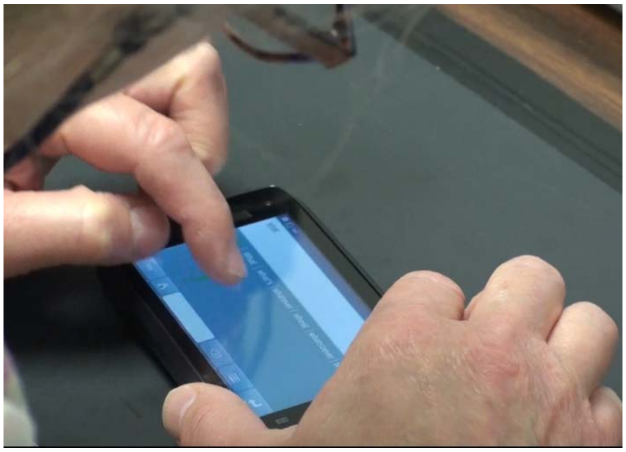


Figure O8. Most common technique used by older adults to enter text with Handwriting.

APPENDIX P

TABLE P1.1

CORRELATIONS BETWEEN MEASURES YOUNGER ADULT FEMALES

	Digit Symbol	ATCQ	Finger Tapping	Jebsen-Taylor	Grooved Pegboard	Keyboard
Attitudes Towards Computers Questionnaire	-.08	-				
Finger Tapping Test	.17	.23	-			
Jebsen-Taylor (dominant)	-.24	-.10	-.19	-		
Grooved Pegboard (dominant)	-.40	.20	-.04	-.24	-	
Keyboarding (WPM)	.05	.11	**.56***	-.20	-.17	-
Hand Length	.38	.19	-.06	.20	.14	-.26
Hand Width	.24	-.01	-.22	.07	.15	-.09
Thumb Length	.37	.40	-.05	.05	.23	-.29
Thumb Width	.37	.19	-.33	.01	-.07	-.04
Thumb Circumference	.28	.07	-.28	-.07	.01	-.10
Index Length	.19	.24	-.15	.11	.20	-.25
Index Width	.25	-.05	-.25	-.22	-.01	.04
Index Circumference	.28	.05	-.19	-.25	.12	-.02
Thumb Reach	.05	-.04	-.03	.12	.26	.09
Functional Range of Motion	-.08	.28	.15	-.32	.10	.04
Speech Rate (WPS)	**.72****	-.20	.23	**-.59***	**-.51***	**-.01**

Note: * indicates statistical significance $p < .05$, ** indicates $p < .01$, both two-tailed, $n = 17$.
Also note that listed values for the Jebsen-Taylor were back-transformed.

TABLE P1.2

CORRELATIONS BETWEEN MEASURES YOUNGER ADULT FEMALES

	Hand Length	Hand Width	Thumb Length	Thumb Width	Thumb Circum	Index Length	Index Width	Index Circum	Thumb Reach	FRoM
Hand Width	**.77****	-								
Thumb Length	**.65****	.40	-							
Thumb Width	**.63****	**.74****	.32	-						
Thumb Circumference	**.66****	**.82****	.24	**.92****	-					
Index Length	**.93****	**.71****	**.66****	**.60***	**.62****	-				
Index Width	.48	**.75****	.09	**.83****	**.94****	.45	-			
Index Circumference	**.53***	**.84****	.17	**.79****	**.90****	**.49***		-		
Thumb Reach	.05	-.24	.29	-.12	-.24	.09	-.29	-.35	-	
Functional Range of Motion	.11	.16	.46	-.15	-.05	.15	-.09	.00	-.03	-
Speech Rate (WPS)	.03	-.03	.21	-.08	-.09	-.10	-.07	.03	-.04	.18

Note: * indicates statistical significance $p < .05$, ** indicates $p < .01$, two-tailed, $n = 17$.

APPENDIX P (continued)

TABLE P2.1

CORRELATIONS BETWEEN MEASURES YOUNGER ADULT MALES

	Digit Symbol	ATCQ	Finger Tapping	Jebsen-Taylor	Grooved Pegboard	Keyboard
Attitudes Towards Computers Questionnaire	.35	-				
Finger Tapping Test	.66	.61	-			
Jebsen-Taylor (dominant)	.29	.26	.16	-		
Grooved Pegboard (dominant)	-.41	**-.76***	-.07	**-.92****	-	
Keyboarding (WPM)	.70	.11	.65	.20	-.23	-
Hand Length	.21	**.78***	.57	.66	-.46	.37
Hand Width	.19	.49	.57	.50	-.17	.28
Thumb Length	-.15	**.81***	.08	.60	-.48	-.09
Thumb Width	-.41	.06	-.07	.01	.34	-.38
Thumb Circumference	-.38	-.10	-.13	.10	.25	-.44
Index Length	.20	**.81***	.47	**.81***	-.70	.35
Index Width	-.45	.19	-.35	.32	-.02	-.50
Index Circumference	-.32	.06	-.39	.39	.30	-.47
Thumb Reach	-.11	.53	.29	.35	-.57	-.19
Functional Range of Motion	-.22	-.58	-.68	-.07	-.01	-.02
Speech Rate (WPS)	**.78***	.47	.61	.26	-.42	.33

Note: * indicates statistical significance $p < .05$, ** indicates $p < .01$, both two-tailed, $n = 8$.
Also note that listed values for the Jebsen-Taylor were back-transformed.

APPENDIX P (continued)

TABLE P2.2

CORRELATIONS BETWEEN MEASURES YOUNGER ADULT MALES

	Hand Length	Hand Width	Thumb Length	Thumb Width	Thumb Circum	Index Length	Index Width	Index Circum	Thumb Reach	FRoM
Hand Width	**.82***	-								
Thumb Length	**.73***	.41	-							
Thumb Width	.29	.51	.31	-						
Thumb Circumference	.06	.50	.11	**.85****	-					
Index Length	**.88****	.58	**.75***	-.03	-.07	-				
Index Width	.24	.51	.41	**.88****	**.86****	.08	-			
Index Circumference	.00	.33	.18	.64	**.87****	.01	**.89****	-		
Thumb Reach	.48	.16	.42	-.19	-.21	.63	-.15	-.23	-	
Functional Range of Motion	-.57	-.42	-.49	-.19	.02	-.44	.06	.30	-.57	-
Speech Rate (WPS)	.13	.13	.09	-.48	-.35	.30	-.49	-.32	.26	-.52

Note: * indicates statistical significance $p < .05$, ** indicates $p < .01$, two-tailed, $n = 8$.

TABLE P3.1

CORRELATIONS BETWEEN MEASURES OLDER ADULT FEMALES

	Digit Symbol	ATCQ	Finger Tapping	Jebsen-Taylor	Grooved Pegboard	Keyboard
Attitudes Towards Computers Questionnaire	.29	-				
Finger Tapping Test	.32	.30	-			
Jebsen-Taylor (dominant)	**-.55***	-.24	**-.51***	-		
Grooved Pegboard (dominant)	**-.57***	-.27	-.42	.29	-	
Keyboarding (WPM)	.44	**.61***	.46	-.28	**-.52***	-
Hand Length	.13	.47	-.03	-.15	.00	.11
Hand Width	.24	-.12	-.12	-.06	-.03	.19
Thumb Length	.02	.23	-.09	-.13	.13	-.16
Thumb Width	-.01	-.02	-.10	.09	.30	.03
Thumb Circumference	-.32	-.10	-.24	-.02	.31	-.23
Index Length	.12	.32	-.21	.01	.16	-.11
Index Width	-.33	-.27	-.04	.09	-.08	-.37
Index Circumference	-.42	-.13	.15	-.16	.19	-.43
Thumb Reach	-.17	**.51***	.24	-.19	.28	.28
Functional Range of Motion	.27	-.37	-.12	.04	.21	-.37
Speech Rate (WPS)	**.52***	.22	.30	-.46	**-.55***	**.68***

Note: * indicates statistical significance $p < .05$, two-tailed, $n = 16$.
Also note that listed values for the Jebsen-Taylor were back-transformed.

APPENDIX P (continued)

TABLE P3.2

CORRELATIONS BETWEEN MEASURES OLDER ADULT FEMALES

	Hand Length	Hand Width	Thumb Length	Thumb Width	Thumb Circum	Index Length	Index Width	Index Circum	Thumb Reach	FRoM
Hand Width	.43	-								
Thumb Length	**.79****	.39	-							
Thumb Width	**.64****	.37	**.56***	-						
Thumb Circumference	-.22	-.45	-.03	.20	-					
Index Length	**.86****	.34	**.84****	**.62***	-.03	-				
Index Width	-.04	-.07	.10	.02	.14	.07	-			
Index Circumference	.11	-.18	.22	.01	-.05	.07	**.67****	-		
Thumb Reach	**.56***	-.01	**.56***	**.54***	.10	.40	-.24	.19	-	
Functional Range of Motion	-.20	.08	.21	-.11	-.17	-.01	-.26	-.01	-.06	-
Speech Rate (WPS)	.08	.26	-.36	.10	-.12	-.20	-.27	-.44	-.13	-.40

Note: * indicates statistical significance $p < .05$, ** indicates $p < .01$, two-tailed, $n = 16$.

TABLE P4.1

CORRELATIONS BETWEEN MEASURES OLDER ADULT MALES

	Digit Symbol	ATCQ	Finger Tapping	Jebsen-Taylor	Grooved Pegboard	Keyboard
Attitudes Towards Computers Questionnaire	.05	-				
Finger Tapping Test	.46	-.18	-			
Jebsen-Taylor (dominant)	-.18	.51	**-.86****	-		
Grooved Pegboard (dominant)	-.48	.18	-.10	-.01	-	
Keyboarding (WPM)	.24	.10	-.53	.63	**-.67***	-
Hand Length	-.17	-.26	.09	-.47	.44	-.49
Hand Width	.12	-.19	.40	-.59	.56	**-.80***
Thumb Length	-.23	-.17	.39	-.42	-.33	-.29
Thumb Width	-.12	.14	.52	-.57	.45	-.64
Thumb Circumference	-.31	.32	.41	-.38	.61	**-.70***
Index Length	-.26	-.13	-.19	-.09	-.33	.08
Index Width	-.60	-.18	.30	-.58	.45	**-.78***
Index Circumference	-.30	.13	.57	-.52	.45	**-.73***
Thumb Reach	.10	.32	-.30	.22	-.10	.01
Functional Range of Motion	.34	.26	.49	-.38	-.24	-.08
Speech Rate (WPS)	.39	.08	.51	-.47	-.43	-.20

Note: * indicates statistical significance $p < .05$, ** indicates $p < .01$, two-tailed, $n = 9$.
Also note that listed values for the Jebsen-Taylor were back-transformed.

APPENDIX P (continued)

TABLE P4.2

CORRELATIONS BETWEEN MEASURES OLDER ADULT MALES

	Hand Length	Hand Width	Thumb Length	Thumb Width	Thumb Circum	Index Length	Index Width	Index Circum	Thumb Reach	FRoM
Hand Width	**.77***	-								
Thumb Length	-.16	-.09	-							
Thumb Width	.62	.54	.16	-						
Thumb Circumference	.41	.43	.08	**.90****	-					
Index Length	.45	.03	.38	.27	-.04	-				
Index Width	.57	.51	.56	.62	.62	.39	-			
Index Circumference	.25	.34	.30	**.84****	**.93****	-.05	.66	-		
Thumb Reach	.36	.29	.04	.13	-.10	.64	.07	-.26	-	
Functional Range of Motion	.25	.14	.05	.57	.40	.37	.14	.40	.20	-
Speech Rate (WPS)	.15	.27	.64	.32	.03	.46	.28	.08	.52	.41

Note: * indicates statistical significance $p < .05$, ** indicates $p < .01$, two-tailed, $n = 9$.