A Comparison of Touchscreen and Mouse for Real-World and Abstract Tasks with Older Adults

SUNG-HEE KIM, Dong-eui University
KAILUN ZHANG, JOANNA MCGRENERE, and KELLOGG S. BOOTH,
University of British Columbia
CLAUDIA JACOVA, Pacific University

16

Computer technology is increasingly being used to facilitate the timely identification of cognitive impairment in older adults. Our **Cognitive Testing on Computer (C-TOC)** project aims to develop a self-administered online test for older adults to take at their home. Due to the freedom of devices they can use, it is important to investigate whether different input devices can impact test performance. We compared touchscreen and mouse input on both abstract and real-world pointing and dragging tasks: classic Fitts's Law tasks and tasks drawn from C-TOC. The abstract and real-world tasks were designed to require equivalent motor skills. Our research goals were to determine (1) if performance on computerized cognitive tasks are affected by input device, and (2) if performance differences due to input device can be explained by those observed on Fitts's Law tasks. Sixteen older adults completed both types of tasks using a touchscreen and a mouse. We found that input device affected speed on three out of four cognitive tasks while only affecting accuracy on one task. Secondarily, our results suggest that Fitts's Law results of differences in mouse and touch cannot be used to predict device differences in the performance on C-TOC tests. As an additional research goal, we looked into the movement patterns of one real-world dragging task—the C-TOC Pattern Construction task—to see if they could provide richer performance measures, beyond speed and accuracy. Such measures could compensate for the lack of a clinician observer who is typically present in comparable paper-based cognitive tests. We found that older adults naturally adopted different movement patterns on the two devices: they tended to make shorter moves and a greater number of moves on a touchscreen than with a mouse. Altogether, our results suggest that careful device-based performance calibration will be needed in computerized tests.

CCS Concepts: • **Human-centered computing** → *Usability testing*; **Pointing devices**; • **Hardware** → **Touch screens**;

Additional Key Words and Phrases: Input device, pointing, dragging, touchscreen, mouse, experiment, older adults, cognitive testing

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under grants RGPIN-116412-11, RGPIN-429595-12, RGPIN-2016-04326, and RGPIN-2017-04549, and by the Graphics, Animation and New Media Network of Centres of Excellence.
Authors' addresses: S.-H. Kim, Dong-eui University, 176 Eomgwang-ro, Busan, 47340, Republic of Korea; email: sh.kim@deu.ac.kr; K. Zhang, J. McGrenere, and K. S. Booth, University of British Columbia, 201-2366 Main Mall, Vancouver, BC, V6T 1Z4, Canada; emails: zhkalu@gmail.com, {joanna, ksbooth}@cs.ubc.ca; C. Jacova, Pacific University, 2043 College Way, Forest Grove, OR 97116; email: cjacova@pacificu.edu.

1936-7228/2020/10-ART16 $15.00
https://doi.org/10.1145/3418057

ACM Reference format:
Sung-Hee Kim, Kailun Zhang, Joanna McGrenere, Kellogg S. Booth, and Claudia Jacova. 2020. A Comparison of Touchscreen and Mouse for Real-World and Abstract Tasks with Older Adults. *ACM Trans. Access. Comput.* 13, 4, Article 16 (October 2020), 26 pages.
https://doi.org/10.1145/3418057

1 INTRODUCTION

The timely identification of cognitive impairment in older adults is considered a health care priority. Such identification relies on objective cognitive testing, ranging from brief screening to comprehensive neuropsychological testing. Traditionally, cognitive testing is done in the clinic and requires trained personnel for administration and scoring. In-clinic cognitive testing can be a burden both for patients and health care professionals. It takes away from already limited consultation times for diagnosis and management of cognitive impairment.

1.1 Traditional Paper-Based Cognitive Testing

Cognitive concerns in older adults have traditionally been assessed with two types of paper-based tests [20]. Brief screening tests are designed to detect and monitor impairment by presenting basic memory, orientation, language, and visuo-constructional tasks. Examples of such tests are the Mini-Mental State Examination (MMSE) [13], the Montreal Cognitive Assessment (MoCA) [32], and the General Practitioner Assessment of Cognition (GPCOG) [3]. In many cases, it is then necessary to determine the etiology of the cognitive impairment, specifically, whether a neurodegenerative disorder is present, and whether it is Alzheimer's disease (AD) or a non-AD dementia. This is done through more comprehensive cognitive testing that may range from intermediate-length batteries, such as the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) [36], to full-length neuropsychological batteries that draw from a multitude of standardized tests to assess in detail each domain of cognitive functioning [20]. This testing approach is validated and reliable but places a burden on clinic resources, and it may not be readily accessible.

1.2 Computer-Based Cognitive Testing

One solution to address limited clinic resources is the administration of cognitive testing using computer technology. Computerized testing has known advantages including automated task presentation, minimization of examiner subjectivity, and the potential of allowing adaptive testing for individuals [49, 53]. Several computer-based cognitive tests have been developed for older adults. These have required a technician to be present in the room for support in a clinic setting on a standard desktop computer, with a keyboard and mouse [9, 11, 39, 45]. In recent years, remote or telehealth approaches to cognitive testing have emerged. One proposed option is to deliver the tests online such that test takers can self-administer the test on their own personal computer in a location of their choosing, such as home [48]. Our own Cognitive Testing on Computer (C-TOC) project, described in Section 2, is aimed at developing, validating, and refining a test battery that older adults can self-administer online at their home prior to attending a clinic visit [22].

A possible concern with this approach is that personal computing has changed significantly in the past 10 years such that the computers that test takers have at home may be quite diverse. Although use of the mouse is still common, touch-based input has quickly gained popularity with the wide adoption of mobile or tablet devices. There is also an increasing preference of touchscreen technology in health care settings as the best interface for older test takers [35]. For example, CAMCI is administered on tablet and CANTAB online or on tablet [28]. Touch tablets such as the iPad are known to be particularly popular with seniors. When people are instructed to take a cognitive test online remotely from home, they will use whatever computing device they have there and are most comfortable with. This means that different people may use different devices, and even the same person may use different devices on different occasions.

These novel testing contexts raise questions regarding the comparability of test results achieved with different input devices, and regarding the applicability of normative data obtained with an input device different from the one used by the test taker. Are test results valid if the baseline performance is assessed with one type of input device and follow-up with a different type of input device? If one test taker completes a cognitive test using a mouse for input and another test taker completes the same test with touch input, to what extent are the results comparable? Eventually, inter-individual and intra-individual comparisons are both clinically very important.

1.3 Speed and Accuracy: Extending from Abstract to Real-World Tasks

The HCI literature should be able to help us answer these questions, at least to some extent. Studies have shown that touch is faster for pointing for the general population and also for older adults [8, 33]. These results are largely based on what are considered to be "abstract tasks" such as those used in Fitts's Law experiments. If these device effects also exist for "real-world tasks," one possible solution would be to calibrate performance on one device to performance on another device. This would require a strong, linear relationship between Fitts's Law task performance differences due to the device and the same device differences on real-world tasks. Fitts's Law is known to be robust for modeling human movement for reaching to a target area, where the area is typically a colored region such as a blue square, a red circle, or a green bar on a white background, and the task is simply to move from a starting position to the target. This type of task is often referred to as an abstract task because it is typically not a real task that one would want to do [30, 34]. It is intended to, for example, mimic the effort to reach to a known widget such as icon on a computer screen, which is a real-world task. Our interest is to make a comparison between the performance on abstract tasks and the performance of people on practical tasks that have application in the real world such as C-TOC tasks (henceforth referred to as real-world tasks). The C-TOC tasks require somewhat more cognitive effort; the motor skills required for pointing or dragging an object are only part of the problem-solving process.

In this study, we address whether performance on Fitts's Law tasks would allow the development of a predictive algorithm for cognitive tasks, which calibrates between input devices. For example, if a particular cognitive test involves selection through pointing alone, can a test taker's performance using a mouse be used to predict how that same test taker would have performed on the same cognitive test using a touch tablet, by applying a speedup based on the size and the distance of the targets? If the prediction is strong, then we could use the results of Fitts's Law tasks to calibrate one device against another, and it would simply be the case that the cognitive test would need to know which device was used and the calibration could be automatic.

1.4 Research Questions

Our primary research goal had two components: (1) to determine if performance on computerized cognitive tasks is affected by input device, and (2) if performance differences due to input device can be explained by those observed on Fitts's Law tasks. To achieve this, we chose four C-TOC subtests that span both pointing and dragging interaction, as well as relatively low- and high-precision tasks. We then mapped these to abstract Fitts's tasks, controlling for index of difficulty throughout. Sixteen older adult participants completed all four of the real-world C-TOC subtests, as well as the abstract tasks that were deemed to be the equivalent, from a motor perspective, using both touchscreen and mouse-based devices.

Because C-TOC is computer based, logging test takers' detailed interaction throughout C-TOC is feasible. This type of data capture may partially compensate for the biggest disadvantage of at home, self-administered computerized testing—the lack of observation from the human examiner who is present during standard paper-based cognitive testing. We were curious to know what other interaction metrics, beyond speed and accuracy, might be available through software logging to evaluate participants' cognitive performance while taking C-TOC and whether these would be device sensitive.

Thus, the secondary research goal was to explore measures other than speed and accuracy that might be valuable for evaluating participants' cognitive performance while taking C-TOC, and to determine whether those measures are device sensitive. In a similar vein, we wanted to clarify any subjective differences in test-taker experience between touchscreen and mouse interfaces.

The contributions of this work are as follows:

- We are the first, to the best of our knowledge, to systematically compare between touchscreen and mouse for computerized cognitive tests, and we find that touch is faster for three of four of the cognitive tests tested but that accuracy is only affected for one test.
- We find that the difference in performance (speed and accuracy) between the two devices cannot be accurately predicted by Fitts's Law performance on tasks with comparable precision level. This cautions against automatic calibration of performance for the device and underscores the complexity in extending Fitts's Law research to computerized cognitive testing.
- We replicate previous Fitts's Law research for both pointing and dragging tasks, reinforcing its applicability to older adults: the touchscreen is faster than the mouse but less accurate in high-precision pointing tasks.
- We uncover considerably different movement patterns between devices in a cognitive dragging task: touchscreen yields nearly 50% more moves compared to the mouse, but this did not translate into differences in total task completion time between the two devices. We further investigate the difference in movement patterns by coding participants' individual dragging moves into a set of categories and find that participants, instead of making just single movements, often separate a move into multiple shorter moves on a touchscreen.

This article reports and extends research described in the masters' thesis of K. Zhang [51], the second author of this article.

2 RELATED WORK

We begin by summarizing research in computerized cognitive testing being developed in the C-TOC Project. Next we cover previous studies in input device comparison among older adults, followed by studies on the effect of aging on performance for specific devices.

2.1 Computerized Cognitive Testing and C-TOC

Cognitive assessments have traditionally been conducted in clinical settings using paper-based tools [27] such as the MoCA [32] and MMSE [13]. These traditional clinic-based cognitive tests require administration and scoring by a clinician or other trained staff, with time and cost implications that may exceed clinic resources. This stands in the way of a major goal in the current care and management of AD and non-AD dementias: to achieve early recognition and manage contributing health issues [37]. A preliminary cognitive test that can be taken independently at home without extra personnel present may ease the process.

Multiple computer-based cognitive tests for older adults have been developed and validated for detection of cognitive impairment [9, 11, 39, 45]. Compared to traditional paper-based cognitive tests, typically these computerized tests do not require a clinician to take on an active role in test administration, but they do rely on a technician being present in the room for support [9]. None of these computerized cognitive test batteries have been validated for at-home self-administration, but recent studies have shown that this testing format is feasible among older adults, and potentially psychometrically sound [38, 47, 49, 53].

The C-TOC Project was designed with an ultimate goal to be a self-administered, web-based, computerized test, which older adults can take independently at home. C-TOC includes test paradigms that require the test taker to provide complex responses moving and assembling stimuli on the screen rather than just acknowledging stimuli by pressing a key or the mouse. C-TOC has been designed under three iterative development cycles to ensure usability among older adults [22]. Three types of intended older-adult end users have been included: those deemed Not Cognitively Impaired (NCI), those diagnosed with Cognitive Impairment Not Dementia (CIND), also referred to as Mild Cognitive Impairment (MCI), and those with mild AD or non-AD dementia. In each development cycle, participants were asked about specific UI features including graphics and images, screen layout, and navigation controls. They were also asked about the feasibility of self-administering each subtest in a unsupervised setting of their own home. Validated against neuropsychological tests, which are considered to be the gold standard in cognitive testing, C-TOC has shown good validity in detecting cognitive impairment [21].

Research on C-TOC has also appeared in the HCI literature. Brehmer et al. [2] studied the impact of varying levels of interruptions on C-TOC test takers' performance, as a first step toward understanding the effect of disruptions that naturally occur in an uncontrolled domestic setting. Haddad et al. [15] designed two different functionally equivalent C-TOC subtest interfaces according to cultural models, then evaluated the impact of those designs on test takers from two different cultures.

As a self-administered test, C-TOC will not have the benefit of a human examiner. Clinicians are always present in paper-and-pencil versions of neuropsychological tests and can collect important behavioral observations to complement test scores [27, 44]. Finding ways both to make up for that missing observational data and provide additional performance measures to complement the raw scores is an important research challenge.

2.2 Older Adults and Device Comparisons

There has been research done on age-related differences in handheld device use, including comparing subjective ratings on perceived ease of device use based on their visual aspects, such as font size, icon size, and feedback [52], as well as actual performance differences for touch and pen interaction [16, 52]. To overcome these age-related differences, researchers have investigated how to better support older adults' computer usage that employs novel or less common input devices, including light pen [7], eye-gaze [31, 42], EZ ball [50], and rotary encoder [40]. Given the C-TOC context, we take a pragmatic approach and only compare two mainstream input devices: mouse and touchscreen.

It is well known that the relative advantage of an input device depends on task and context [5, 7, 40]. Different tasks or even different contexts may require different types of interaction. Findlater et al. [11] found that older adults were 35% faster using a touchscreen compared to a mouse, but speed gain was much bigger in some interaction types (pointing and crossing) than others (dragging and steering). We focus on pointing and dragging—the only two interaction techniques used in C-TOC.

Performing pointing tasks on a touchscreen is known to be significantly faster than using a mouse but much more error prone, not only for the general population [8, 14, 41] but also for older adults: Ng et al. [33] found that pointing on a touchscreen was 100% faster than with a mouse for older adults. However, other research has shown that accuracy on a touchscreen suffers as a result [17, 25, 26, 46]. Touchscreens are especially inaccurate for small target sizes. These require higher-precision movements to select accurately. Kobayashi et al. [25] found a target width of 30 pixels (px) was too small for older adults to point to with a finger. The error rate for a target of this size was 13.6% for iPad and 39% for iPod. Performance did not improve even after a week of practice. However, the same high error rate was not found for target sizes just a bit larger, indicating that performance may not degrade smoothly. Another factor that affected pointing performance other than size was spacing between targets [17]. When the target size was small (5 and 8 mm), the pointing performance was higher with 3-mm-wide target spacing compared to 1 mm.

For dragging tasks, studies have had inconsistent results when comparing the touchscreen and mouse. Findlater et al. [12] found comparable dragging times for older adults, but Wood et al. [50] found that the touchscreen was 40% slower for older adults. However, this difference in results is likely explained by the 32-px icon size of Wood et al. [50], which is too small according to Kobayashi et al. [25]. Toy et al. [46] conducted an experiment on steering tasks that are similar to dragging tasks but differ in that they have an explicit path to follow. They found that for steering task targets, 64 px in size was more effective than 32 px, and shorter-lasting steering tasks for 6 seconds had lower levels of error compared to 12-second tasks [46].

All studies comparing computer input devices, with older adults as the participants, have exclusively used abstract tasks that have little or no cognitive component, with one exception. Rogers et al. [40] used an Entertainment System Simulator to evaluate performance of a touchscreen and a little-known device—a rotary encoder. Neither device was a clear winner.

2.3 Effect of Age and Dexterity on Input Device Performance

Aging typically affects performance with input devices negatively, due to the various functional declines associated with aging, although the degree to which aging affects performance differs across devices. For pointing tasks, aging has a lesser effect on task performance using a touchscreen than using a mouse [19, 33]. To the best of our knowledge, no previous research has studied the effect of aging for dragging tasks.

One of the functional abilities closely related to aging is manual dexterity. Previous studies have tried to isolate the effect of manual dexterity on input device performance. For example, Jin et al. [23] reported that older adults with lower manual dexterity spent significantly longer performing pointing on a touchscreen. The effect of manual dexterity, however, has not been studied in the context of dragging tasks, nor compared between different input devices.

3 TASK DESIGN

We investigated two types of tasks: real-world C-TOC tasks and abstract tasks, each spanning two interaction types (pointing and dragging) and requiring both high and low precision. C-TOC tasks were drawn from actual C-TOC subtests. Abstract tasks were traditional Fitts's Law tasks that were chosen to approximately match the precision required by the corresponding C-TOC tasks. We used the idea of task precision from Fitts's Law, namely index of difficulty (ID), to estimate task precision of the selected C-TOC tasks. ID is calculated from target width (W) and movement amplitude (A). We used the Shannon formulation, $ID = \log_2(A/W + 1)$, recommended by MacKenzie [29]. We start by explaining C-TOC tasks and then how we estimated the ID for each C-TOC task to determine the task precision levels for the abstract tasks.

3.1 Real-World C-TOC Tasks

We selected four C-TOC subtests for the experiment: *Picture-Word Pairs*, *Arithmetic*, *Sentence Comprehension*, and *Pattern Construction*. Figure 1 shows how each subtest corresponds to a precision

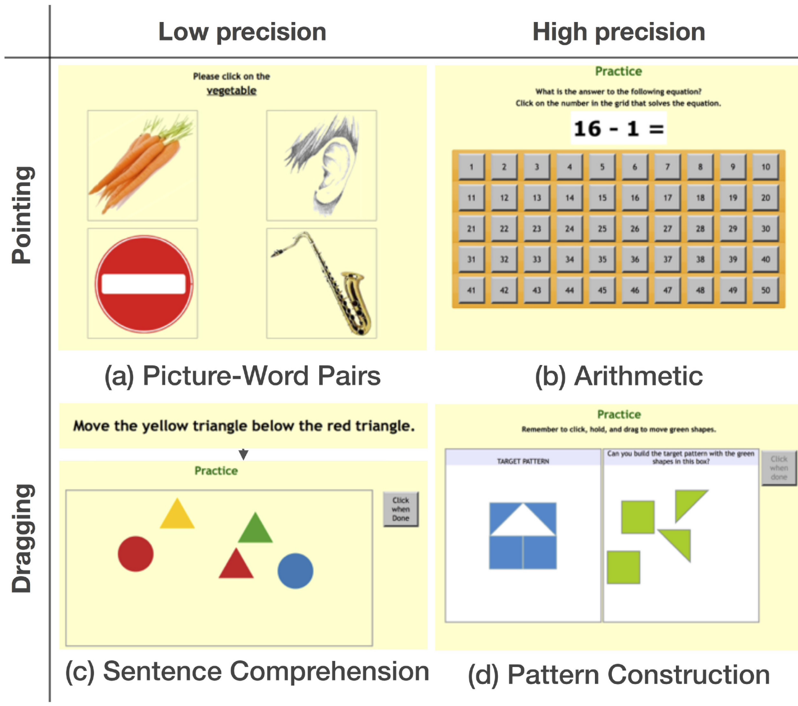

Fig. 1. The four C-TOC subtests used in the experiment. Each subtest corresponds to a task type (pointing or dragging) and a precision level (low or high).

level (low or high) and task type (pointing or dragging). C-TOC scoring for subtests depends on either accuracy alone or a combination of accuracy and speed. In our performance analysis, for simplicity and clarity, we report accuracy and speed individually instead of reporting C-TOC scores. The subtests and estimated task precisions are described in the following sections.

3.1.1 Low-Precision Pointing. The *Picture-Word Pairs* task (Figure 1(a)) is a memory-encoding task. The participant is presented with four images and an instruction such as "Please click on the vegetable." The participant must click/tap on one of the four images, which ends the trial. Each trial starts with the participant clicking an "OK" button in a pop-up window in the middle of the screen. This ensures that the mouse cursor or the finger always starts from the same position. Task precision is $ID \approx 1.0$, which was calculated by the width of the images (250 px) and amplitude, which is the distance between the start and the end point in the task (250 px), where the end point is the center of each object.

3.1.2 High-Precision Pointing. The *Arithmetic* task (Figure 1(b)) assesses numeracy with simple arithmetic problems and four basic operators ($+ - \times \div$). To answer, a grid of clickable buttons corresponding to numbers from 1 to 50 is provided. Each trial starts with the participant clicking an "OK" button in a pop-up window in the middle of the screen (the same as for the *Picture-Word Pairs* subtest) and ends when the participant clicks one of the number buttons. Task precision is $ID \approx 2.5$, based on the width of the number buttons (70 px) and the distance between the start and end point for the task (250–400 px).

3.1.3 Low-Precision Dragging. The *Sentence Comprehension* task (Figure 1(c)) assesses working and immediate memory. It has two stages: (a) memorize the instruction given on the screen, such

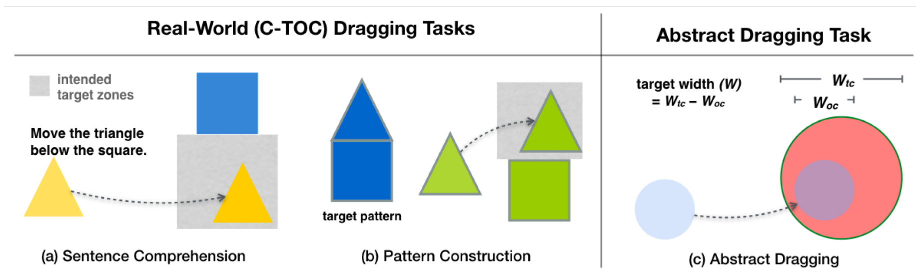

Fig. 2. For real-world dragging tasks, we defined an intended target zone. The intended target zone size is larger in the low-precision task (a) *Sentence Comprehension* compared to the high-precision task (b) *Pattern Construction*. The abstract dragging task (c) is adjusted to be comparable to the real-world tasks, in which participants were asked to drag the blue object circle (W_{oc}) fully into the red target circle (W_{tc}).

as "Move the yellow triangle below the red triangle" and then click the "Next" button to transit to the second stage on a new screen; (b) then, among the movable shapes, drag the shapes as per the instruction and then click the button "Click when done." Two types of time periods were measured: (1) total task completion time, which is the period between clicking "Next" and clicking the "Click when done" button, and (2) individual times for each dragging movement. Task precision is calculated based on width of the intended target zone (Figure 2(a)) and movement amplitude. The width of the movable shapes is 80 to 100 px, the width of intended target zones is 200 to 400 px, and movement amplitude is 200 to 400 px, varying across trials. The width of intended target zones and the amplitudes were verified in pilot tests. The intent of the task was to check verbal comprehension of instructions (e.g., "Move the yellow triangle below the red triangle."). Participants could place the yellow triangle close, but not necessarily perfectly underneath, the red triangle to be scored as correct. The pixel range was determined to capture the level of accuracy commensurate with checking whether participants cognitively understood the instructions for each question. Task precision is $ID \approx 1.0$. Older adults have large variance in perceptual and cognitive abilities [4], so perceived width of intended target zone can vary across each individual, resulting in discrepancies estimating task precision.

3.1.4 High-Precision Dragging. The *Pattern Construction* task (Figure 1(d)) is a visuospatial test. The participant is asked to drag a set of movable shapes to match a reference target pattern that remains visible throughout the test (Figure 2(b)). Shapes can be translated but not rotated. Two types of time periods were measured: (1) total task completion time that starts when the screen appears showing the target pattern and movable shapes, and ends when the participant clicks the "Click when done" button, and (2) individual times for each dragging movement. Due to flexibility in constructing patterns with multiple objects, the precision required for a specific dragging movement could be low or high, but the maximum precision is estimated as $ID \approx 2.5$. The width of movable shapes is 80 to 160 px. For high-precision movements, the intended target zone is 0 to 30 px wider than the movable shape. Movement amplitude is up to 150 to 200 px.

3.2 Abstract Tasks

Abstract tasks were multi-directional pointing and dragging tasks, implemented based on ISO:9241-400 [18] (Figure 3). For pointing, the participant is asked to click or tap on a target object. For a dragging task, we modified the standard Fitts's law dragging task. The participant is asked to drag an object circle fully into, as opposed to partially overlapping with, a target circle to successfully complete the task (Figure 2(c)). The circumference of the target circle highlights in

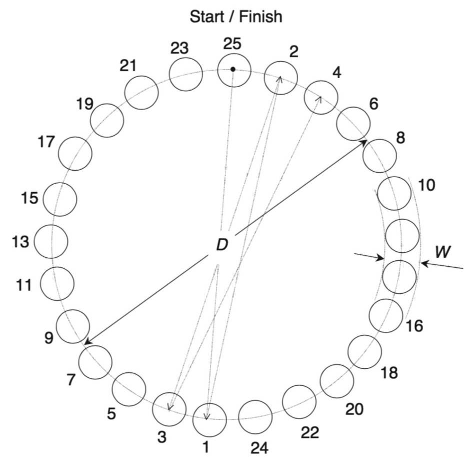

Fig. 3. Paradigm for multi-directional abstract pointing and dragging tasks. Figure copied from Soukoreff and MacKenzie [43].

green once the object circle is fully within the target circle. The modification was to better mimic C-TOC dragging tasks in which participants drag an object shape into an intended target zone. This circle was shown on a white background, so it would have appeared black/dark grey for any participants with color vision deficiencies. For dragging, the target width (W) is defined as the difference between the object circle width (W_{oc}) and target circle width (W_{tc}).

For both the pointing and dragging abstract tasks, amplitude (A) is 250 px to ensure that the largest target circle (approximately 250 px) fits within the tablet (iPad) screen (768 px on the short edge). We determined W values from the amplitude 250 px and the set of *ID*s that reflect task precisions in C-TOC subtests. We had two object widths (50 and 80 px) in the dragging tasks so we could test for an effect of object width.

We chose three task precisions: *ID*s of 1.0, 2.5, and 3.0. The first two approximated the precisions in the low- and high-precision C-TOC tasks. $ID = 3.0$ was included to cover a wider range for trend analysis. An *ID* higher than 3.0 was excluded because the target width would be too small for touchscreens [25]. Although all *ID*s are considered low precision compared to typical *ID* values of 2 to 8 for abstract tasks [18], interfaces designed for older adults typically require lower task precision (i.e., larger targets) compared to interfaces for the general population. For each precision, we added two close variants (e.g., for $ID = 1.0$, 0.9, and 1.1 were added) for a total of nine *ID*s (*A*-*W* pairs), to allow flexibility in estimating task precision and to ensure sufficient power for regression modeling.

4 METHOD

This section discusses the detailed methodology used in the experiment. Tasks have already been described in Section 3.

4.1 Participants

Sixteen people (10 female) ages 57 to 88 years ($M = 71.81, SD = 9.60$) participated in the study. The recruitment flyer was posted throughout the community. Locations included the UBC campus, Vancouver Public Library branches, Vancouver community and senior centers, and senior housing complexes. All participants were right-handed, none with any diagnosed cognitive impairment.

Table 1. Summary of Purdue Pegboard Score for All Participants

	Min	Max	Mean	Std. Dev.
Right hand	8	17	12.25	2.46
Left hand	8	17	11.38	2.47
Both hands	5	15	9.63	2.55
Assembly	16	40	25.44	7.31

Table 2. Summary of Frequency of Device Usage for All Participants

	Mouse	Touchscreen
Several times per day	8	6
Once or twice per day	0	1
Several times per week	2	1
Once or twice per week	1	1
Less than once per week	5	7

We used the participants' score on the first Purdue Pegboard Test as an indication of their manual dexterity (Table 1 presents a summary of detailed scores).

Fourteen participants reported no conditions that would affect motor ability. Two participants reported having arthritis, but their Pegboard results were above the normative data for their age group [10], so we included their data. All participants were right-handed, and they used only their right-hand during the experiment.

No participant had a significant drop in Pegboard Test score after completing the experiment, indicating fatigue was not likely an issue. Results from the eyesight test showed no visual deficiency for any participant that might affect performance. Five out of 16 participants reported that they use a computer with a mouse less than once per week; seven out of 16 participants reported that they use a touch-based device less than once per week (Table 2).

4.2 Apparatus

The experiment was implemented in JavaScript, HTML, and PHP and built with the Raphaël vector graphics library. It ran on an iPad 4th-generation (touchscreen condition) and a 13-inch MacBook Pro with a Logitech Wireless Mouse M310 (mouse condition). Both devices had retina displays with resolution of 1024×768 px (iPad) and 1280×800 px (MacBook Pro). The experiment was run in the Safari browser on both devices (version 8.0 on iPad under iOS 8.3 and version 8.0.4 on MacBook under OS X Yosemite). The experiment was held in an office. The iPad was set in landscape orientation and tilted at a fixed 20-degree angle for all participants. Once the participant located the iPad comfortably at the beginning, we recommended not moving the position of the device during the experiment. During the experiment, we recorded the screen of the devices, participants' hands interacting with the touchscreen, and audio of the interview sessions.

4.3 Design

The experiment included four factors: task, input device, interaction type, and task precision. Each participant completed (1) eight C-TOC subtest conditions: 2 (touchscreen vs. mouse) $\times$ 2 (pointing vs. dragging) $\times$ 2 (precision levels), and (2) four abstract task conditions: 2 (touchscreen vs. mouse) $\times$ 2 (pointing vs. dragging); each abstract task condition contained nine precision levels (0.90, 1.00, 1.10, 2.40, 2.50, 2.60, 2.90, 3.00, 3.10) that were fully randomized across six repetitions in pointing and three repetitions in dragging with two object sizes, which achieved the same number of trials for each interaction type.[1] Optional break times were evenly distributed within each condition. The orders of task, input device, and interaction type were fully counterbalanced. For C-TOC tasks, the order of task precision was fully counterbalanced, and we fully randomized two isomorphic sets of trials to ensure that participants did not see the same trials in both the touchscreen and the mouse conditions.

Although the experimental design included four factors, our primary interest was to understand the effect of device on the factors of task (abstract vs. real world), interaction type (pointing vs. dragging), and precision level. We were not interested in directly comparing interaction types to each other (it is well known that pointing is faster than dragging) nor in directly comparing the C-TOC tasks to each other, as they involve very different types of cognitive skills.

4.4 Procedure

After signing a consent form, a participant completed a demographic questionnaire about age, gender, motor and visual impairments, and frequency of computer usage. The frequencies of touchscreen and mouse usage were collected using 5-point Likert scales as part of the questionnaire, followed by administration of the Purdue Pegboard Test and the Snellen Vision Test to measure manual dexterity and eyesight, respectively.

Participants used one device for all tasks before switching to the other device. Within each device, they first performed all tasks of a single interaction type (pointing or dragging) before tasks of the other interaction type (counterbalanced, as noted earlier). Participants alternated between abstract and C-TOC tasks (counterbalanced). They had practice trials throughout and were offered breaks between each task.

After completing all trials, the Purdue Pegboard test was administered a second time to check for fatigue. A session concluded with an interview asking for the preferred device for each task and why it was preferred. Total duration of a study session was approximately 1.5 hours.

5 RESULTS

5.1 Performance Comparison Between Input Devices

Our main interest is to investigate whether performance on computerized cognitive (real-world) tasks is affected by the input device by analyzing speed and accuracy on C-TOC tasks. We also analyze these measures for Fitts's Law abstract tasks to see if our results are consistent with previous findings, as well as to enable us to compare performance on the real-world tasks and abstract tasks.

5.1.1 C-TOC Tasks.

5.1.1.1 Speed. Pointing task. For low-precision pointing (*Picture-Word Pairs*), participants performed 32% faster on a touchscreen than with a mouse (1,696 ms vs. 2,246 ms, $F_{1,15} = 9.9, p = .006, \eta_G^2 = .40$). For high-precision pointing (*Arithmetic*), there was a trend with large effect size that using a touchscreen was about 12% faster than using a mouse (5,602 ms vs. 6,273 ms, $F_{1,15} = 3.8, p = .069, \eta_G^2 = .20$).

Dragging task. For both time measures in the dragging tasks, mean time on a touchscreen was always the same or faster compared to a mouse, but not all comparisons were significant. For low-precision dragging (*Sentence Comprehension*), task completion time was significantly faster on a

[1] Dragging tasks have two object sizes with each target width (W), whereas pointing tasks have only one object size in each W. To achieve the same total number of trials per task, dragging tasks have half the number of repetitions compared to pointing tasks.

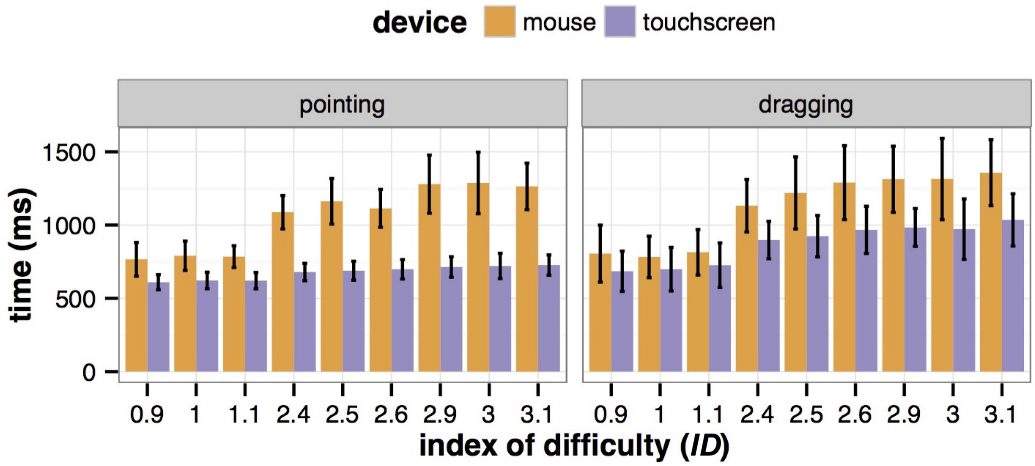

Fig. 4. Speed for abstract tasks by *device* and *task precision*. Error bars show the 95% confidence interval.

touchscreen compared to a mouse (9.7 seconds vs. 10.9 seconds, $F_{1,15} = 4.76, p = .04, \eta^2_G = .24$). But there was no significant effect of device in duration of individual dragging moves (1.2 seconds vs. 1.4 seconds, $F_{1,15} = 1.98, p = .18, \eta^2_G = .12$). For high-precision dragging (*Pattern Construction*), it was the opposite. Duration of individual dragging moves was significantly faster on a touchscreen (1.2 seconds vs. 2.2 seconds, $F_{1,15} = 18.80, p = .001, \eta^2_G = .63$), with a 83% increase of time on a mouse. As will be explained in Section 5.3, we found that participants made moves of shorter distance on the touchscreen and longer distance with the mouse. When we normalized for distance moved, the actual speed difference between devices is reduced to 53%. Despite this difference in speed, there was no significant difference in total task completion time (both 62 seconds, $F_{1,15} < 0.01, p = .99, \eta^2_G < .01$).

5.1.1.2 Accuracy. Pointing and dragging tasks. High-precision pointing (*Arithmetic*) was the only C-TOC task that had a significant difference of accuracy between devices (90% for mouse vs. 82.5% for touch, $F_{1,15} = 5.87, p = .028, \eta^2_G = .28$). There were no significant differences for low-precision pointing (*Picture-Word Pairs*) ($F_{1,15} = 1, p = .33, \eta^2_G = .06$), low-precision dragging (*Sentence Comprehension*) ($F_{1,15} = 0.03, p = .86, \eta^2_G < .01$) or high-precision dragging (*Pattern Construction*) ($F_{1,15} = .62, p = .44, \eta^2_G = .04$).

5.1.2 Abstract Tasks. We describe the outlier removal process, then the results for speed, accuracy, and the regression analysis for a Fitts's model. Two criteria were used for detecting spatial outliers. We eliminated trials in which movement amplitude was (1) less than half the trial amplitude, or (2) greater than three standard deviations from the mean distance for the set of trials with the same participant, *device*, *task*, and *target width*. Outliers accounted for 1.9% of all trials. In the following sections, analyses of speed and error exclude all outliers. For the *pointing task*, we used a 2×9 RM-ANOVA: *input device* by *task precision* (nine *IDs*). For the *dragging task*, we used a $2 \times 2 \times 9$ RM-ANOVA for factors *input device*, *object width*, and *task precision*.

5.1.2.1 Speed. Pointing task. For the pointing task, the decrease in speed as task precision increased was significantly larger for the mouse than for the touchscreen: a *device* × *task precision* interaction dominated ($F_{8,120} = 31.16, p < .001, \eta^2_G = .33$). Overall, touch was faster: main effect of *device* ($F_{1,15} = 75.90, p < .001, \eta^2_G = .73$). As *task precision* increased, speed decreased (Figure 4,

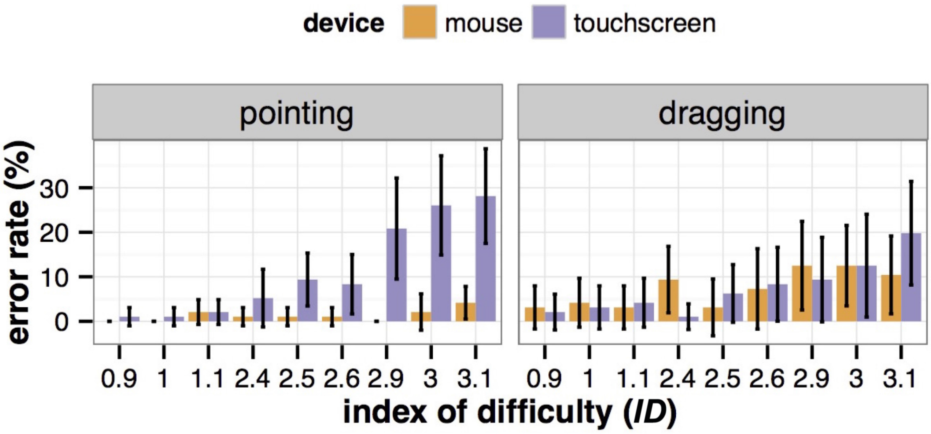

Fig. 5. Error rate for abstract tasks by *device* and *task precision*. Error bars show the 95% confidence interval.

left): main effect of *task precision* ($F_{8,120} = 72.25, p < .001, \eta_G^2 = .53$). Post hoc tests revealed that the touchscreen had no significant increase on pointing time across all *ID* levels, and it was always faster than the mouse at the same *ID* level ($p < .001$). Pointing using a mouse for *ID* higher than 2.4 was slower than pointing using touchscreens regardless of *ID* ($p < .001$).

Dragging task. For the dragging task, similar to pointing, the decrease in speed as task precision increased was greater for the mouse than for touchscreen: a *device* × *task precision* interaction dominated ($F_{8,120} = 5.60, p < .001, \eta_G^2 = .07$). Overall, touch was fastest: main effect of *device* ($F_{1,15} = 18.84, p < .001, \eta_G^2 = .27$). As *task precision* increased, speed decreased (Figure 4, right): main effect of *task precision* ($F_{8,120} = 61.18, p < .001, \eta_G^2 = .45$). Larger objects yielded faster speed (975 ms) compared to smaller objects (1,016 ms): main effect of *object width* ($F_{1,15} = 5.35, p = .035, \eta_G^2 = .01$). There was no significant interaction between object width and the other two factors. Post hoc tests revealed that the touchscreen had no significant increase of time once $ID \geq 2.4$. The touchscreen and mouse had comparable speed for $ID \leq 1.1$ but significantly faster speed using a touchscreen after $ID \geq 2.4$ ($p < .001$). Dragging using a mouse for *ID*s higher than 2.5 is slower than dragging using a touchscreen regardless of the *ID* ($p < .001$).

5.1.2.2 Accuracy. Pointing task. The overall error rate was 5.59% for the pointing task, but it was dependent on *input device* and *task precision*: a *device* × *task precision* interaction dominated ($F_{8,120} = 7.56, p < .001, \eta_G^2 = .18$). There was a relatively constant error rate across precision levels for the mouse but an abrupt increase in error rate as *task precision* increased for the touchscreen. The touchscreen was particularly inaccurate for high-precision tasks in which target width was around 4 mm (33–40 px). Overall, touch had more errors: main effect of *device* ($F_{1,15} = 20.28, p < .001, \eta_G^2 = .20$). As precision increased, so did the rate of errors (Figure 5, left): main effect of *task precision* ($F_{8,120} = 12.50, p < .001, \eta_G^2 = .23$). Post hoc tests showed the error rate for the touchscreen pointing in high precision ($ID \geq 2.9$) was always significantly higher than the error rate in (1) any other precision level of touchscreen pointing and (2) all mouse pointing regardless of task precision ($p < .001$).

Dragging task. The overall error rate was 4.24% for dragging tasks. The error rate increased as *task precision* increased, independent of input device (Figure 5, right): a main effect of *task precision* ($F_{8,120} = 4.89, p < .001, \eta_G^2 = .08$). There was also an interaction between *object width* and *device* ($F_{1,15} = 6.37, p = .02, \eta_G^2 < .01$) and an interaction between *task precision* and *device*

Table 3. Regression Analysis for Each Device-Task Combination

Device	R^2	Intercept (ms)	Slope (ms/bit)	Throughput (bits/s)
		Regression Coefficients		
Pointing				
Mouse	0.97	488	198	5.1
Touchscreen	0.92	412	93	10.8
Dragging				
Mouse	0.97	472	234	4.3
Touchscreen	0.86	414	159	6.3

Note: Throughput is calculated by $1/b$, where b is the slope of the model.

Table 4. Performance Comparison on Task Type and Precision Level

<table>
<tr><th colspan="2" rowspan="2">Precision Level</th><th colspan="2">Pointing</th><th colspan="2">Dragging</th></tr>
<tr><th>Low</th><th>High</th><th>Low</th><th>High</th></tr>
<tr><td rowspan="2">Speed</td><td>C-TOC</td><td>Touch faster* (p = .006, η^2_G = .40)</td><td>Marginal, touch faster (p = .069, η^2_G = .20)</td><td>No effect (p = .18, η^2_G = .12)</td><td>Touch faster* (p = .001, η^2_G=.63)</td></tr>
<tr><td>Fitts's Law</td><td colspan="2">Touch faster* (p < .001, η^2_G = .73)</td><td colspan="2">touch faster* when $ID \geq 2.4$ (p < .001, η^2_G = .27)</td></tr>
<tr><td rowspan="2">Accuracy</td><td>C-TOC</td><td>No effect (p = .33, η^2_G = .06)</td><td>Mouse higher* (p = .028, η^2_G = .28)</td><td>No effect (p = .86, η^2_G < .01)</td><td>No effect (p = .44, η^2_G = .04)</td></tr>
<tr><td>Fitts's Law</td><td colspan="2">Mouse higher* when $ID \geq 2.9$ (p < .001, η^2_G = .20)</td><td colspan="2">No effect (p < .001, η^2_G = .08)</td></tr>
</table>

*Shows statistical significance with $p < 0.05$.

($F_{8,120} = 2.41, p = .02, \eta^2_G = .02$). However, both interactions had very small effect sizes and thus need careful interpretation. Unlike the pointing task, we did not find a dramatic increase in error rate for high-precision tasks on the touchscreen, possibly because during each dragging move, participants could continuously see and adjust the circle they grabbed with their finger until it reached the desired position, whereas each pointing move was a task with a single and instant attempt.

5.1.2.3 Regression Analysis. Pointing task. A regression was performed of time on the effective index of difficulty (ID_e) that had been computed from W_e and A_e. As expected, the touchscreen was much more efficient (107%) than the mouse in pointing tasks in ID ranges from 1 to 3, with throughput of the touchscreen (10.8 bits/s) double the throughput of the mouse (5.1 bit/s). The R^2 values are reported in Table 3.

Dragging task. The performance was 46% more efficient for the touchscreen than with the mouse in dragging tasks with ID ranges from 1 to 3. Throughputs for the touchscreen and mouse were 6.3 bits/s and 4.3 bits/s, respectively. Similar to pointing tasks, we computed ID_e according to the adjusted W_e and A_e. There were high correlations between time and ID_e for the mouse ($R^2 = 0.97$). R^2 for the touchscreen was slightly lower at 0.86 (Table 3).

5.1.3 Summary. We summarize in Table 4 the impact of input device on speed and accuracy for each interaction type (pointing and dragging) and each precision level (low and high) for the real-world C-TOC tasks and the abstract Fitts's Law tasks. First considering speed, we found that

Table 5. Correlation Analysis Between C-TOC and Fitts's Law Tasks for Time Measurements ($N = 16$)

Pointing		Dragging	
Low	High	Low	High
$R^2 = .46, p = .073$	$R^2 = .078, p = .774$	$R^2 = .29, p = .276$	$R^2 = 0.32, p = .227$

input device affects speed on three of four cognitive tasks (note the borderline p-value for *Arithmetic*, the second task). More specifically, for pointing tasks, people are overall faster on a touch screen than with a mouse, and for dragging tasks, as the precision level gets higher, touch is faster. Second, for accuracy, the input device affects only on one of four cognitive tasks: the *Arithmetic* task (high precision) had higher accuracy with a mouse than touch. For the other three cognitive tasks, accuracy does not differ significantly based on input device. These findings fully parallel the findings on Fitts's Law tasks, as can be seen in the table. In other words, the impact of the device on cognitive tasks is on the whole in line with those on Fitts's Law tasks. Note that the cognitive task for C-TOC had two levels of precision (low and high) and for the Fitts's Law task there were three levels of task precisions (i.e., *ID*s of 1.0, 2.5, and 3.0), and *ID*s of 1.0 and 2.5 reflect the precision in the low- and high-precision C-TOC tasks (see Section 3.2).

5.2 Analysis on Factors to Predict Input Device Performance Difference

5.2.1 Fitts's Law Tasks. We wanted to investigate whether the performance differences on C-TOC tasks due to the input device could be explained by those differences observed on Fitts's Law tasks. Therefore, we operationalized the input device effect as the difference for each individual between the individual's task completion time using a mouse and that using a touchscreen. For each participant, ($Mouse_{Timespent} - Touch_{Timespent}$) was calculated, which gives an individual impact of device for both C-TOC tasks and for Fitts's Law tasks. Note that the cognitive task had two levels of precision (low and high), and for the Fitts's Law task, *ID*s of 1.0 and 2.5 reflect the precision in the low- and high-precision C-TOC tasks (see Section 3.2). Therefore, a Pearson correlation coefficient was computed only for these two levels so that we can understand the strength of the relationship between the cognitive task and the Fitts's Law task.

The Pearson's coefficient, R^2 value, and significance values between C-TOC and Fitts's Law Tasks are reported in Table 5. From this analysis, we conclude that we cannot take Fitts's Law results of difference in mouse and touch to predict the device difference in the performance on C-TOC tests. There is a systematic relationship between device differences on cognitive and Fitts's Law tasks, but it is insufficient to make predictions regarding device differences on the cognitive tasks. More specifically, less than 50% of the variance (7.8% to 46%) in difference on cognitive tasks is accounted for by performance on the comparable Fitts's Law tasks. There is no noticeable difference in predictive power between pointing and dragging: correlations appear to be of similar magnitude and not sufficient for accurate predictions. The correlation between mouse and touch seems to disappear when cognitive difficulty increases, as in the *Arithmetic* task, which was the pointing task with high precision in C-TOC.

5.2.2 Motor Dexterity Test. We also analyzed whether dexterity can explain different performance for individuals for C-TOC tasks and Fitts's Law tasks (Table 6). The Pegboard data were based on right-hand performance, and these data were used for the regression analysis. Tasks were also performed with the right hand only—the dominant hand for all participants. The R^2 values suggest that we can predict some of the variance in device differences based on motor dexterity, but this seems to apply more for pointing than for dragging (for both cognitive and Fitts's Law tasks).

Table 6. Correlation Analysis with Motor Dexterity Test and Task Performances ($N = 16$)

	Pointing		Dragging	
	Low	High	Low	High
C-TOC Tasks	−0.609* ($R^2 = 0.370, p = .047$)	0.018 ($R^2 = 0.000, p = .947$)	−0.323 ($R^2 = 0.104, p = .222$)	−0.296 ($R^2 = 0.088, p = .266$)
Fitts's Law Tasks	−0.503* ($R^2 = 0.253, p = .047$)	−0.608* ($R^2 = 0.369, p = .012$)	−0.102 ($R^2 = 0.010, p = .707$)	−0.344 ($R^2 = 0.119, p = .192$)

For C-TOC tasks specifically, motor dexterity accounts for variance in device differences on cognitive tasks requiring low-precision pointing but not on tasks requiring high-precision pointing.

5.3 Test Performance Beyond Speed and Accuracy

Differences due to device need not be limited to speed and accuracy. One of the most important limitations of self-administered computerized cognitive testing is the lack of qualitative performance data that an examiner might collect, such as the number of attempts made, trial and error approach to task, and tendency to undo work. Such behaviors have established clinical significance in paper-and-pencil tests. There have been relatively recent proposals to log computer interaction behaviors, such as mouse trajectories, key-stroke intervals, and work area on screen, as analogues to paper-and-pencil test behaviors [24]. Here we were interested in the moves an individual makes during dragging to see whether their interaction patterns were comparable across devices. We note that to date such interaction measures have not been utilized in cognitive testing and that their development is work in progress.

We make a first attempt at this for the C-TOC dragging task *Pattern Construction*. Due to the considerable flexibility in completing the task, *Pattern Construction* exhibited multiple behaviors because participants could use different strategies to construct a target pattern.

We observed that participants seemed to make more dragging moves with the touchscreen than with a mouse. Analysis of the data showed that there was indeed a significant difference across devices (mean of 24 moves vs. 15 moves per trial, $F_{1,15} = 13.54, p = 0.002, \eta^2_G = .47$).

Comparing the mean for distances, participants also had shorter moving distances with the touchscreen compared to the mouse (50 px vs. 106 px, $F_{1,15} = 19.19, p < .001, \eta^2_G = .56$).

We wondered why participants performed more dragging moves on a touchscreen than with a mouse. We examined the log data and video recording for *Pattern Construction* and generated a set of categories for the dragging moves. We coded individual dragging moves into these categories to see if the classification would reveal any differences in movement pattern between devices.

5.3.1 Coding Method. Each logged dragging move was matched with its paired screen and hand movement video recordings. Based on the videos, we added a *fail-to-grab* category, which was an important movement that was not always captured by the log data.

The second author coded six trials for multiple participants on both devices and designed the coding scheme. A second rater used the scheme and independently coded the same six trials. The inter-rater reliability was found to be good, with kappa = 0.83. The two raters slightly modified the coding scheme after validation. The second author then coded the rest of the trials. We selected two trials from *Pattern Construction* to code—one more complex than the other. In total, 64 trials were coded (2 trials × 2 devices × 16 participants).

Table 7. Summary of the Total Number of Each Type of Move for Each of the Devices

Type of Moves		Total Number of Moves			
		Mouse	Touch	ANOVA	
Single move		273	205	$p = .09$	$\eta_G^2 = .17$
Sub-move		45	376	$p < .001$	$\eta_G^2 = .63$
Precision adjustment		93	108	$p = .63$	$\eta_G^2 = .02$
Trial & error		38	9	$p = .06$	$\eta_G^2 = .21$
De-construction		14	7	$p = .22$	$\eta_G^2 = .10$
Make-way move		53	70	$p = .42$	$\eta_G^2 = .04$
Rotation attempt		0	4	$p = .16$	$\eta_G^2 = .13$
Accidental click		18	1	$p = .06$	$\eta_G^2 = .21$
Constrained move		19	1	$p = .007$	$\eta_G^2 = .39$
Unknown		19	31	$p = .26$	$\eta_G^2 = .08$
Fail-to-grab	Logged	6	24	$p < .001$	$\eta_G^2 = .56$
	Unlogged	15	109		
Total Number of Logged Moves		565	836	$p = .002$	$\eta_G^2 = .47$

Note: Logged fail-to-grab attempts are ones in which participants tap on a shape or click with cursor on a shape. Unlogged fail-to-grab attempts are ones in which participants tap or click on the canvas but not within a shape.

5.3.2 Classification of Types of Moves. We classified each dragging move into one of the following nine categories:

(1) *Target-oriented move* is when participants move a shape to a specific target position. There are three types of moves under this category: *single move, sub-move*, and *precision adjustment*.
   (a) *Single move* is when participants move a shape directly to the target position.
   (b) *Sub-move* is a step in a sequence of two or more steps that together move a shape to the target position.
   (c) *Precision adjustment* is a move for fine tuning the precise location of a shape that was largely already in target position (and is not in a sequence of *sub-move*).
(2) *Trial & error move* is when participants attempt to move a shape to a target position, realize that the position is incorrect before releasing, and either attempt at a new target position or move the shape aside.
(3) *De-construction move* is when participants move one shape out of the already built pattern.
(4) *Make-way move* is when participants move a shape away from its current position to make room for other shapes.
(5) *Rotation attempt* only happened on the touchscreen. It is an action in which participants drag the mouse or their finger in a circular trajectory with the intention to rotate a shape. This is often accompanied with verbal articulation.
(6) *Accidental click* is when participants click unintentionally.
(7) *Constrained move* is when participants try to move a shape beyond the canvas boundary.
(8) *Unknown move* is a move logged by the system whose intention could not be inferred.
(9) *Fail-to-grab* is when participants attempt to grab a shape with mouse cursor or finger but fail to do so.

Table 7 gives a summary of the total number of moves by category across devices. We found that *sub-moves* and *fail-to-grab* move contributed the most toward the high count of dragging moves on touchscreen among all computer-logged moves.

5.3.3 Sub-Move on a Touchscreen. Participants were much more likely to separate a single *target-oriented move* into smaller consecutive moves (*sub-moves*) on a touchscreen device compared to a mouse ($F_{1,15} = 26.46, p < 0.001, \eta_G^2 = .63$), resulting in a shorter distance on a touchscreen and longer distance with a mouse. Apart from the decomposition of dragging moves on the touchscreen, the opposite is observed on the mouse: participants tend to make more *single moves* ($F_{1,15} = 3.26, p = 0.09, \eta_G^2 = .17$) and *trial & error* ($F_{1,15} = 4.05, p = 0.06, \eta_G^2 = .21$) using a mouse. Both *single moves* and *trial & error* require participants to drag and hold a shape to attempt one or more target positions without releasing the shape in the middle of moving.

There are at least three contributing factors for why there are more sub-moves on touchscreen, which are presented next.

Occlusion. When participants move a shape toward a target position that was occluded by the hand, they drop the shape partway and check the target position (P5).

Friction between finger and screen. P7 and P9 reported intentionally using the *sub-move* strategy on the touchscreen by dropping the shape in the middle of a movement trace because moving a finger over a long distance on screen causes "too much friction." Neither participant used the same strategy with a mouse.

Fail-to-grab move may have caused participants to do more sub-moves on the touchscreen. Participants had significantly more *fail-to-grab* moves on the touchscreen compared to the mouse ($F_{1,15} = 19.22, p < 0.001, \eta_G^2 = .56$). In addition, *fail-to-grab* is correlated with *sub-move* ($R_2 = .71, p < .001$). *Fail-to-grab* moves happened when participants were unable to acquire a shape when they initiated dragging. The imprecise nature of pointing with a finger on a touchscreen (vs. precise pointing with a mouse) explains why there were more *fail-to-grab* moves on the touchscreen. From our observation, when participants were unable to acquire a shape in their first few attempts (*fail-to-grab*), they would continuously "brush" their finger over the shape trying to acquire it. Among one of those "brushing" moves, they could acquire the shape and drag it partway.

Although there were consistent *fail-to-grab* moves observed across all participants, some participants had very different attitudes toward this type of move. Participants not used to the touchscreen became especially annoyed and anxious if they could not grab a shape. Yet participants with touchscreen experience reported they did not mind it at all.

Most participants were not aware that they made more moves on a touchscreen. All participants, with the exception of P7 and P9 (who reported intentionally using the *sub-moves* strategy on the touchscreen), reported that they "did not feel compelled to make more moves on touchscreen," and they had made "less or about the same number of moves on touchscreen."

The move classification for dragging revealed some interesting issues about device affordances. Some participants performed a rotation gesture (*rotation attempt*), but only for those who received the touchscreen conditions before the mouse conditions, suggesting that touchscreen is a natural device to afford complex gestures, such as rotation. Participants also performed significantly more *constrained moves* with a mouse, trying to move shapes out of the canvas boundary.

To summarize, we found participants had considerably different movement patterns across the two devices in *Pattern Construction*. Of particular note is that on a touchscreen, participants tended to make multiple *sub-moves*, resulting in a higher number of moves but shorter distances in each move. The reverse was observed when participants used a mouse.

5.4 Subjective Preference

A summary of participants' preference of device by interaction type (pointing and dragging) and task type (abstract and real-world) is presented in Table 8.

For the analysis, we excluded counts for a tie (no preference of device). We first looked into whether participants had different device preferences for pointing and dragging tasks by

Table 8. Participants' Subjective Preference of Device by Task (N = 96)

	Tasks	Mouse	Touch	Tie
Pointing	Abstract	0	14	2
	Real-world low precision	4	10	2
	Real-world high precision	1	13	2
Dragging	Abstract	3	10	3
	Real-world low precision	2	5	9
	Real-world high precision	7	6	3
	Totals	17	58	21

collapsing votes across task type. The chi-square test revealed that preference did differ by interaction type, $\chi^2_{(1, N=75)} = 4.50, p = 0.03$. Although participants expressed a preference for the touchscreen over a mouse for both interaction types, the magnitude of their preference was much stronger for pointing tasks (37 vs. 5, with 6 ties) compared to dragging tasks (21 vs. 12, with 15 ties).

We were also interested to know if participants had different device preferences for abstract versus real-world tasks. We similarly collapsed the votes across interaction type. A chi-square test revealed no difference in preference of device for either type of task, $\chi^2_{(1, N=75)} = 2.01, p = 0.15$.[2] The collapsed votes for both tasks indicate that participants expressed a strong preference for the touchscreen over a mouse regardless of task type (real world: 34 vs. 14, with 16 ties; abstract: 24 vs. 3, with 5 ties).

In self-reports, participants preferred the touchscreen because it was "fast," "direct," "intuitive to use," and "easier to point." In tasks where participants preferred a mouse over a touchscreen, the main reasons were the "high precision" of the mouse cursor, "no occlusion of finger on screen," and familiarity with a mouse. A tie in the preference for device occurred when participants reported that the task took a lot of mental effort. For example, participants reported they did not mind whether *Sentence Comprehension* is on a touchscreen or a mouse because "the hard part of the test is memorizing the instruction."

6 DISCUSSION

We start the discussion with performance results—speed and accuracy—for the abstract and the real-world C-TOC tasks. We then look more closely at the different movement strategies participants adopted in the *Pattern Construction* test. Drawing on the analysis, we suggest some implications for touchscreen design and then discuss limitations of the study.

6.1 Reliability of Measures for Computerized Cognitive Tasks

The performance results we obtained for the C-TOC tasks are mostly consistent with those for the corresponding Fitts's Law abstract tasks in terms of differences between devices. A touchscreen speed advantage was found for both pointing and dragging on the C-TOC tests, but not all time differences were statistically significant or even going in the same direction. Therefore, we should

[2] The analysis for pointing versus dragging is significant, but abstract versus real world is not. In fact, the ratios are similar between the two for the first analysis (37 vs. 5 approximately $7:1$; 21 vs. 12 approximately $2:1$), and for the second analysis (34 vs. 14 approximately $2.5:1$; 24 vs. 3 approximately $8:1$). Yet the first is significant and the second is not. We followed up by running Fisher's exact test, which is another non-parametric distribution test (similar to chi-square). The outcome for the abstract versus real-world comparison was $p = .09$, thus borderline significant. Altogether, this points to the need for further research on subjective preferences for devices for different types of tasks.

be cautious when interpreting time to completion as a measure of cognitive ability since performance speed on pointing and dragging tasks is dependent also on input device. This is especially the case for tasks that require movement actions, such as those in Fitts's Law tasks, as well as cognitive tasks that require a rather simple movement as in *Sentence Comprehension*, in which individuals performed the task significantly faster on a touchscreen compared to a mouse. Compared to these tasks, the relatively more complicated cognitive task *Pattern Construction* did not have a significant difference in the total task completion time. It may be that the cognitive demands of the task mask or wash out the input device difference in terms of the total time spent to finish the task. In terms of accuracy differences between devices, the increased error rate on the touchscreen for high-precision abstract pointing (Figure 5, left) was reflected in the high-precision pointing C-TOC task (*Arithmetic*), although the magnitude of the error rate differences between devices was smaller with C-TOC compared to abstract pointing. There was no device difference for the other three C-TOC tasks.

To manage the impact of the input device differences when taking a self-administrated test, it would be useful to be able to explain the performance difference for performance calibration. We attempted to see whether Fitts's Law task measures could explain the difference, and we found that device-related performance differences on Fitts's Law tasks were only weakly to moderately predictive of device-related performance differences on C-TOC tasks. We observed weak correlations for all tasks except for the high-precision pointing task (*Arithmetic*). It is likely that the strategy participants used when solving cognitive tasks (some striving for accuracy, others aiming for speedy performance) played a role in reducing the relationship between C-TOC and Fitts's Law tasks. We would need to conduct a larger-scale study, with tight experimental control over participants' task strategy, to conclude whether Fitts's Law tasks measures have meaningful predictive power.

6.2 Different Movement Patterns on the Touchscreen and Mouse

There were different behavioral patterns between the touchscreen and mouse for *Pattern Construction.* On the touchscreen, participants performed twice as many dragging movements as they did with a mouse. Dragging distance on a touchscreen was relatively shorter, and participants were more likely to decompose a single target-oriented move into smaller consecutive moves (*sub-moves*).

One reason that device type contributes to a difference in movement patterns is a bigger overhead in using a mouse during dragging. This comes from two sources: (1) extra workload in pressing down with a mouse compared to just a finger for a single click, and (2) longer time for re-acquiring a shape if it is moved only partway and then moved again. Reacquiring a shape is similar to a pointing task—it is much slower using a mouse than using a touchscreen, according to the pointing task results. The larger overhead of a mouse discourages users from dropping a shape during dragging and then reacquiring it. One extreme case is the *trial & error* move, which is a single dragging movement that moves a shape to more than one target position. Data showed that *trial & error* moves happen often with a mouse: participants, once having acquired a shape, did not release the mouse button until the shape reached a final target position, often after two or more attempts at placing it. We saw much less of this behavior on the touchscreen.

Another reason for the substantial number of moves on the touchscreen was the *sub-move* strategy, which makes dragging tasks much easier. By separating a single target move into n steps of *sub-moves*, task precision for each move is significantly reduced: the distance for each *sub-move* is arbitrarily decided, with an average of A/n (A being the total distance for the move). Dragging on a touchscreen has relatively little overhead compared to a mouse, and thus there is little cost if a single dragging move is decomposed into multiple movements, each with shorter distance.

Despite the dramatic difference in the total number of moves between devices, no participant seemed aware of the difference. Most participants reported that they "don't think [they] have made more or less moves in either device." Some participants even felt they had more moves using a mouse. Most, but not all, reported having no recollection of making *sub-moves*, even after we demonstrated it. The discrepancy between what participants did and what they believed they did might be due to cognitive chunking when registering *sub-moves*. Users might implicitly chunk all *sub-moves* into a single target-oriented move. More research is needed to better understand the mechanism of *sub-moves* and the cognitive chunking behind it.

6.3 Implications for Touchscreen Interface Design and Deployment

In our study, we validated findings on pointing and dragging tasks with different input devices and extended them to real-world tasks. In this section, we suggest three implications for design and two for deployment regarding task designs for touchscreen-based interaction.

Design 1: Utilize more pointing and less dragging, to better support aging users. Compared to dragging tasks, pointing tasks on a touchscreen are known to be faster but can be more error prone for older adults [33]. This is consistent with our results for both abstract and real-world tasks. Interfaces that are designed for users with lower dexterity or older age might best take advantage of touchscreens by adopting more pointing gestures and fewer dragging gestures, so long as target size is sufficiently large (see next design point). This will provide the intended users with an easier and faster interaction experience.

Design 2: Have a simpler interface with bigger buttons that have a reasonable spacing. For high-precision pointing tasks there is a difference between devices, so it would be better to reduce the precision that is needed for accomplishing the task (e.g., provide larger buttons). For abstract tasks, our results show that the touchscreen was particularly inaccurate for high-precision tasks in which the target width was around 4 mm (33–40 px), which is in line with Kobayashi et al. [25], who found that 30 px was too small for older adults on pointing tasks. Making buttons bigger than 40 px would help to get rid of the accuracy discrepancy between the touchscreen and mouse. Furthermore, due to smaller touchscreen screen sizes compared to the screens used with a mouse in most commercial products, older adults are more likely to be overwhelmed by cluttered interfaces on a touchscreen.

Design 3: Provide support for decomposable dragging tasks. The *sub-move* strategy adopted by some participants indicates a natural tendency to drag differently with a touchscreen than with a mouse. Especially for real-world tasks with dragging, there was no significant difference in total task completion time between the mouse and the touchscreen. However, participants made moves with a shorter distance more quickly on a touchscreen. Similar to dragging, Toy et al. [46] showed that shorter-lasting steering tasks had lower levels of error compared to longer tasks (6 seconds vs. 12 seconds). Therefore, touchscreen interfaces should support decomposable dragging by allowing objects to "hang" when they are dropped instead of going all the way back to the starting point if users prefer to use the *sub-move* strategy. For example, when a file is dragged into a folder, users could "pause" in the middle of the move and then initiate another drag to complete the move, without the file snapping back to its original position.

Deployment 1: Give explicit instruction about capacitive touchscreen usage so users can harness the full power of touchscreen sensitivity. Our experiment used a capacitive touchscreen—a fourth-generation iPad. During the pilot, we noticed that older adults still treated the iPad as if it were a resistive touchscreen: they pressed hard to acquire objects. More importantly, most of them, even some who have iPads at home, were not aware that capacitive touchscreens, like the iPad, depend on the conductive nature of human body. They sometimes attempted to use their fingernail to point to smaller objects and later blamed the insensitivity of the screen. Wood et al. [50] similarly detected performance issues in dragging interactions for older adults, especially on resistive touchscreens. This reinforces the need for explicit instructions about capacitive touchscreens. During our actual experiment, all participants were instructed to use their fingertip, not fingernail, to point or drag on a touchscreen. Participants reported that they found “the instruction is especially useful” and the touchscreen was “easier to use” after hearing the instruction.

Deployment 2: To deploy a cognitive test on both a PC and tablet, normative data will likely be needed. Our research demonstrates that older adults’ computerized cognitive test performance is significantly impacted by the input device they use. We speculate that each particular type of input device creates specific task demands that are different for a touchscreen and mouse. The touchscreen appears to invite faster but less accurate responses to task stimuli, and this should be considered in task design. For example, a task like our *Arithmetic* test may require larger response boxes on a touchscreen and/or instructions to respond both accurately and quickly. Additional touchscreen-specific normative performance data will likely also be needed.

6.4 Limitations of the Study

We discuss four limitations noted by the referees.

6.4.1 Study Environment. The study was conducted in a regular university office with no special soundproofing or other acoustical treatment. Our ultimate goal is a self-administered online test for older adults to take at their home. We expect that homes have a range of noise levels from a variety of sources. Our study provides encouraging evidence that testing in a normal acoustic environment works well enough to serve as a screening step. Future studies will need to determine if home environments with high noise levels pose a problem.

6.4.2 Fatigue. Participants each spent about 1.5 hours in the study. Fatigue, as measured by the PegBoard task, was not an issue. Nevertheless, fatigue could be a concern in self-administered home-based testing. Participants spent time completing the pre-trials demographics questionnaire, the Purdue Pegboard Test, and the Snellen Vision Test, as well as the post-trials Pegboard Test and interview. These would not be required in actual deployment, so duration ought to be less than an hour. Further research might find ways to decrease the time required or to break testing into shorter modules to reduce the likelihood of fatigue.

6.4.3 Disability. The older adults in our study may not be representative of the target population in terms of disabilities or levels of manual dexterity. None of our participants had noticeable motor deficiencies. Further research will be required to determine whether this poses a limitation. Two participants in our study did report having arthritis. We did not compare their performance against other participants’ performance beyond an initial review of the data for anomalies (none were found) and removal of outlier trials for all participants.

6.4.4 Outliers in the Data. During data analysis, we eliminated outlier trials in which movement amplitude was (1) less than half the trial amplitude or (2) greater than three standard deviations from the mean distance for the set of trials with the same participant, device, task, and target width. Outlier removal is common for Fitts’s Law studies. Casiez [6] instructs students to “remove the outliers by removing trials two standard deviations away from the mean” before testing movement time as a function of amplitude and target width and in another study [1] employs a threshold of three standard deviations to remove outliers.

7 CONCLUSION AND FUTURE DIRECTIONS

Our experiment revealed that input devices have impact on speed and accuracy performance measures for computerized cognitive tests. While designing the four subtests to reflect precision level of a similar Fitts's Law task, three subtests were faster on the touchscreen, and the mouse had higher accuracy for one subtest. This study has shown that when computerized cognitive tests are done using different input devices, there are device-related performance differences that need to be taken into consideration when evaluating the performance. We attempted to use Fitts's Law performance to predict the device effect on cognitive task performance with comparable precision levels and did not find strong predictive power. However, we recognize that our sample size was small. Our goal was to obtain proof of concept evidence of device-related performance differences rather than precise effect sizes. Larger-scale studies replicating and extending these findings would be desirable in the future.

We also found that older adults naturally adopted different movement patterns between devices in one of our C-TOC tasks: on a touchscreen, they decomposed a single dragging movement into multiple movements, resulting in shorter individual moves but a greater number of moves compared to a mouse. Future work should examine specifically the relationship between movement categories and cognitive functioning. If movement patterns have clinical utility, coding could be automated and eventually be integrated into C-TOC.

Our work provides important insights for the C-TOC Project. The device-specific difference in performance by older adults on C-TOC tasks suggests that a larger-scale study will be needed to find valid performance calibrations across devices.

ACKNOWLEDGMENTS

The authors thank the anonymous referees for comments and suggestions that improved the article.

REFERENCES

[1] Marc Baloup, Thomas Pietrzak, and Géry Casiez. 2019. RayCursor: A 3D pointing facilitation technique based on raycasting. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI'19)*. ACM, New York, NY, 1–12. DOI: https://doi.org/10.1145/3290605.3300331

[2] Matthew Brehmer, Joanna McGrenere, Charlotte Tang, and Claudia Jacova. 2012. Investigating interruptions in the context of computerised cognitive testing for older adults. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 2649–2658.

[3] Henry Brodaty, Dimity Pond, Nicola M. Kemp, Georgina Luscombe, Louise Harding, Karen Berman, and Felicia Huppert. 2002. The GPCOG: A new screening test for dementia designed for general practice. *Journal of the American Geriatrics Society* 50, 3 (2002), 530–534.

[4] Niamh Caprani, Noel E. O'Connor, and Cathal Gurrin. 2012. Touch screens for the older user. In *Assistive Technology*. InTech, Rijeka, Croatia, 95–118.

[5] Stuart K. Card, Thomas P. Moran, and Allen Newell. 1983. *The Psychology of Human-Computer Interaction*. L. Erlbaum Associates, Hillsdale, NJ.

[6] Géry Casiez. 2018. Fitts' Experiment Analysis. Retrieved September 16, 2020 from https://cs.uwaterloo.ca/~gcasiez/cs889f18/exercises/e5_fittslaw.html

[7] Neil Charness, Patricia Holley, Jeffrey Feddon, and Tiffany Jastrzembski. 2004. Light pen versus mouse for a menu selection task: Age, hand, and practice effects. *Human Factors* 46 (2004), 373–384.

[8] Andy Cockburn, David Ahlström, and Carl Gutwin. 2012. Understanding performance in touch selections: Tap, drag and radial pointing drag with finger, stylus and mouse. *International Journal of Human-Computer Studies* 70, 3 (2012), 218–233.

[9] David Darby, P. Maruff, A. Collie, and M. McStephen. 2002. Mild cognitive impairment can be detected by multiple assessments in a single day. *Neurology* 59, 7 (2002), 1042–1046.

[10] Johanne Desrosiers, Réjean Hébert, Gina Bravo, and Elisabeth Dutil. 1995. The Purdue Pegboard Test: Normative data for people aged 60 and over. *Disability and Rehabilitation* 17, 5 (1995), 217–224.

[11] Glen M. Doniger, David M. Zucker, Avraham Schweiger, Tzvi Dwolatzky, Howard Chertkow, Howard Crystal, and Ely S. Simon. 2005. Towards practical cognitive assessment for detection of early dementia: A 30-minute computerized battery discriminates as well as longer testing. *Current Alzheimer Research* 2, 2 (2005), 117–124.

[12] Leah Findlater, Jon E. Froehlich, Kays Fattal, Jacob O. Wobbrock, and Tanya Dastyar. 2013. Age-related differences in performance with touchscreens compared to traditional mouse input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'13)*. ACM, New York, NY, 343–346. DOI: https://doi.org/10.1145/2470654.2470703

[13] Marshal F. Folstein, Susan E. Folstein, and Paul R. McHugh. 1975. "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* 12, 3 (1975), 189–198.

[14] Clifton Forlines, Daniel Wigdor, Chia Shen, and Ravin Balakrishnan. 2007. Direct-touch vs. mouse input for tabletop displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'07)*. ACM, New York, NY, 647–656. DOI: https://doi.org/10.1145/1240624.1240726

[15] Shathel Haddad, Joanna McGrenere, and Claudia Jacova. 2014. Interface design for older adults with varying cultural attitudes toward uncertainty. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 1913–1922.

[16] Juan Pablo Hourcade and Theresa R. Berkel. 2008. Simple pen interaction performance of young and older adults using handheld computers. *Interacting with Computers* 20, 1 (2008), 166–183.

[17] Hwan Hwangbo, Sol Hee Yoon, Beom Suk Jin, Young Suk Han, and Yong Gu Ji. 2013. A study of pointing performance of elderly users on smartphones. *International Journal of Human-Computer Interaction* 29, 9 (2013), 604–618.

[18] ISO:9241-400. 2007. *Ergonomics of Human-System Interaction—Part 400: Principles and Requirements for Physical Input Devices.* ISO 9241-400:2007(E). International Organization for Standardization, Geneva, Switzerland.

[19] Hirokazu Iwase and Atsuo Murata. 2003. Empirical study on the improvement of the usability of a touch panel for the elderly—Comparison of usability between a touch panel and a mouse. *IEICE Transactions on Information and Systems* 86, 6 (2003), 1134–1138.

[20] Claudia Jacova, Andrew Kertesz, Mervin Blair, John D. Fisk, and Howard H. Feldman. 2007. Neuropsychological testing and assessment for dementia. *Alzheimer's & Dementia* 3, 4 (2007), 299–317.

[21] Claudia Jacova, Joanna McGrenere, Hyunsoo Lee, William Wang, Sarah Le Huray, Matthew Brehmer, Samantha Feldman, et al. 2012. Cognitive Testing on Computer (C-TOC): Design, usability evaluation and validation of a novel computerized testing tool. *Alzheimer's & Dementia* 8, 4 (2012), P540.

[22] Claudia Jacova, Joanna McGrenere, Hyunsoo S. Lee, William W. Wang, Sarah Le Huray, Emily F. Corenblith, Matthew Brehmer, et al. 2015. C-TOC (Cognitive Testing on Computer): Investigating the usability and validity of a novel self-administered cognitive assessment tool in aging and early dementia. *Alzheimer Disease & Associated Disorders* 29, 3 (2015), 213–221.

[23] Zhao Xia Jin, Tom Plocher, and Liana Kiff. 2007. Touch screen user interfaces for older adults: Button size and spacing. In *Universal Access in Human Computer Interaction. Coping with Diversity*. Springer, Beijing, China, 933–941.

[24] Jeffrey A. Kaye, Shoshana A. Maxwell, Nora Mattek, Tamara L. Hayes, Hiroko Dodge, Misha Pavel, Holly B. Jimison, Katherine Wild, Linda Boise, and Tracy A. Zitzelberger. 2011. Intelligent systems for assessing aging changes: Home-based, unobtrusive, and continuous assessment of aging. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 66, Suppl. 1 (2011), i180–i190.

[25] Masatomo Kobayashi, Atsushi Hiyama, Takahiro Miura, Chieko Asakawa, Michitaka Hirose, and Tohru Ifukube. 2011. Elderly user evaluation of mobile touchscreen interactions. In *Human-Computer Interaction—INTERACT 2011*. Springer, Lisbon, Portugal, 83–99.

[26] Roxanne Leitão and Paula Alexandra Silva. 2012. Target and spacing sizes for smartphone user interfaces for older adults: Design patterns based on an evaluation with users. In *Proceedings of the 19th Conference on Pattern Languages of Programs (PLoP'12)*. Article 5, 13 pages.

[27] Muriel D. Lezak, Diane B. Howieson, Erin D. Bigler, and Daniel Tranel. 2012. *Neuropsychological Assessment*. Oxford University Press, Oxford, UK.

[28] Monica Luciana. 2003. Practitioner review: Computerized assessment of neuropsychological function in children: Clinical and research applications of the cambridge neuropsychological testing automated battery (CANTAB). *Journal of Child Psychology and Psychiatry* 44, 5 (2003), 649–663.

[29] I. Scott MacKenzie. 1992. Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction* 7, 1 (March 1992), 91–139. DOI: https://doi.org/10.1207/s15327051hci0701_3

[30] Karyn Moffatt and Joanna McGrenere. 2009. Exploring methods to improve pen-based menu selection for younger and older adults. *ACM Transactions on Accessible Computing* 2, 1 (2009), 3.

[31] Atsuo Murata. 2006. Eye-gaze input versus mouse: Cursor control as a function of age. *International Journal of Human-Computer Interaction* 21, 1 (2006), 1–14.

[32] Ziad S. Nasreddine, Natalie A. Phillips, Valérie Bédirian, Simon Charbonneau, Victor Whitehead, Isabelle Collin, Jeffrey L. Cummings, and Howard Chertkow. 2005. The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society* 53, 4 (2005), 695–699.

[33] Ho-Chuen Ng, Da Tao, and Calvin K. L. Or. 2013. Age differences in computer input device use: A comparison of touchscreen, trackball, and mouse. In *Advances in Information Systems and Technologies*. Springer, Algarve, Portugal, 1015–1024.

[34] Emmanuel Pietriga, Caroline Appert, and Michel Beaudouin-Lafon. 2007. Pointing and beyond: An operationalization and preliminary evaluation of multi-scale searching. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'07)*. ACM, NY, 1215–1224. DOI: https://doi.org/10.1145/1240624.1240808

[35] Chethan Ramprasad, Leonardo Tamariz, Jenny Garcia-Barcena, Zsuzsanna Nemeth, and Ana Palacio. 2017. The use of tablet technology by older adults in health care settings—Is it effective and satisfying? A systematic review and meta analysis. *Clinical Gerontologist* 30 (2017), 1–10.

[36] Christopher Randolph, Mary C. Tierney, E. Mohr, and Thomas Newell Chase. 1998. The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): Preliminary clinical validity.*Journal of Clinical and Experimental Neuropsychology* 20 3 (1998), 310–319.

[37] Ivar Reinvangm. 2012. Mild Cognitive Impairment—The Gøteborg-Oslo (GO) Project (Completed). Retrieved September 16, 2020 from http://www.sv.uio.no/psi/english/research/projects/go-project/index.html.

[38] Dorene M. Rentz, Maria Dekhtyar, Julia Sherman, Samantha Burnham, Deborah Blacker, Sarah L. Aghjayan, Kathryn V. Papp, et al. 2016. The feasibility of at-home iPad cognitive testing for use in clinical trials. *Journal of Prevention of Alzheimer's Disease* 3, 1 (2016), 8.

[39] Trevor W. Robbins, Merle James, Adrian M. Owen, Barbara J. Sahakian, Lynn McInnes, and Patrick Rabbitt. 1994. Cambridge Neuropsychological Test Automated Battery (CANTAB): A factor analytic study of a large sample of normal elderly volunteers. *Dementia and Geriatric Cognitive Disorders* 5, 5 (1994), 266–281.

[40] Wendy A. Rogers, Arthur D. Fisk, Anne Collins McLaughlin, and Richard Pak. 2005. Touch a screen or turn a knob: Choosing the best device for the job. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 47, 2 (2005), 271–288.

[41] Farzan Sasangohar, I. Scott MacKenzie, and Stacey D. Scott. 2009. Evaluation of mouse and touch input for a tabletop display using Fitts' reciprocal tapping task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 53, 12 (2009), 839–843.

[42] Nicole Schneider, Janet Wilkes, Morten Grandt, and Christopher M. Schlick. 2008. Investigation of input devices for the age-differentiated design of human-computer interaction. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 52, 2 (2008), 144–148.

[43] R. William Soukoreff and I. Scott MacKenzie. 2004. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies* 61, 6 (2004), 751–789.

[44] Jennifer C. Thompson, Cheryl L. Stopford, Julie S. Snowden, and David Neary. 2005. Qualitative neuropsychological performance characteristics in frontotemporal dementia and Alzheimer's disease. *Journal of Neurology, Neurosurgery & Psychiatry* 76, 7 (2005), 920–927.

[45] Jane B. Tornatore, Emory Hill, Jo Anne Laboff, and Mary E. McGann. 2005. Self-administered screening for mild cognitive impairment: Initial validation of a computerized test battery. *Journal of Neuropsychiatry and Clinical Neurosciences* 17, 1 (2005), 98–105.

[46] Katelyn Toy, Edward O'Meara, Ravi Kuber, and Sidas Saulynas. 2017. An investigation of ways to support older adults when using mobile interfaces. In *Proceedings of iConference 2017*. 11.

[47] Lisa D. Van Mierlo, Hans Wouters, Sietske A. M. Sikkes, Wiesje M. Van der Flier, Niels D. Prins, Jonne A. E. Bremer, Teddy Koene, and Hein P. J. Van Hout. 2017. Screening for mild cognitive impairment and dementia with automated, anonymous online and telephone cognitive self-tests. *Journal of Alzheimer's Disease* 56, 1 (2017), 249–259.

[48] Keith A. Wesnes, Helen Brooker, Clive Ballard, Laura McCambridge, Robert Stenton, and Anne Corbett. 2017. Utility, reliability, sensitivity and validity of an online test system designed to monitor changes in cognitive function in clinical trials. *International Journal of Geriatric Psychiatry* 32, 12 (2017), e83–e92.

[49] Katherine Wild, Diane Howieson, Frank Webbe, Adriana Seelye, and Jeffrey Kaye. 2008. Status of computerized cognitive testing in aging: A systematic review. *Alzheimer's & Dementia* 4, 6 (2008), 428–437.

[50] Eileen Wood, Teena Willoughby, Alice Rushing, Lisa Bechtel, and Jessica Gilbert. 2005. Use of computer input devices by older adults. *Journal of Applied Gerontology* 24, 5 (2005), 419–438.

[51] Kailun Zhang. 2015. *A Comparison of Touchscreen and Mouse for Real-World and Abstract Tasks with Older Adults*. Master's Thesis. University of British Columbia. https://open.library.ubc.ca/cIRcle/collections/ubctheses/24/items/1.0216481.

[52] Jia Zhou, Pei-Luen Patrick Rau, and Gavriel Salvendy. 2014. Age-related difference in the use of mobile phones. *Universal Access in the Information Society* 13, 4 (2014), 401–413.
[53] Stelios Zygouris and Magda Tsolaki. 2015. Computerized cognitive testing for older adults: A review. *American Journal of Alzheimer's Disease and Other Dementias* 30, 1 (2015), 13–28.

Received November 2019; revised July 2020; accepted July 2020