Slippery When Incorrect: Breaking Down Cognitive Activity during Errors with Fine-Finger Tracking on Touchscreens

SUJIT KUMAR KAMARAJ, University of Utah, USA

VINEET PANDEY, University of Utah, USA

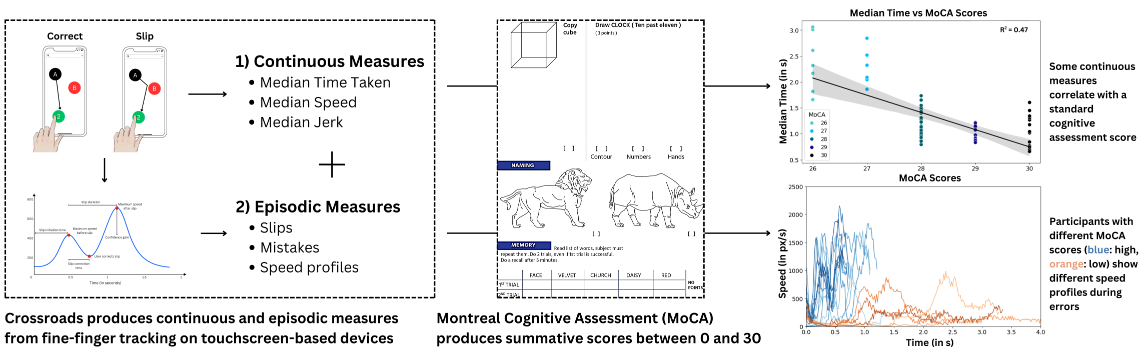


Fig. 1. Our work proposes to understand cognitive activity. To do this, we provide a new tool *Crossroads* that develops episodic measures for infrequent errors made by users and continuous measures from small finger trajectories. A study with 11 participants found that many measures are reliable and correlate well with standard cognitive assessment scores. Additionally, speed profiles during errors provide information that is complementary to standard assessments.

Recent research endeavors have proposed digital health tools that can assess physiological functioning, like lung function and motor performance. Similar research efforts to assess cognitive performance are largely missing. Traditional cognitive assessments require in-person interaction with experts, making them infrequent and not widely available. Furthermore, such cognitive assessments produce summative scores that do not yield substantial insights into underlying cognitive performance. Our work provides a human-centered approach to understanding cognitive activity by focusing on *errors* made by people while performing touchscreen-based tasks. We contribute a novel tool, *Crossroads*, for fine-finger tracking that generates measures that are *episodic* (*e.g.*, speed profile during errors) or *continuous* (*e.g.*, median speed over the entire finger trajectory). A user study demonstrates *Crossroads'* feasibility in tracking cognitive performance. Crossroads produces speed-time profiles when users commit errors: people with different cognitive scores demonstrate varying profiles. Multiple continuous measures demonstrate high test-retest reliability and correlate well with summative scores from a standard assessment. Participants report high acceptability for the tool. The median usage time for Crossroads is less than two minutes. Our work suggests that fine-finger tracking provides a novel opportunity to understand cognitive performance in ways that go beyond summative scores provided by current cognitive assessments.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing systems and tools**.

Additional Key Words and Phrases: Errors, Fine-finger tracking, Cognitive performance, Empirical study

Authors' Contact Information: Sujit Kumar Kamaraj, University of Utah, USA, sujit.kamaraj@utah.edu; Vineet Pandey, University of Utah, USA, vineet.pandey@utah.edu.

Manuscript submitted to ACM

1 Introduction

Health and well-being involve multiple components, including physiological functioning and cognitive performance [27, 43]. Increasingly, multiple digital tools assess physiological functioning. For example, SpiroCall uses exhalation sounds made over a phone call to assess lung function [25], BiliCam uses a smartphone's camera to assess newborn jaundice [16], and Hevelius uses a mouse-based pointing task to assess motor performance [38]. However tools for cognitive performance are under explored. We find this to be an important are of investigation since fluctuations in cognitive performance can signal stress, fatigue, medication side effects, or the early stages of neurological disorders—factors that often go unnoticed without systematic tracking [5, 35, 36, 44].

Cognitive performance includes multiple aspects such as memory, perception, language, reasoning, and decision-making [26]. Traditional assessments exist for these aspects of cognitive performance, but they provide summative scores and require in-person interaction with trained experts who have limited availability. Additionally, almost all these tests are pen-and-paper based. When digital variants exist, they copy existing pen-and-paper assessments [13]. How might we design digital tools that assess some cognitive performance in a way that complements existing cognitive assessments, while making use of the affordances provided by ubiquitous digital tools? We believe developing a computational understanding of errors provides a clue.

Errors provide important insights into cognitive performance. They can be used to track changes in cognitive performance and track individual differences [48]. Current cognitive performance assessments make no effort to track and analyze errors. We consider this a missed opportunity. Our work provides *Crossroads*, a novel tool which tracks cognitive performance using *fine-finger tracking*–continuously tracking the fluctuations in the finger location as the participant performs the task [19]. Finger trajectory on a touchscreen is decomposed into continuous measures and errors made while using the tool. Unlike prior contributions, our work develops speed-time profiles from errors to provide fine-grained insights. Crossroads also tracks measures like time taken, speed, acceleration, pause duration, and jerk.

An in-person lab study evaluated Crossroads with 11 participants. The findings demonstrates multiple promising results. Speed-time profiles from the errors provide rich insights into each participant's cognitive performance. Multiple continuous measures of Crossroads show high reliability over a single session. Many of these reliable measures also demonstrate a strong correlation with a standard cognitive assessment. Participants took a median time of less than two minutes to perform our task compared to a median time of nine minutes for a standard cognitive assessment.

Our work makes two contributions to human-centered computing in healthcare literature [53]:

(1) an artifact contribution with a novel self-administered tool that uses fine-granular information about errors to assess cognitive performance within minutes.
(2) an empirical contribution with a mixed-methods lab study that assesses the feasibility of this approach with measures of reliability, relevance, and acceptability.

Our tool and the empirical findings showcase promising directions in digital health research where web-based tasks on touchscreen-based devices can yield useful measures of cognitive performance.

2 Related Work

Our research draws inspiration from the strengths and limitations of cognitive assessments. Our ideas build on prior work around tracking errors that have been useful in multiple domains to understand cognitive activity.

2.1 Cognitive performance assessments provide summative scores and require in-person interaction with experts

Current cognitive assessments provide useful outcomes but struggle with multiple challenges. Traditional pen-and-paper screening tests, like the Montreal Cognitive Assessment (MoCA), provide a summative score on a scale of 0 to 30 [37]. These tests usually involve tasks such as memorizing a short list of words, naming objects shown in pictures, copying shapes, and arithmetic problems. Screening tests, like MoCA, *require* experts to both administer and evaluate the test. Furthermore, standard thresholds used to screen participants for cognitive impairments are not agreed upon by experts [12]. Therefore, summative scores provided by traditional cognitive assessments require interpretation across different contexts, often leading to ad-hoc decision-making by experts. For example, although the standard MoCA cutoff for mild impairment is 26/30, experts recommend lowering it to 23/30 for older adults or those with less education, highlighting how highly summative scores fail to account for individual and contextual differences [12]. There have been cases of *false negatives* where an individual actually has a cognitive impairment, but the assessment fails to detect it due to rigid threshold criteria [40]. Furthermore, due to their reliance on expert presence, traditional cognitive assessments are difficult to access for many people and prevent frequent testing. Experts need to supervise and evaluate traditional screening and comprehensive cognitive assessments. Therefore, participants need to travel to the clinic or labs to access such assessments.

Beyond issues of access, traditional cognitive assessments by their design miss out on important details. For example, summative scores do not differentiate between participants who struggle, hesitate, or make multiple attempts to complete a task [8, 28]. Additionally, current assessments do not capture the degree of correctness or the nature of errors made. For instance, tests like MoCA do not differentiate between slight mistakes, which might indicate a small change in in cognitive performance, and a completely incorrect response, which could suggest a severe decrease in cognitive performance. For example, in the Montreal Cognitive Assessment (MoCA), participants need to draw a clock with specific features in the clock-drawing task. The evaluation is based solely on the completed clock. If two participants make errors–such as misplacing a number or positioning the clock hands incorrectly–and correct these errors during the task, both receive full scores on this task and the differences in errors and subsequent detection is not tracked. We believe that the research and clinical community's understanding of cognitive performance will benefit from tracking such rich process-level information (i.e. *how* participants perform a task) in conjunction with outcome-level information (i.e. *what* was a participant's output for a task).

Overall, we argue that developing cognitive assessment tools that reduce reliance on experts, take less time, and provide more specific measures of cognitive performance–like errors–can benefit multiple stakeholders in the healthcare process.

2.2 Self-administered digital tools designed for cognitive assessments do not track cognitive processes like error detection and correction

Many digital cognitive assessments are digital copies of traditional pen-and-paper assessments that provide summative scores [13]. For example, digital copies of the Montreal Cognitive Assessment provide summative scores out of 30 and rely on a trained expert to supervise and interpret the results [6, 13, 30, 49, 54].

Some digital variants of traditional cognitive assessments, like the digital Trail Making Test and digital Whack-a-Mole, provide multiple measures rather than a single summative score (Figure 2a) [21, 50]. For example, in the traditional (pen-and-paper) Trail Making Test, the participant is asked to connect the dots labeled with either a letter or number

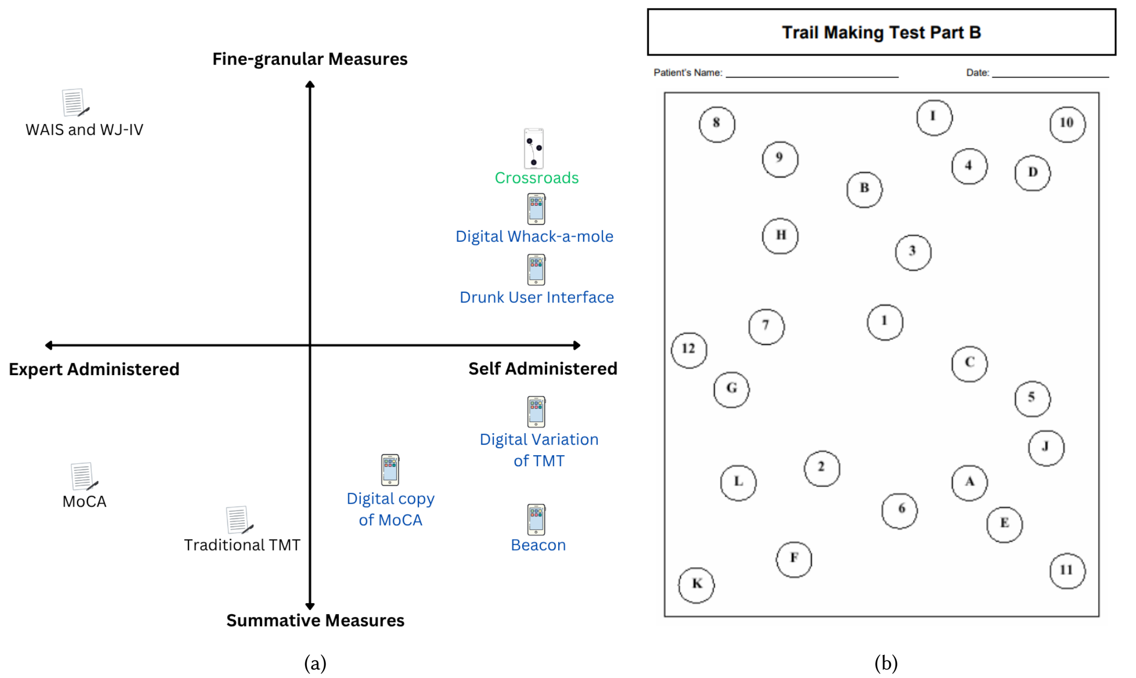


Fig. 2. a) A design space summarizes different cognitive assessments along two dimensions: 1) who administers the assessment, and 2) nature of measures tracked. Traditional assessments, like the Trail Making Test [9], produce a single summative score (time taken). Cognitive assessments are typically administered by experts. Some, like Wechsler Adult Intelligence Scale [51], require substantial expert time (60 to 90 minutes) and produce detailed measures. Our work contributes Crossroads which assesses cognitive activity and produces descriptive measures without the need for experts to administer the task.
b) The Trail Making Test (TMT) is a screening test used to screen for cognitive impairments based on the time taken to complete the test. Participants connect the dots while alternating between letters and numbers in ascending order. Some digital version of the Trail Making Test measures other features like the number of pauses, pause duration, lift duration, and number of errors.

(Figure 2b) [9]. The participant needs to connect the dots in ascending order while alternating between numbers and letters (1 to A, A to 2, 2 to B, and so on). The time taken to complete the pen-and-paper Trail Making Test is used to assess cognitive functioning. The digital version of the Trail Making Test tracks additional measures like the number of pauses, pause duration, lifts, lift duration, number of errors, time inside each circle, and time between circles [21]. Such additional measures provided by digital variants are useful but only provide details about the summative performance. They still overlook infrequent but meaningful occurrences—like self-corrections, pauses, or exploratory movements—that reveal cognitive processes and individual strategies beyond what summative measures capture. Additionally, they fail to track participants' divergent heuristics for the same task, thereby masking individual differences that may be critical for personalized assessment. For example, when a participant makes an error, details about error detection, processing, and correction are not studied even though understanding these aspects can provide critical insights into a person's cognitive performance [7]. Some digital assessments use multiple tasks to estimate summative scores provided by traditional assessments. For example, the Drunk User Interfaces (DUI) uses tasks like swiping, typing, and two reaction tasks to measure how alcohol affects a person's motor coordination and cognition [33]. DUI uses various measures–based on distance, speed, time, and accuracy–to estimate a single summative score (for blood alcohol level). Other digital tools have made improvements to the traditional assessment. For example, traditional tools for capturing critical flicker frequency–a known proxy for cognitive performance–work only under specific ambient light intensities [20]. Beacon–a portable and self-administered tool–overcomes this limitation by adjusting the intensity of the light source in the device with respect to the ambient light [29]. While such approaches are promising, the need for dedicated hardware can still impose some limits on who gets to use such innovative tasks.

Recent digital tools have improved upon traditional assessments by offering multiple performance measures and supporting self-administration. However, they continue to prioritize summative outcomes while overlooking fine-grained cognitive processes—such as error recognition, correction, and task-solving strategies—that are critical for understanding how individuals perform. This highlights the need for digital tools that move beyond outcome metrics to capture the underlying cognitive processes during task execution.

2.3 Errors can provide useful insights, but are not tracked by cognitive performance assessments

One missing aspect across cognitive performance assessments is tracking error detection and correction. Errors can track the progression of cognitive performance over time [15, 34]. People with different neurological disorders make different numbers of errors and take varying durations to correct their errors [2, 39]. Errors demonstrate the potential to discriminate among groups with different cognitive performance [2]. Since motor performance–the ability to control and execute a movement task–reduces over time [46], older people take longer on cognitive performance assessments like the Trail Making Test that rely on motor performance [9]. However, errors on such tasks can still be less susceptible to factors like age [2]. Focusing on errors–*e.g.* by dividing them into concrete stages–can provide richer information than summative measures over the entire task [7]. For instance, understanding the breakdown of time duration across multiple phases in an error (initiation, detection, fix, completion) provides useful information that a measure like median speed over the entire trajectory does not [7].

Even when errors are tracked, doing so requires experts who observe people's behavior during an assessment [7]. Relying on experts to track errors comes with multiple challenges: it requires expert time, and observers might miss subtle errors due to inherent limits to the human visual perception system. Digital assessments can track errors automatically without the need for an expert. For example, the digital Trail Making Test automatically tracks the number of mistakes made [21]. Drunk User Interface (DUI) categorizes errors in a typing task based on whether it was corrected or not and tracks movement variability/inaccuracy [33].

Closer to our work, fine-finger tracking has been used to analyze the underlying cognitive processes during task performance. By examining continuous finger trajectories, different stages of decision-making and error processing can be inferred. For example, variations in movement speed and direction have been shown to reflect changes in confidence while performing a task and shifts in decisions during dragging tasks [18, 19]. These findings highlight the potential of using detailed finger trajectory data to uncover cognitive processes beyond what summative scores reveal.

Our research contributes ways to go beyond current error-tracking methods that track the number of mistakes/inaccurate attempts or rely on human perception. We use speed-time profiles to analyze different stages of performance—such as error detection, error correction, and strategies used—without relying on human perception. Such approaches offer a more nuanced understanding of cognitive performance beyond aggregate error counts.

3 The Crossroads System

Crossroads is a web-based tool for fine-finger tracking on touchscreen-based devices to assess cognitive performance. Crossroads is self-administered, takes less time compared to traditional cognitive assessments, and provides quantitative

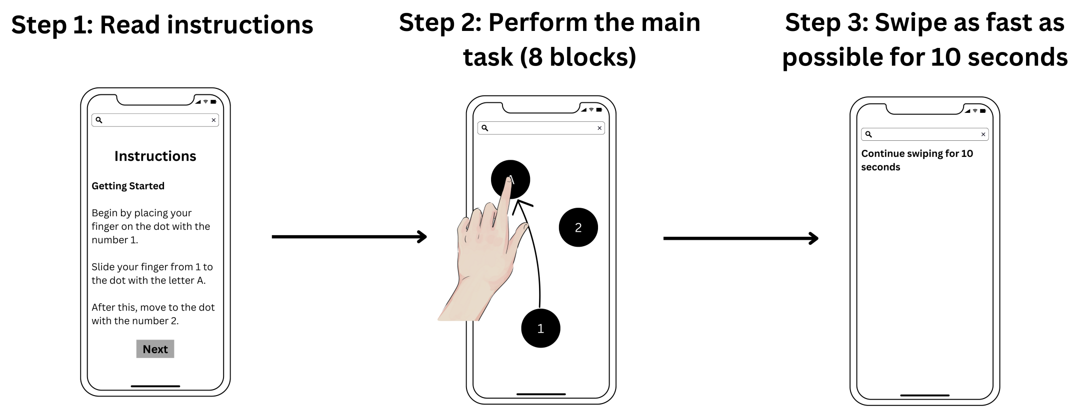


Fig. 3. Crossroads is a web-based touchscreen task that assesses cognitive activity without the need for expert administration. a) Crossroads provides instructions to perform the task, b) The main task is repeated 8 times to measure cognitive activity, c) The swiping task is used to calculate the maximum speed when cognitive load is minimal.

details about errors and multiple measures from the overall trajectory. Crossroads' design derives from the traditional Trail Making Test (TMT) [9]. We chose Trail Making Test as the base task for Crossroads because of two reasons: 1) TMT is a key part of many cognitive assessments [17, 24, 52] and 2) scores on TMT discriminates between different cognitive performance [41].

3.1 Task design

Crossroads comprises two tasks: 1) the *main task* where the goal for the user is to reach the final dot while alternating between numbers and letters; and 2) a *swiping task* where the user swipes their finger on the screen for 10 seconds as quickly as possible (Figure 3, Step 3). Crossroads tracks the trajectory and time for each finger movement. In Crossroads, a *drag* means pressing and moving a finger across the screen. Users drag the primary finger of their dominant arm between dots on the screen.

3.1.1 Main Task. The main task is a directed finger dragging task, where the user needs to reach a correct dot. The user alternatively drags their finger to a dot labeled with a number and then a dot labeled with a letter. The numbers and letters are organized in ascending order and the user needs to reach the final dot without lifting their finger. The main task is organized as 8 blocks (Figure 3, Step 2). Each drag (number-to-letter or letter-to-number) is considered to be a single *trial*. Each *block* consists of 10 trials (1-A-2-B-3-C-4-D-5-E-6).

Crossroads first displays instructions on how to perform the task. Instructions include the order in which finger must be dragged, a warning against lifting their finger during the task, and the need to complete the task correctly and quickly(Figure 3, Step 1).

The task starts with 3 dots labeled '1', 'A', and '2' (Figure 4a). Users drag their primary finger from the dot labeled '1' to the dot labeled 'A' (Figure 4b). When users enter the dot labeled 'A,' the dots labeled '1' and '2' disappear, and new dots labeled '2' and 'B' appear in different positions (Figure 4c). Users then need to drag their finger from 'A' to '2' (Figure 4d). This sequence goes on till the user reaches the dot labeled '6'. After completing a block (a sequence of 10 trials from 1 to 6), the user proceeds to the next block, or to the swiping task if all eight blocks are over.

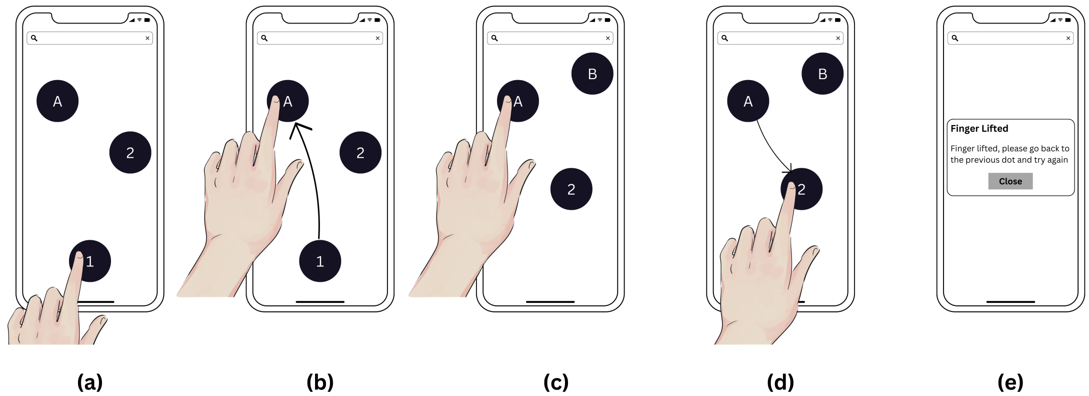


Fig. 4. Crossroads displays three dots labeled with numbers or letters. Users need to drag their fingers from one dot to another while alternating between numbers and letters in ascending order. Users are asked to restart from the previous dot if they lift their finger or touch an incorrect dot.

The position of the dots for a block is identical across all blocks and across all users. This enables easier comparisons between participants and between blocks without needing to normalize the position of the dots and the distances between them. Each dot has a fixed diameter of 80 pixels. Crossroads shows an error message when the finger is lifted or if the user touches the incorrect dot (Figure 4e). A trial restarts (i.e., goes back to the previous dot) after an error.

3.1.2 Swiping Task. The user is instructed to swipe their primary finger on a blank screen as fast as possible for ten seconds (Figure 3, Step 3). There is no goal or direction for this rapid, swiping movement. This data is used to estimate each user's maximum speed when there is minimal cognitive load. This maximum speed is then used to normalize speed values during data analysis. This operation helps Crossroads reduce the effects of individual-level motor performance during the main task.

3.2 Episodic measures

We define episodic measures as measures that are recorded for specific events that happen infrequently during finger movement. Specifically, Crossroads generates episodic measures for errors. In Crossroads, errors are categorized as mistakes and slips. *Mistakes* occur when the participant touches the incorrect dot (Figure 5a). For example, a mistake occurs when the participant drags their finger from the dot labeled '1' to '2' instead of 'A'. *Slips* occur when the user initiates movement towards an incorrect dot but changes direction to reach the correct dot (Figure 5a). For each block, such measures include: 1) number of mistakes, 2) number of slips, 3) slip duration, 4) slip detection time, 5) slip correction time, 6) slip initiation time, and 7) confidence gain.

3.2.1 Computing slip measures. Slips are calculated using two angular measures - *angle with the correct dot* and *angle with the incorrect dot* angle (Figure 6). Consider the case where P2 is the current position on the trajectory, P1 is the previous position on the trajectory, C is the center of the correct dot, and I is the center of the incorrect dot. For this case, the angle between the lines P2-P1 and P1-I is the *angle with the incorrect dot* and the angle between the lines P2-P1 and P1-C is the *angle with the correct dot.* The user moves in the *incorrect* direction when the *angle with the incorrect dot* is less than the *angle with the correct dot.* Similarly, the user moves in the *correct* direction when the *angle with the*

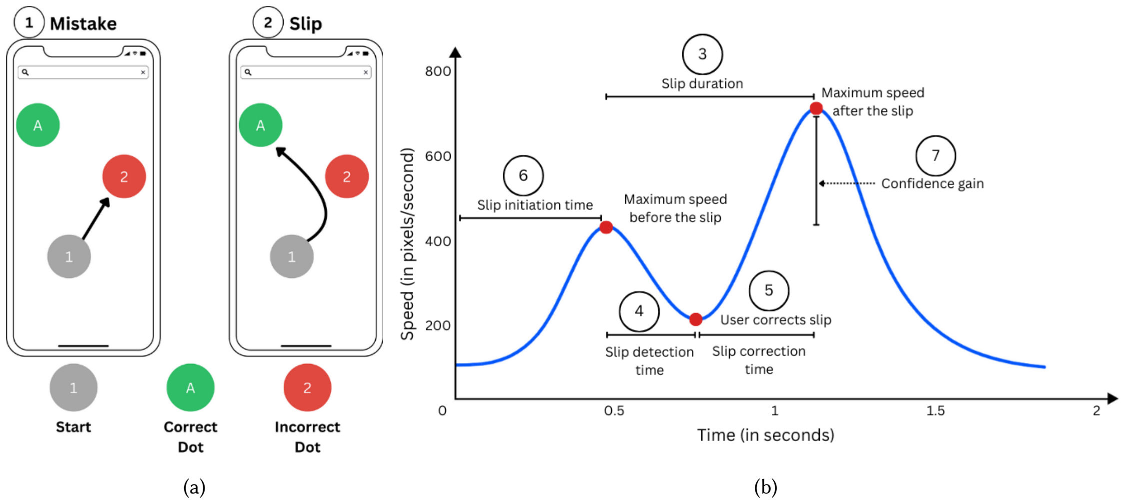


Fig. 5. Crossroads generates episodic measures for errors that include mistakes (1) and slips (2). A mistake occurs when the user drags their finger onto an incorrect dot. A slip occurs when the user moves their finger towards the incorrect dot and then changes direction to reach the correct dot. By analyzing changes in speed Crossroads tracks different stages of error processing, like slip duration (3), slip detection time (4), slip correction time (5), slip initiation time (6), and confidence gain (7) (figure inspired by [23]).

incorrect dot exceeds the *angle with the correct dot*. A slip occurs when the user moving in the *incorrect* direction starts moving in the *correct* direction. Crossroads assumes that a user moves towards either the correct or incorrect dot and not randomly away from both dots.

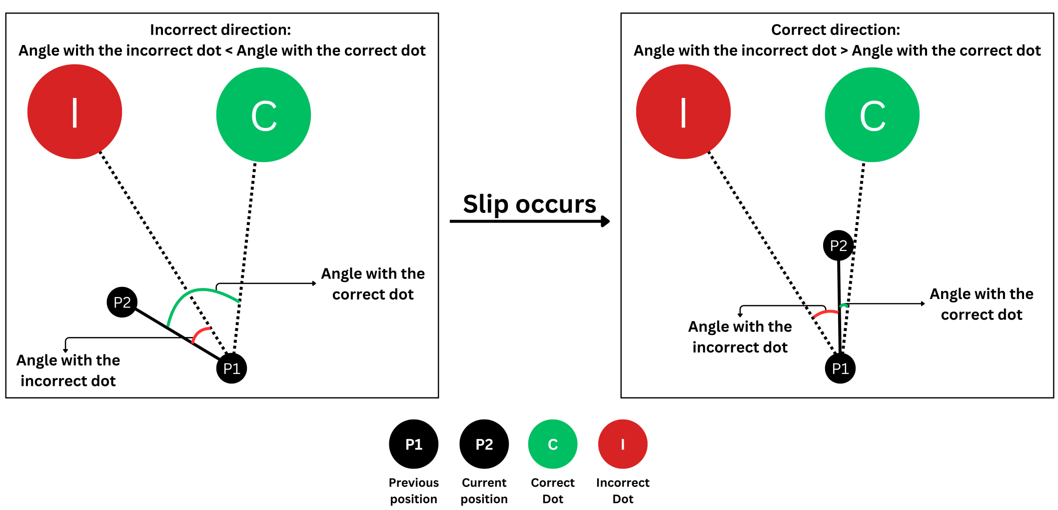


Fig. 6. Slips are identified using two angular measures: 1) *angle with the incorrect dot*, which is the angle between P2, P1 and I, and 2) *angle with the correct dot*, which is the angle between P2, P1 and C. User is moving in the incorrect direction when the *angle with the correct dot* is greater than the *angle with the incorrect dot*. Slips occur when a user moving in the incorrect direction starts moving in the correct direction

3.2.2 *Computing confidence gain and detection time.*

A slip has two segments: detection and correction (Figure 5b). The detection segment starts at the point where the speed starts decreasing and ends at the point where the speed reaches the minimum. The correction segment starts at the point of minimum speed and continues until the speed reaches a maximum. Decrease and increase in speeds have been empirically shown to correlate with a user's confidence in their movement [19].

Initiation time is the time taken to the point when the speed starts decreasing from the start of the trial (Figure 5b-(6)). This measure estimates the duration when the user was moving in the incorrect direction.

Detection time is the time taken from the maximum speed before a slip to the minimum speed during a slip (Figure 5b-(4)). This measure estimates how long it takes for a user to start correcting themselves after likely detecting the slip.

Correction time is the time taken to go from the minimum speed during a slip to the maximum speed after a slip (Figure 5b-(5)). This measure estimates how long it takes for a user to correct themselves after they detected the slip.

Slip duration is the sum of detection time and correction time (Figure 5b-(3)).

Confidence gain is the difference between the maximum speed before and after a slip (Figure 5b-(7)). Since different users can have different speeds when using Crossroads, the difference is normalized (per user) by dividing it by the user's maximum speed during the swiping task.

$$Confidence\ gain = \frac{maximum\ speed\ after\ slip - maximum\ speed\ before\ slip}{maximum\ speed\ from\ swiping}$$

3.3 Continuous measures

Continuous measures are calculated from the finger trajectory when the user's finger is in contact with the screen. Calculated at a sampling frequency of 43Hz, continuous measures are used to track the overall performance of a participant on the task. Crossroads measures the following continuous measures for each trial (Table 2): 1) time, 2) speed, 3) acceleration, 4) jerk, and 5) pause duration. For all continuous measures, we compute the minimum, maximum, mean, and median values. Additionally, we calculate the number of pauses.

4 Study

A user study tested the feasibility of Crossroads' design (task, user flow, and measures) to assess cognitive activity with a focus on characterizing errors. Concretely, the study was designed to answer questions around four themes:

(1) *Value of episodic measures*: What insights do episodic measures provide?
(2) *Reliability of continuous measures*: Do some continuous measures demonstrate high test-retest reliability across the blocks? Which ones?
(3) *Relevance of all measures*: Do some measures demonstrate a high correlation with scores on Montreal Cognitive Assessment? Which ones?
(4) *Acceptability of the tool*: What did participants think of Montreal Cognitive Assessment and Crossroads?

4.1 Methods

4.1.1 *Participants.*

Participants were students or staff members over the age of 18 at an educational institution. A total of 11 individuals took part in the study. Of these, 7 were aged 18–25 and 4 were aged 25–30. The group included 5 females and 6 males. In terms of race and ethnicity, 5 participants identified as Asian or Asian American, 4 as White, 1 as Hispanic/Latino/a/Latinx, and 1 as belonging to more than one racial or ethnic group. None of the participants reported any cognitive or motor impairments. Each participant received a $20 Amazon gift card for participating in the study.

4.1.2 Material. MoCA Duo–a digital version of Montreal Cognitive Assessment (MoCA)–was used to generate a summative score on MoCA for each participant [6]. Crossroads was used on a 10.9 inch iPad (10th generation) provided by the research team.

4.1.3 Procedure. The study was conducted in a quiet conference room at a university. The study followed the following steps: 1) obtaining consent to participate and being video recorded, 2) a digital version of Montreal Cognitive Assessment (MoCA) on the iPad, 3) Crossroads (main task across eight blocks and a swiping task), 4) a short interview, and 5) a survey. Across steps 2 and 3, consecutive participants were counter-balanced across using Crossroads (or MoCA) before the other. Other steps were the same for all participants. The survey and interview focused on participant experience, contextual information that might affect their performance (*e.g.*, the amount of sleep, or caffeine) and demographic information (*e.g.*, their age and education level). Each participant session was videotaped for future analysis.

4.1.4 Data Analysis. Speed vs time profiles were analyzed to better understand slips made by each participant. Kruskal-Wallis test was used to assess differences in speed values during slips among participants grouped by Montreal Cognitive Assessment scores.

Similar to prior work [38], we computed the Intraclass Correlation Coefficient, or ICC (single rating, absolute-agreement, 2-way mixed-effects model) to quantify the reliability of the different measures captured by Crossroads [4, 31]. We computed the reliability using ICC in three ways: for one block at a time, median of two blocks at a time, and median of three blocks at a time. The first block was used as a calibration block and excluded during all three ICC calculations. The last block (8th block) was excluded from data analysis when calculating ICC for the median of two blocks and the median of three blocks. We interpreted ICC using commonly-accepted thresholds [31]: 1) below 0.50: poor; 2) between 0.50 and 0.75: moderate; 3) between 0.75 and 0.90: good; 4) above 0.90: excellent.

Pearson correlation was computed between all measures and scores on Montreal Cognitive Assessment. We interpreted correlation values using commonly-accepted thresholds [45]: 1) below 0.10: negligible; 2) between 0.1 and 0.39: weak; 3) between 0.4 and 0.69: moderate; 4) between 0.7 and 0.89: strong; and 5) above 0.9: very strong. Acceptability of Crossroads was studied using participant's interview responses.

5 Results

Overall, Crossroads provides a rich understanding of slips and associated cognitive activities with its episodic measures. Crossroads produces reliable measures that also correlate with scores from standard Montreal Cognitive Assessment. Participants overwhelmingly found Crossroads to be acceptable, with many reporting that they enjoyed the task. Participants took median 27 mins (`max`: 39 mins, `min`: 19 mins, `mean`: 28 mins) to complete the entire session. Using Crossroads took median 1 min 23 sec (`min`: 56 sec, `max`: 2 mins 57 sec) The Montreal Cognitive Assessment took a median of 8 mins 25 sec (`min`: 6 mins 55 sec, `max`: 9 mins 09 sec).

5.1 Simple touchscreen-based dragging tasks provide a rich understanding of errors

Median speed during a slip increased with higher MoCA scores. Additionally, the number of errors (mistakes and slips) varied among participants with the same MoCA score, indicating individual differences beyond traditional cognitive assessment scores.

5.1.1 Participants showed different profiles during a slip. Analysis of the speed vs time profiles demonstrated that participants with lower scores on Montreal Cognitive Assessment (MoCA) moved slower and for longer duration while participants with higher scores on MoCA moved faster and for shorter duration (Figure 7). For example, P8 and P10 made multiple slips, but P10 showed a much longer slip duration. Note that P8 scored 30/30 on MoCA while P10 scored 26/30.

Median speed values showed an increasing trend across MoCA scores: 139 px/s (for MoCA = 26), 312 px/s (for MoCA = 27), 419 px/s (for MoCA = 28), 411 px/s (for MoCA = 29), and 783 px/s (for MoCA = 30) (Figure 7 f). A Kruskal-Wallis test indicated that there was a significant difference in speed across five scores of Montreal Cognitive Assessment (MoCA), χ^2 (4, N = 5) = 9.48, p = < .001.

Confidence gain was positive for all participants' slips (Figure 8a). The slip detection time decreased with the increase in scores on Montreal Cognitive Assessment (R^2 = 0.30, Figure 8b).

5.1.2 Trends during slip detection. Moderate correlation was found between scores on Montreal Cognitive Assessment and detection time (R^2 = 0.22, Table 1). Weak correlation was found between scores on Montreal Cognitive Assessment and median jerk during detection (R^2 = 0.03, Table 1).

Feature	R^2 **Value**	**Strength**
Detection time	0.22	Moderate
Slip duration	0.15	Weak
Correction time	0.15	Weak
Median jerk during a slip	0.07	Weak
Median jerk during correction	0.07	Weak
Median speed during a slip	0.05	Weak
Median speed during correction	0.05	Weak
Number of errors	0.04	Weak
Median jerk during detection	0.03	Weak
Number of slips	0.01	Negligible
Median speed during detection	0.01	Negligible

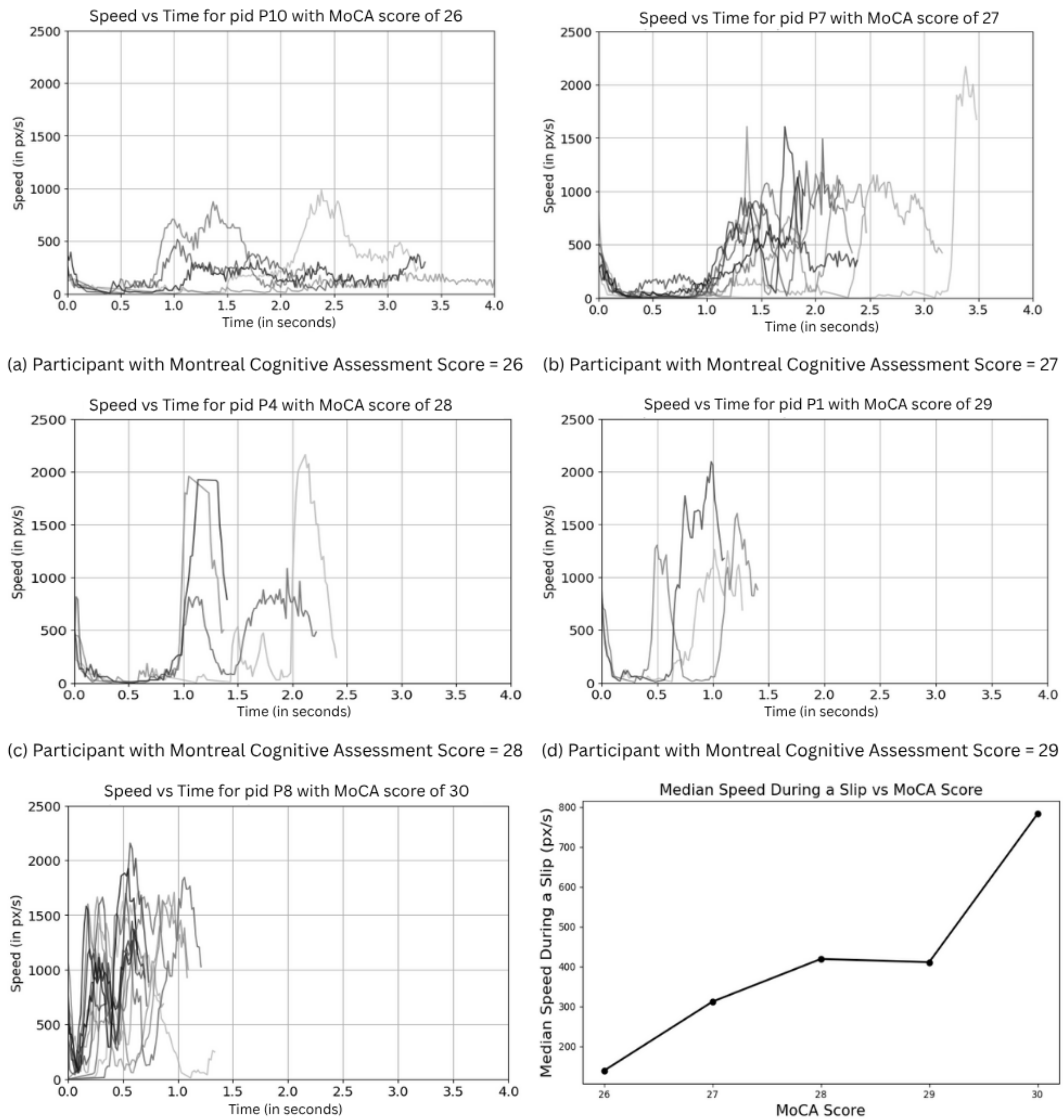
Table 1. R^2 value for features related to slips and its different segments

5.1.3 Trends during slip correction. Weak correlation was found between scores on Montreal Cognitive Assessment and correction time (R^2 = 0.15), median jerk (R^2 = 0.07), and median speed (R^2 = 0.05) (Table 1) during the correction segment of a slip.

5.1.4 Participants with same scores on Montreal Cognitive Assessment showed different numbers of slips. The number of errors (mistakes + slips) varied among participants with the same score on Montreal Cognitive Assessment. For example, two participants (P8 and P9) with a score of 30 on Montreal Cognitive Assessment 12 and 1 slips respectively. P8 moved as quickly as possible without pausing at each target. P9 moved slowly and paused at most target leading to a larger number of pauses (Figure 9) and lower median speed (Figure 10).

5.1.5 No correlation was found between scores on Montreal Cognitive Assessment and number of errors, slips, mistakes, and the continuous measures during slips. No correlation was found between scores on Montreal Cognitive Assessment and the number of errors (R^2 = 0.04) or slips (R^2 = 0.01) (Table 1)[1]. Weak correlation was found between scores on

[1]Number of errors is the sum of mistakes and slips



(a) Participant with Montreal Cognitive Assessment Score = 26 (b) Participant with Montreal Cognitive Assessment Score = 27

(c) Participant with Montreal Cognitive Assessment Score = 28 (d) Participant with Montreal Cognitive Assessment Score = 29

(e) Participant with Montreal Cognitive Assessment Score = 30 (f) Median speed during slip vs MoCA Score

Fig. 7. (a-e) Speed vs time profiles of individual slips demonstrate varying performance. Different colored lines in a plot show different trajectories for the same MoCA score. People with scores on Montreal Cognitive Assessment = 26 and 27 moved slower and for longer duration while people with higher scores on Montreal Cognitive Assessment moved faster and for shorter duration. We have selected representative users for each MoCA category for this plot. Supplementary material shows speed profiles for all users. (f) Median speed during slip increased with MoCA score

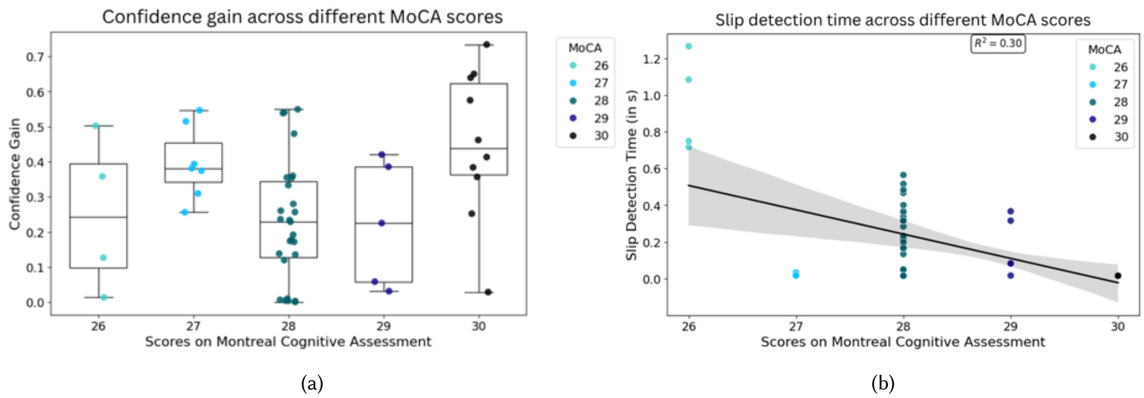


Fig. 8. Confidence gain is always positive, and participants of different scores on Montreal Cognitive Assessment show varying slip detection time.

Montreal Cognitive Assessment and slip duration ($R^2 = 0.15$), median jerk ($R^2 = 0.07$), and median speed ($R^2 = 0.05$) (Table 1). Only two participants (P7 and P11) made mistakes while using Crossroads. Since mistakes are rare in our study, we have focused our error analysis chiefly on slips.

5.2 Touchscreen-based dragging tasks provide reliable measures

When calculated one block at a time, five measures showed good reliability (0.90 > ICC > 0.75) (Table 2): median jerk (0.87), median time taken (0.87), median speed (0.86), mean speed (0.8), and mean time taken (0.79). When the median of two blocks was used, one measure showed excellent reliability (ICC > 0.90): median jerk (0.95), and six other features showed good reliability: median time taken (0.89), median speed (0.88), mean speed (0.86), mean time taken (0.84), error

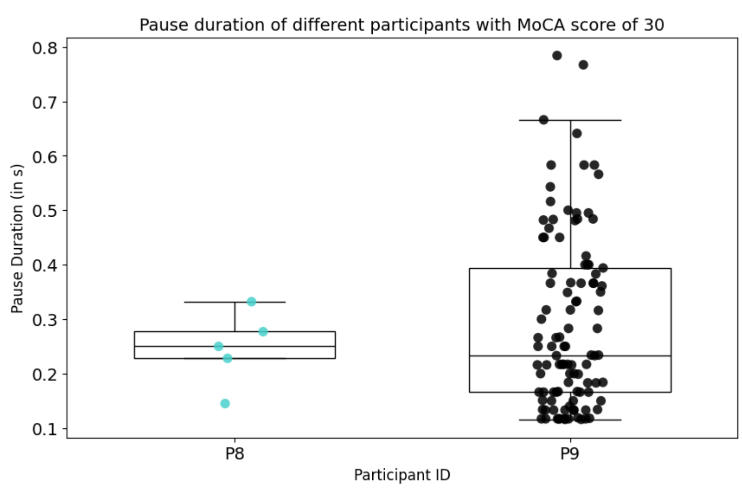


Fig. 9. P8 and P9 both have a score on Montreal Cognitive Assessment of 30 but show varying numbers of pauses and pause duration.

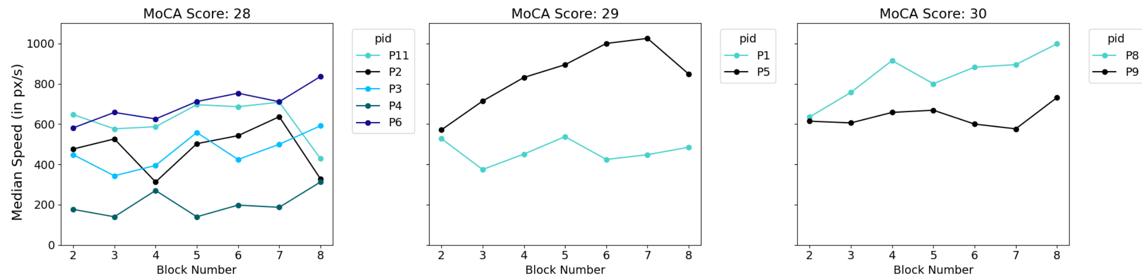


Fig. 10. Participants with same scores on Montreal Cognitive Assessment show different changes in median speed across blocks. Participant with Montreal Cognitive Assessment score of 26 and 27 are not shown since there were one participant each in these categories.

rate (0.81), and minimum jerk (0.81). When the median of three blocks was used, median jerk (0.94) and median time taken (0.9) showed excellent reliability.

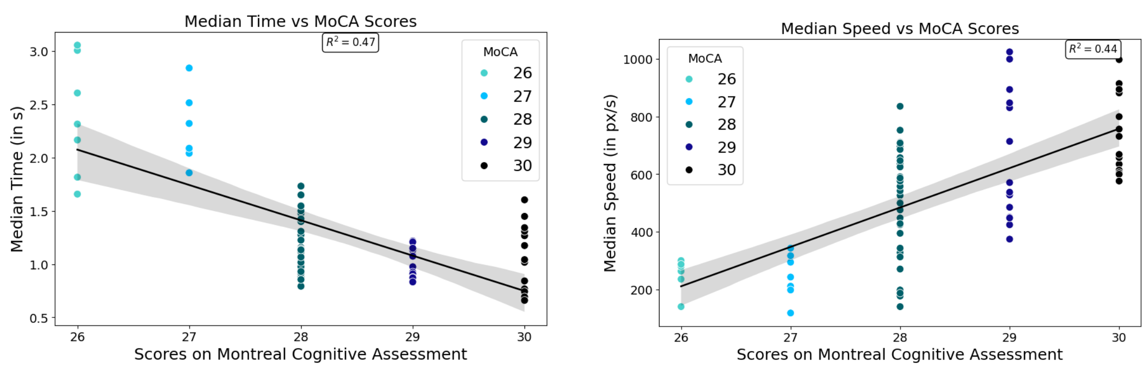
Measure	**Description**	**1 at a time**	**2 at a time**	**3 at a time**
Median jerk	Median of the rate of change of acceleration during a block	0.87	0.95	0.94
Median time taken	Median of the time taken to complete a trial during a block	0.87	0.89	0.9
Median speed	Median of the movement speed during a block	0.86	0.88	0.88
Minimum speed	Minimum movement speed during a block	0.81	0.94	0.92
Mean speed	Average movement speed during a block	0.8	0.86	0.85
Mean time taken	Average time taken to complete a trial during a block	0.79	0.84	0.82
Error rate	Percentage of distance traveled in the incorrect direction during a block	0.71	0.81	0.8
Minimum jerk	Minimum rate of change of acceleration during a block	0.71	0.81	0.77
Number of pauses	Number of times speed was 0 pixels/second for over 100 milliseconds during a block	0.59	0.61	0.63
Median acceleration	Median of the rate of change of speed during a block	0.53	0.75	0.65
Maximum pause duration	Maximum duration for which speed was 0 pixels/second for over 100 milliseconds during a block	0.51	0.5	0.71
Maximum time taken	Maximum time taken to complete a trial during a block	0.45	0.61	0.77
Number of errors	Number of errors (mistakes or slips) made during a block	0.37	0.56	0.57
Number of slips	Number of slips made during a block	0.3	0.53	0.45

Number of mistakes	Number of mistakes during a block	0.27	0.47	0.88
Median pause duration	Median duration for which speed was 0 pixels/second for over 100 milliseconds during a block	0.2	0.7	0.78
Mean pause duration	Mean duration for which speed was 0 pixels/second for over 100 milliseconds during a block	0.2	0.63	0.73
Maximum speed	Maximum movement speed during a block	0	0	0
Minimum acceleration	Minimum rate of change of speed during a block	0	0	0
Maximum jerk	Maximum rate of change of acceleration during a block	0	0	0
Mean jerk	Average rate of change of acceleration during a block	0	0	0
Maximum acceleration	Maximum rate of change of speed during a block	0	0	0
Mean acceleration	Mean rate of change of speed during a block	0	0	0

Table 2. Reliability scores for all measures. ICC scores were calculated for three different groups - individual blocks, two consecutive blocks, and three consecutive blocks. The table is sorted by descending order of individual block's ICC scores.

5.3 Measures from Crossroads show strong correlation to scores on Montreal Cognitive Assessment

Six of the measures are strongly correlated with the Montreal Cognitive Assessment (MoCA) scores (Table 3). These measures are maximum time taken (-0.816), median time taken (-0.786), mean time taken (-0.785), minimum time taken (-0.762), maximum pause duration (-0.760), and median speed (0.713). Figure 11 plots median time and median speed with scores on Montreal Cognitive Assessment.



(a) (b)

Fig. 11. Median time and median speed per block for every participant shows a moderate correlation to Montreal Cognitive Assessment ($R^2 = 0.47,\ 0.44$ respectively)

Measure (per participant)	Pearson correlation	Strength	Direction
Maximum time taken	**-0.816**	**Strong**	Negative
Median time taken	**-0.786**	**Strong**	Negative
Mean time taken	**-0.785**	**Strong**	Negative
Minimum time taken	**-0.762**	**Strong**	Negative
Maximum pause duration	**-0.760**	**Strong**	Negative
Number of pauses	-0.615	Moderate	Negative
Slip detection time	-0.394	Moderate	Negative
Number of mistakes	-0.296	Weak	Negative
Number of error	-0.212	Weak	Negative
Median acceleration	-0.179	Weak	Negative
Number of slips	-0.112	Weak	Negative
Maximum acceleration	-0.076	Negligible	Negative
Maximum jerk	-0.076	Negligible	Negative
Mean jerk	-0.075	Negligible	Negative
Maximum speed	-0.074	Negligible	Negative
Mean acceleration	-0.073	Negligible	Negative
Minimum acceleration	0.077	Negligible	Positive
Error rate	0.182	Weak	Positive
Mean pause duration	0.21	Weak	Positive
Confidence gain	0.262	Weak	Positive
Median jerk	0.471	Moderate	Positive
Median pause duration	0.563	Moderate	Positive
Minimum jerk	0.625	Moderate	Positive
Mean speed	0.676	Moderate	Positive
Median speed	**0.713**	**Strong**	Positive

Table 3. Crossroads produces six measures that have a strong Pearson correlation with scores on Montreal Cognitive Assessment.

5.4 Acceptability of Crossroads and Montreal Cognitive Assessment

Participants enjoyed using Crossroads and typically spent less than two minutes using the tool without needing any help from the research team.

5.4.1 Crossroads requires very little participant time. Participants took a median of 1 min 23 sec (min: 56 sec, max: 2 mins 57 sec) to perform the dragging task on Crossroads (Figure 12). Participants took a median of 8 mins 25 sec (min: 6 mins 55 sec, max: 9 mins 09 sec) to complete the Montreal Cognitive Assessment.

5.4.2 All participants used Crossroads without needing any help from the research team. Many reported enjoying the task. All participants used Crossroads without any additional instructions or research team's assistance. When asked if they had any difficulty understanding any of the instructions provided, none reported any issues.

> "It was relatively simple, and the instructions were pretty straightforward." –P10

When asked about their general experience using Crossroads, four of eleven participants mentioned that they enjoyed the task.

> "When I was doing the dots, honestly, it felt more of like a fun challenge than anything." –P11

> "I think especially once I noticed that it was the same thing and trying to figure out the fastest way through it, it was sort of in the moment engaging and interesting." –P8

The interviews and video analysis showed that most of the participants prioritized either speed or accuracy over both.

> "To connect the things in the right order. Go fast. But that was secondary to correctness." –P6

> "My goal was to do it as quickly as possible. And then I think the very first time, the first thing I did was a mistake. And so I was like, okay, let's let's try and do it as fast as possible, but still try to be accurate." –P8

Some participants reported trying to focus on both but still prioritized one over the other.

> "To do it as fast as possible and as accurately as possible. . . ..I even slowed down at certain points when I thought I was going too fast." –P11

Eight participants noticed that the dots were in the same location across different blocks. Most realized this in the middle of the task (around the third or fourth block).

> "I think I did. I'm not sure, but for a second, I thought the numbers and letters were moving to the same spot every time I did the task...I noticed it after like the third or fourth time." –P2

5 participants noticed that sometimes their hands would cover a target, and they would have to move around to find it.

> " Sometimes, my hand would cover certain parts of the screen. So I'd have to like either like move it weird or sort of move out of the way and not be following that. Just to make sure I could see all the targets." –P8

5.4.3 Challenges with standardized assessments. When asked, some participants mentioned that the memory task in Montreal Cognitive Assessment made them anxious.

> "I think when I had to remember the words, that was throughout, I was trying to remember them while doing other tasks." –P11

> "I was like slightly worried. When you said, remember these words, and then I'll ask them for you at the end" –P8

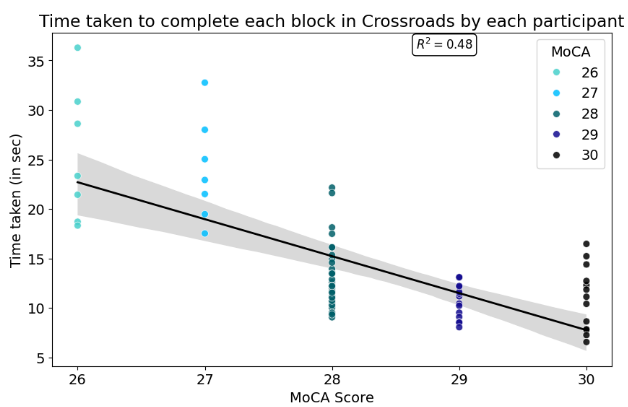


Fig. 12. Time taken by participants to use Crossroads decrease with increase in Montreal Cognitive Assessment scores.

Some participants found the Montreal Cognitive Assessment to be too easy and weren't sure what they were being assessed for-

> "I didn't really know what you were testing as much as, if I'm looking at like dragging, you're probably testing, you know, accuracy and speed and all that. But if I'm drawing a clock, I don't really know like what you're looking for" –P5 (Note: Drawing a clock is a task in the Montreal Cognitive Assessment)

> "Yeah, it was fine. I was very curious what it was measuring, I'd say, because it wasn't like hard." –P6

5.4.4 External factors affecting performance. One participant reported consuming caffeine right before the task. Some participants felt like the time of the day had an effect on their performance.

> "I had a long day today. So maybe I could have done better. Like the thing about recollecting the words, maybe if I just got up and I was totally fresh or something like that, I could have guessed the fifth word that I missed out too." –P4

> "No, I guess this is the middle of the day, and I'm probably in my prime cognitive space right now. If this was at 10 pm, I'd do worse." –P6

Some participants reported that other daily activities like upcoming exams might have influenced their performance.

> "Maybe a little stress about the exam, but not significant." –P9

6 Discussion

Our research demonstrates that error profiles developed with a touchscreen-based task offer insights into cognitive performance. Additionally, fine-finger tracking yields reliable measures that correlate strongly with standard cognitive assessments. In this section, we discuss how Crossroads-like efforts can complement traditional assessments by providing rich performance data beyond summative scores. Additionally, we explore design possibilities to improve the effectiveness of digital tools for different healthcare needs and across different populations.

6.1 By providing information about errors, digital tools can complement traditional assessments and meet the needs of experts and participants

Unlike traditional methods, Crossroads tracks and analyzes errors automatically without relying on human perception. In this section, we discuss how such measures can be useful and propose ways to use such data.

6.1.1 Measures provided by Crossroads overcome some challenges faced by traditional assessments. Analysis of the different measures produced by Crossroads helps us understand the different strategies used by participants. For example, P8 made multiple slips, moved fast, and did not pause often, while P9 made only one slip, moved slowly, and paused more frequently. Despite these drastic differences in tool usage, both participants had the same MoCA score of 30. This highlights how current cognitive assessments (like MoCA)'s summative scores might hide underlying participant strategies. Such differences can be important for clinical purposes; for instance, in the real-world, such differences might be related to *compensatory strategies*–techniques or actions used to overcome or adapt to limitations or challenges in order to maintain effective performance or functioning–that are common among people with early cognitive decline [47]. Current assessments lack the possibility of tracking compensatory strategies. For example, the pen-and-paper Trails Making Test measures only the time taken and does not track compensatory strategies like pausing at each dot to avoid making errors.

In our study, participants demonstrated different slip profiles. For example, those with higher scores on the Montreal Cognitive Assessment made quicker slips with higher average speeds (Figure 7). Additionally, slip detection time correlated moderately with scores on the Montreal Cognitive Assessment, implying that people with higher scores on the Montreal Cognitive Assessment spent less time detecting their slips. By tracking a participant's detailed performance metrics and error patterns *during* the task (and not just at the end), digital tools can enable more precise monitoring of cognitive changes than traditional summative scores.

A total of eleven measures–such as median time taken and median speed–showed excellent or good reliability scores. Several of these measures strongly correlated with MoCA scores. These results suggest that digital tools can obtain reliable fine-granular measures that can complement traditional assessments. Additionally, acceptability results with Crossroads shows that digital tools (when designed appropriately) can be self-administered. During the study, participants used instructions provided in the tool without any issues or interventions by the research team. Reducing the need for trained experts can lower costs associated with cognitive assessments, making frequent monitoring more feasible. By taking less than 2 minutes to use, our digital tool shows that such monitoring can be easy to perform and enables frequent testing.

Slower finger or hand movements often reduce overall task speed, particularly in older adults who commonly experience motor impairments or age-related motor decline. Without accounting for this, cognitive assessments risk mixing motor limitations with cognitive limitation, leading to inaccurate interpretations. To separate these factors, it is important to include motor-only baseline tasks that isolate motor performance with minimal cognitive load. For instance, the swiping task in Crossroads estimates a user's maximum movement speed by providing minimal constraints and cognitive load, effectively providing a proxy for motor-dominant performance. Measurements from such motor-dominant tasks can be used to normalize cognitive performance metrics, as demonstrated by Crossroads' approach of normalizing confidence gain based on maximum swiping speed.

6.1.2 Limitations and next steps. Our research offers insights into how errors tracked with fine-finger trajectory on a touchscreen-based task can be used to assess cognitive performance. We list some limitations to Crossroads and ways to improve. Crossroads' current limitations fall along three dimensions: 1) generating clinical scores, 2) tracking confounding variables; and 3) study population.

Crossroads provides multiple measures that are reliable and correlate with standard cognitive scores but does not directly estimate standard cognitive assessment scores from these measures. Estimating clinical scores from tracked measures (like Hevelius [38]) will enhance the tool's relevance in healthcare settings. Future work can involve developing interpretable regression models that estimate standard cognitive scores (like MoCA scores) from Crossroads' measures.

Participants reported that factors such as time of day, caffeine intake, and academic stress (e.g., upcoming exams) may have influenced their performance. Future longitudinal at-home studies could track and evaluate these effects by collecting self-reports on contextual factors and administering tasks under varying conditions. Analyzing within-subject variance might help isolate the impact of real-world factors on cognitive performance. Fatigue, a common confound in lengthy assessments, may further impair performance. Beyond tracking factors like fatigue, digital task design can also mitigate fatigue effects by reducing the number of blocks or trials while studying its trade-offs with overall data quality.

This study involved a sample of 11 participants, all of whom were younger adults (18 to 30 years old). Additionally, MoCA scores were not evenly distributed, with five out of eleven participants scoring 28 out of 30. Future studies will likely benefit from larger and more diverse samples—both in age and cognitive scores—through broader deployment or targeted recruitment based on clinical screening.

6.2 Which changes to such digital tools' design could improve their use for healthcare needs across varied demographics and longitudinal deployment?

In this section, we discuss how digital tools might help in healthcare processes. Our understanding and proposals are rooted in the user study (e.g., participant responses) and clinical needs.

6.2.1 Refining the tool. Digital cognitive assessments–like digital version of whack-a-mole and the Rapid Visual Information Processing task [14, 50]–often instruct participants to optimize both speed and accuracy to better reflect real-world performance. This balance is critical, as focusing solely on accuracy—for example, moving more slowly to reduce errors—can obscure true cognitive ability. At one end, some users may deliberately move slowly—either as a compensatory strategy or due to motor impairments—resulting in fewer errors that mask underlying cognitive challenges [32]. At the other end, other users may prioritize speed, leading to a higher error rate that exaggerates cognitive impairment. Enforcing minimum and maximum speed thresholds and providing real-time feedback via the tool's interface can help mitigate these distortions and yield realistic estimates of performance.

One possibility for Crossroads-like digital tools is remote longitudinal use for frequent health assessments. In our study, multiple participants noticed that the dots were identically positioned across the blocks. Noticing the identical position of dots may lead to improved performance over time due to familiarity rather than actual cognitive change. This improvement might be a sign of *learning effect*—improvements in performance due to repeated exposure rather than a change in cognition [3]. Learning effects need to be minimized in longitudinal studies. To minimize such learning effects in longitudinal use, varying the task design within and between sessions can prevent improvements driven by repetition. Two strategies include randomizing dot positions across blocks and introducing multiple incorrect options instead of a single incorrect dot. Prior work, such as Hevelius at Home, has successfully applied Fitts' law to dynamically adjust target size and distance, maintaining task difficulty over repeated sessions [38]. Additionally, learning effects themselves may not be entirely unfavorable when properly interpreted. The ability to learn and adapt to new tasks reflects important aspects of cognitive functioning and can serve as a valuable indicator of cognitive performance [22].

6.3 Digital tools can be used to assist different populations and provide opportunities for self-tracking and self-experimenting

Digital tools have the potential to enable scalable and frequent cognitive monitoring, supporting different populations and applications such as early detection, intervention tracking, and self-experimentation.

6.3.1 Older adults and people with Mild Cognitive Impairment can benefit from such digital tools. Two key populations of interest in cognitive performance research are older adults and individuals with mild cognitive impairment (MCI) [1, 11]. Easy to use digital cognitive assessments will benefit both groups. Designing these tools for self-administered use will enable use at home, without the need for expert supervision or clinic visits—especially valuable for those with mobility challenges, transportation barriers, or living in rural and under-served areas. For individuals with MCI, incorporating digital assessments into daily or weekly routines promotes awareness of their cognitive performance and allows for early detection of subtle changes over time. Frequent digital testing can make it possible to detect *presymptomatic* shifts in behavior—such as increasing errors or delayed error correction time—which can prompt timely clinical intervention.

Caregivers play a crucial role in supporting older adults and individuals with mild cognitive impairment, yet relying solely on their observations to assess cognitive changes can be challenging due to variability in observation and potential biases. Digital cognitive assessments can support caregiving work and tracking by providing quantitative scores for cognitive performance.

6.3.2 Digital tools can be used to track efficacy of cognitive interventions. Cognitive interventions use structured tasks and strategies—such as mnemonics for memory, timers for attention, and checklists for planning—to improve specific cognitive skills [10, 42]. Their effectiveness is typically measured by expert assessments, conducted in institutional labs and clinics, before and after the intervention [10]. However, this provides performance numbers at only two points in time. To better understand whether an intervention improves cognitive performance, frequent and ongoing assessments might help. Longitudinal tracking can reveal which interventions are effective and how their impact evolves over time.

7 Conclusion

This paper reports on the design of a novel tool Crossroads that tracks cognitive activity using fine-finger tracking on touchscreen-based devices. A feasibility study with eleven participants evaluated the efficacy of tracking two kinds of measures from finger trajectory: episodic measures focused on errors (slips, mistakes, slip duration/detection/correction), and continuous measures (speed, time, jerk, acceleration and pause duration) from the entire trajectory. Our results demonstrate the utility of our tool: episodic measures and speed-time profiles provide insights into participant behavior during errors. Additionally, multiple continuous features are reliable and correlate strongly with scores on a standard cognitive assessment. Participants overwhelmingly found the tool to be engaging. Median usage time was less than two minutes and none of the participants required help from the research team. We believe tracking fine-finger function with touchscreen devices provides a novel opportunity to understand cognitive activity in ways that go beyond summative scores provided by current cognitive assessments. Our long-term objective continues to be to design and refine such tools. Deploying such digital tools in longitudinal, real-world settings can likely bring benefits to scientific research and clinical work.

References

[1] Nicole D Anderson. 2019. State of the science on mild cognitive impairment (MCI). *CNS spectrums* 24, 1 (2019), 78–87.

[2] Lee Ashendorf, Angela L Jefferson, Maureen K O'Connor, Christine Chaisson, Robert C Green, and Robert A Stern. 2008. Trail Making Test errors in normal aging, mild cognitive impairment, and dementia. *Archives of Clinical Neuropsychology* 23, 2 (2008), 129–137.

[3] Claudia Bartels, Martin Wegrzyn, Anne Wiedl, Verena Ackermann, and Hannelore Ehrenreich. 2010. Practice effects in healthy adults: a longitudinal study on frequent repetitive cognitive testing. *BMC neuroscience* 11 (2010), 1–12.

[4] John J Bartko. 1966. The intraclass correlation coefficient as a measure of reliability. *Psychological reports* 19, 1 (1966), 3–11.

[5] Martin Behrens, Martin Gube, Helmi Chaabene, Olaf Prieske, Alexandre Zenon, Kim-Charline Broscheid, Lutz Schega, Florian Husmann, and Matthias Weippert. 2023. Fatigue and human performance: an updated framework. *Sports medicine* 53, 1 (2023), 7–31.

[6] Jody-Lynn Berg, January Durant, Gabriel C Léger, Jeffrey L Cummings, Ziad Nasreddine, and Justin B Miller. 2018. Comparing the electronic and standard versions of the Montreal Cognitive Assessment in an outpatient memory disorders clinic: a validation study. *Journal of Alzheimer's Disease* 62, 1 (2018), 93–97.

[7] Brianne Magouirk Bettcher, Tania Giovannetti, Laura Macmullen, and David J Libon. 2008. Error detection and correction patterns in dementia: A breakdown of error monitoring processes and their neuropsychological correlates. *Journal of the International Neuropsychological Society* 14, 2 (2008), 199–208.

[8] Alberto Blanco-Campal, Unai Diaz-Orueta, Ana Belen Navarro-Prados, Teresa Burke, David J Libon, and Melissa Lamar. 2021. Features and psychometric properties of the Montreal Cognitive Assessment: Review and proposal of a process-based approach version (MoCA-PA). *Applied Neuropsychology: Adult* 28, 6 (2021), 658–672.

[9] Christopher R Bowie and Philip D Harvey. 2006. Administration and interpretation of the Trail Making Test. *Nature protocols* 1, 5 (2006), 2277–2281.

[10] Verena Buschert, Arun LW Bokde, and Harald Hampel. 2010. Cognitive intervention in Alzheimer disease. *Nature Reviews Neurology* 6, 9 (2010), 508–517.

[11] Roberto Cabeza, Marilyn Albert, Sylvie Belleville, Fergus IM Craik, Audrey Duarte, Cheryl L Grady, Ulman Lindenberger, Lars Nyberg, Denise C Park, Patricia A Reuter-Lorenz, et al. 2018. Maintenance, reserve and compensation: the cognitive neuroscience of healthy ageing. *Nature Reviews*

Neuroscience 19, 11 (2018), 701–710.

[12] Nicole Carson, Larry Leach, and Kelly J Murphy. 2018. A re-examination of Montreal Cognitive Assessment (MoCA) cutoff scores. *International journal of geriatric psychiatry* 33, 2 (2018), 379–388.

[13] Joyce YC Chan, Sarah TY Yau, Timothy CY Kwok, and Kelvin KF Tsoi. 2021. Diagnostic performance of digital cognitive tests for the identification of MCI and dementia: A systematic review. *Ageing Research Reviews* 72 (2021), 101506.

[14] Jennifer T Coull, CD Frith, R Sx J Frackowiak, and PM Grasby. 1996. A fronto-parietal network for rapid visual information processing: a PET study of sustained attention and working memory. *Neuropsychologia* 34, 11 (1996), 1085–1095.

[15] Elizabeth A Crocco, Rosie Curiel Cid, Marcela Kitaigorodsky, Gabriella A Grau, Jessica M Garcia, Ranjan Duara, Warren Barker, Cesar L Chirinos, Rosemarie Rodriguez, and David A Loewenstein. 2021. Intrusion errors and progression of cognitive deficits in older adults with mild cognitive impairment and PreMCI states. *Dementia and geriatric cognitive disorders* 50, 2 (2021), 135–142.

[16] Lilian De Greef, Mayank Goel, Min Joon Seo, Eric C Larson, James W Stout, James A Taylor, and Shwetak N Patel. 2014. Bilicam: using mobile phones to monitor newborn jaundice. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 331–342.

[17] Dean C Delis, Edith Kaplan, and Joel H Kramer. 2001. Delis-Kaplan executive function system. *Assessment* (2001).

[18] Dror Dotan, Florent Meyniel, and Stanislas Dehaene. 2018. On-line confidence monitoring during decision making. *Cognition* 171 (2018), 112–121.

[19] Dror Dotan, Pedro Pinheiro-Chagas, Fosca Al Roumi, and Stanislas Dehaene. 2019. Track it to crack it: dissecting processing stages with finger tracking. *Trends in cognitive sciences* 23, 12 (2019), 1058–1070.

[20] Auria Eisen-Enosh, Nairouz Farah, Zvia Burgansky-Eliash, Uri Polat, and Yossi Mandel. 2017. Evaluation of critical flicker-fusion frequency measurement methods for the investigation of visual temporal resolution. *Scientific reports* 7, 1 (2017), 15621.

[21] Robert P Fellows, Jessamyn Dahmen, Diane Cook, and Maureen Schmitter-Edgecombe. 2017. Multicomponent analysis of a digital Trail Making Test. *The Clinical Neuropsychologist* 31, 1 (2017), 154–167.

[22] Rocío Fernández-Ballesteros, María Dolores Zamarrón, and Lluís Tàrraga. 2005. Learning potential: a new method for assessing cognitive impairment. *International Psychogeriatrics* 17, 1 (2005), 119–128.

[23] Krzysztof Z Gajos, Katharina Reinecke, Mary Donovan, Christopher D Stephen, Albert Y Hung, Jeremy D Schmahmann, and Anoopum S Gupta. 2020. Computer mouse use captures ataxia and parkinsonism, enabling accurate measurement and detection. *Movement Disorders* 35, 2 (2020), 354–358.

[24] Richard C Gershon, Molly V Wagster, Hugh C Hendrie, Nathan A Fox, Karon F Cook, and Cindy J Nowinski. 2013. NIH toolbox for assessment of neurological and behavioral function. *Neurology* 80, 11_supplement_3 (2013), S2–S6.

[25] Mayank Goel, Elliot Saba, Maia Stiber, Eric Whitmire, Josh Fromm, Eric C Larson, Gaetano Borriello, and Shwetak N Patel. 2016. Spirocall: Measuring lung function over a phone call. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 5675–5685.

[26] Philip D Harvey. 2019. Domains of cognition and their assessment. *Dialogues in clinical neuroscience* 21, 3 (2019), 227–237.

[27] Machteld Huber, J André Knottnerus, Lawrence Green, Henriëtte Van Der Horst, Alejandro R Jadad, Daan Kromhout, Brian Leonard, Kate Lorig, Maria Isabel Loureiro, Jos WM Van Der Meer, et al. 2011. How should we define health? *Bmj* 343 (2011).

[28] Edith Kaplan. 1988. The process approach to neuropsychological assessment. *Aphasiology* 2, 3-4 (1988), 309–311.

[29] Ravi Karkar, Rafal Kocielnik, Xiaoyi Zhang, Jasmine Zia, George N Ioannou, Sean A Munson, and James Fogarty. 2018. Beacon: Designing a portable device for self-administering a measure of critical flicker frequency. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–27.

[30] Naomi Kokubo, Yuma Yokoi, Yuji Saitoh, Miho Murata, Kazushi Maruo, Yoshitake Takebayashi, Issei Shinmei, Sadanobu Yoshimoto, and Masaru Horikoshi. 2018. A new device-aided cognitive function test, User eXperience-Trail Making Test (UX-TMT), sensitively detects neuropsychological performance in patients with dementia and Parkinson's disease. *BMC psychiatry* 18 (2018), 1–10.

[31] Terry K Koo and Mae Y Li. 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine* 15, 2 (2016), 155–163.

[32] Kaiyan Ling, Hang Zhao, Xiangmin Fan, Xiaohui Niu, Wenchao Yin, Yue Liu, Cui Wang, and Xiaojun Bi. 2024. Model Touch Pointing and Detect Parkinson's Disease via a Mobile Game. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 2 (2024), 1–24.

[33] Alex Mariakakis, Sayna Parsi, Shwetak N Patel, and Jacob O Wobbrock. 2018. Drunk user interfaces: Determining blood alcohol level through everyday smartphone tasks. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

[34] Daniel C Marson, SM Annis, B McInturff, A Bartolucci, and LE Harrell. 1999. Error behaviors associated with loss of competency in Alzheimer's disease. *Neurology* 53, 9 (1999), 1983–1983.

[35] Kimford J Meador. 1998. Cognitive side effects of medications. *Neurologic clinics* 16, 1 (1998), 141–155.

[36] Alex J Mitchell, Steven Kemp, Julián Benito-León, and Markus Reuber. 2010. The influence of cognitive impairment on health-related quality of life in neurological disease. *Acta Neuropsychiatrica* 22, 1 (2010), 2–13.

[37] 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.

[38] Vineet Pandey, Nergis C Khan, Anoopum S Gupta, and Krzysztof Z Gajos. 2023. Accuracy and Reliability of At-Home Quantification of Motor Impairments Using a Computer-Based Pointing Task with Children with Ataxia-Telangiectasia. *ACM Transactions on Accessible Computing* 16, 1 (2023), 1–25.

[39] R Pezzetta, ME Wokke, SM Aglioti, and KR Ridderinkhof. 2022. Doing it wrong: a systematic review on electrocortical and behavioral correlates of error monitoring in patients with neurological disorders. *Neuroscience* 486 (2022), 103–125.
[40] Janice M Ranson, Elżbieta Kuźma, William Hamilton, Graciela Muniz-Terrera, Kenneth M Langa, and David J Llewellyn. 2019. Predictors of dementia misclassification when using brief cognitive assessments. *Neurology: Clinical Practice* 9, 2 (2019), 109–117.
[41] Ralph M Reitan. 1958. Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and motor skills* 8, 3 (1958), 271–276.
[42] Galeno J Rojas, Veronica Villar, Monica Iturry, Paula Harris, Cecilia M Serrano, Jorge A Herrera, and Ricardo F Allegri. 2013. Efficacy of a cognitive intervention program in patients with mild cognitive impairment. *International Psychogeriatrics* 25, 5 (2013), 825–831.
[43] Luis Salvador-Carulla, Ramona Lucas, José Luis Ayuso-Mateos, and Marta Miret. 2014. Use of the terms" Wellbeing" and" Quality of Life" in health sciences: a conceptual framework. *The European Journal of Psychiatry* 28, 1 (2014), 50–65.
[44] Carmen Sandi. 2013. Stress and cognition. *Wiley Interdisciplinary Reviews: Cognitive Science* 4, 3 (2013), 245–261.
[45] Patrick Schober, Christa Boer, and Lothar A Schwarte. 2018. Correlation coefficients: appropriate use and interpretation. *Anesthesia & analgesia* 126, 5 (2018), 1763–1768.
[46] Charles D Smith, GH Umberger, EL Manning, JT Slevin, DR Wekstein, FA Schmitt, WR Markesbery, Z Zhang, GA Gerhardt, RJ Kryscio, et al. 1999. Critical decline in fine motor hand movements in human aging. *Neurology* 53, 7 (1999), 1458–1458.
[47] Sarah Tomaszewski Farias, Maureen Schmitter-Edgecombe, Alyssa Weakley, Danielle Harvey, Katherine G Denny, Cheyanne Barba, Jason T Gravano, Tania Giovannetti, and Sherry Willis. 2018. Compensation strategies in older adults: association with cognition and everyday function. *American Journal of Alzheimer's Disease & Other Dementias®* 33, 3 (2018), 184–191.
[48] Nash Unsworth, Gene A Brewer, and Gregory J Spillers. 2012. Variation in cognitive failures: An individual differences investigation of everyday attention and memory failures. *Journal of Memory and Language* 67, 1 (2012), 1–16.
[49] Marco Vacante, Gordon K Wilcock, and Celeste A de Jager. 2013. Computerized adaptation of The Placing Test for early detection of both mild cognitive impairment and Alzheimer's disease. *Journal of clinical and experimental neuropsychology* 35, 8 (2013), 846–856.
[50] Bruce Wallace, Frank Knoefel, Rafik Goubran, Philippe Masson, Amanda Baker, Brianna Allard, Victor Guana, and Eleni Stroulia. 2017. Detecting cognitive ability changes in patients with moderate dementia using a modified "whack-a-mole" game. *IEEE Transactions on Instrumentation and Measurement* 67, 7 (2017), 1521–1534.
[51] David Wechsler. 1955. Wechsler adult intelligence scale–. *Archives of Clinical Neuropsychology* (1955).
[52] Kathleen A Welsh, Nelson Butters, Richard C Mohs, D Beekly, S Edland, G Fillenbaum, and A Heyman. 1994. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part V. A normative study of the neuropsychological battery. *Neurology* 44, 4 (1994), 609–609.
[53] Jacob O Wobbrock and Julie A Kientz. 2016. Research contributions in human-computer interaction. *interactions* 23, 3 (2016), 38–44.
[54] Ke Yu, Shangang Zhang, Qingsong Wang, Xiaofei Wang, Yang Qin, Jian Wang, Congyang Li, Yuxian Wu, Weiwen Wang, and Hang Lin. 2015. Development of a computerized tool for the chinese version of the montreal cognitive assessment for screening mild cognitive impairment. *International psychogeriatrics* 27, 2 (2015), 213–219.