Self-Administered Digital Cognitive Assessment for Older Adults via Fine-Finger Tracking on Touchscreen-based Devices

ANONYMOUS AUTHOR(S)

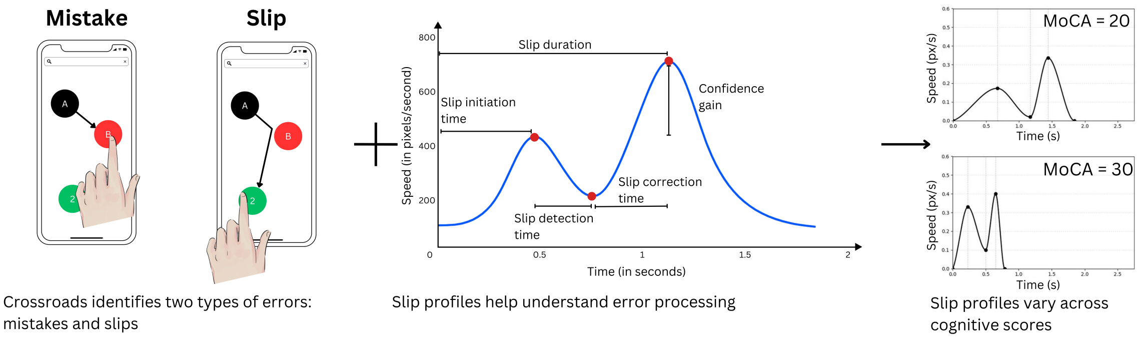


Fig. 1. Our work proposes to understand cognitive performance with a new tool *Crossroads* that tracks errors and continuous measures from fine-finger trajectories. A study with 12 older adults and 15 younger adults found that slip profiles during errors provide information that complements standard assessments. Additionally, many measures are reliable and correlate well with standard cognitive assessment scores.

Can we design digital cognitive assessments that are accessible, low-burden, and capable of tracking more than summative scores? Traditional cognitive assessments require in-person interaction with trained experts, making them infrequent and difficult to access–particularly for older adults facing mobility or transportation barriers. Additionally, current cognitive assessments reduce rich behavioral cues to a summative score, missing the process-level information embedded in how people perform a task. We present *Crossroads*, a self-administered, web-based tool that uses fine-finger tracking on touchscreen devices to assesses cognitive performance. *Crossroads* generates two types of measures: *episodic* measures focused on errors (like slips and mistakes) and pauses; and *continuous* measures (like speed, time, acceleration, and jerk). A user study with 12 older adults (aged 65+ years) and 15 younger adults (aged 18-30 years)–spanning a broad range of scores on a standard cognitive assessment–evaluated the reliability, validity, and acceptability of Crossroads. Multiple continuous measures demonstrate high test-retest reliability across blocks. Two measures–median time taken and number of pauses–show strong correlations with standard cognitive scores (r = -0.72). *Slip profiles* reveal useful differences in error processing across participants with different cognitive scores, offering process-level insight that summative assessments do not provide. Older adults completed Crossroads in a median of 2 minutes 19 seconds. All participants completed the task without additional assistance. Our work demonstrates that fine-finger tracking on touchscreen devices provides new opportunities for understanding cognitive performance in ways that complement and extend what current cognitive assessments offer.

CCS Concepts: • **Human-centered computing → Interactive systems and tools**; *Empirical studies in HCI*.

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

1 Introduction

Cognitive performance often declines with age, leading to reduced attention, processing speed, and memory [15]. Subtle changes in cognition can go unnoticed without systematic fine-granular tracking. Traditional cognitive assessments–like the Montreal Cognitive Assessment (MoCA)–collapse rich behavioral information into summative scores [6, 35]. Additionally, current cognitive assessments require in-person administration by trained clinicians–limiting frequent testing and making them largely inaccessible in everyday life. These barriers are especially pronounced for older adults, who may face mobility limitations, transportation constraints, or limited access to clinical resources [26, 32]. Digital health tools can help make assessments remotely accessible and with low burden [10].

Most digital cognitive assessments are direct copies of pen-and-paper tests that produce the same summative scores [12]. Such digital copies of traditional assessments also require expert supervision. Current self-administered digital variants–such as digital Trail Making Tests and Whack-a-Mole adaptations–provide a richer set of measures (e.g., number of pauses, movement speed, etc.) but continue to reduce performance to outcome-level metrics [20, 47]. None of these tools track rich process-level information; *e.g.*, how a user detects, processes, and corrects an error. Such process-level information can be particularly useful for older adults, since error detection and correction patterns are sensitive markers of cognitive decline and can help distinguish healthy aging from early impairment [5, 13, 34]. Older adults with age-related declines in attention and executive function are less able to detect and correct errors [1, 5]; yet existing digital tools do not support processing errors to understand cognitive performance.

Our research presents *Crossroads* [1], a self-administered, web-based tool that uses fine-finger tracking on touchscreen devices to assess cognitive performance. Crossroads operationalizes the insight that errors provide crucial information about cognitive performance. The speed and movement trajectory of a finger during an error encodes information about how a user detects, processes, and corrects their errors. Crossroads builds on the Trail Making Test–a widely validated cognitive assessment–and decomposes finger trajectories into two types of measures: episodic measures (e.g., slips, mistakes, and pauses), and continuous measures (e.g., speed, time, acceleration, and jerk). A rapid swiping task establishes a user's baseline motor performance, allowing cognitive measures to be normalized with respect to individual motor performance. We evaluated Crossroads in a lab study–using the Montreal Cognitive Assessment as the ground truth–with 12 older adults (65-90 years) and 15 younger adults (18–30 years).

Our study demonstrates that Crossroads produces reliable, and valid measures as a self-administered cognitive assessment tool for older adults. Two measures–median time taken and number of pauses–showed strong correlations with MoCA scores (r=-0.72) and high test-retest reliability. Multiple other features moderately correlated with MoCA scores. Slip profiles revealed useful differences in error processing across participants' slips: initiation time and slip duration varied significantly across MoCA groups, offering process-level insights. Older adults completed the task in a median of 2 minutes 19 seconds. Our work makes two contributions to designing new cognitive assessments for older adults [51]:

(1) *an artifact contribution* with a novel, self-administered tool for tracking cognitive performance within minutes using fine-finger tracking.
(2) *an empirical contribution* with a mixed-methods lab study that assesses the feasibility of this approach for older adults with measures of reliability, validity, and acceptability.

Our tool and the empirical findings point toward a broader design direction: touchscreen-based tasks with fine-finger tracking can complement traditional cognitive assessments by tracking *how they perform*, rather than relying solely on what they produce. Such touchscreen-based tasks can be useful for health monitoring (especially for older adults), where reducing reliance on clinicians, shortening testing time, and enabling self-administration at home are all critical.

[1] The name Crossroads is inspired by the figure of speech "at a crossroads"–which mean a situation that requires making a choice between many options. When using the Crossroads tool, users need to make decisions on which direction to move their finger.

2 Related Work

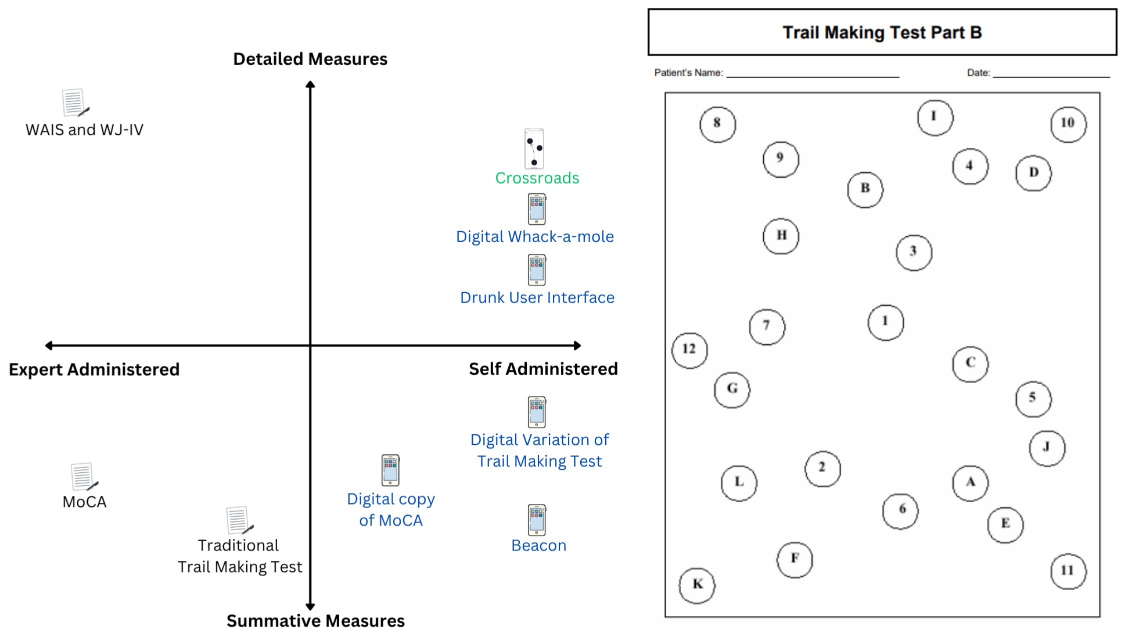
Our research draws inspiration from the strengths and limitations of current cognitive assessments. Our ideas build on prior work around tracking errors that have been useful in multiple domains to understand cognitive performance.

2.1 Cognitive assessments require in-person interaction with experts and do not provide process-level information

Current cognitive assessments provide useful outcome measures but struggle with multiple challenges, especially for older adults. Traditional pen-and-paper screening tests provide a summative score; *e.g.*, the Montreal Cognitive Assessment (MoCA) provides a score between 0 and 30 [35]. These tests usually involve tasks such as memorizing a short list of words, naming objects shown in pictures, copying shapes, and arithmetic problems. Such assessments suffer from multiple challenges. First, participants need to rely on experts to administer and evaluate such screening test like MoCA. Second, standard thresholds used to screen participants for cognitive impairments are not agreed upon by experts [11]; summative scores 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 those with less education [11]. *False negatives* can occur as well: a participant might have a cognitive impairment, but the assessment fails to detect it due to rigid threshold criteria [40]. Finally, accessing traditional cognitive assessments often and on-demand is not possible due to the reliance on expert presence. 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 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 [6, 28]. Additionally, current assessments do not track the degree of correctness or the nature of errors made. Screening tests–like the Montreal Cognitive Assessment (MoCA)–do not differentiate between slight mistakes, which might indicate a small change in cognitive performance, and a completely incorrect response, which could suggest a severe decrease in cognitive performance. For example, in MoCA, participants need to draw a clock with specific features in the clock-drawing task. The evaluation is based solely on the drawing of the completed clock. If two participants make different errors–such as misplacing a number and positioning the clock hands incorrectly–and correct these errors during the task, both receive full scores on the task. The initiation, detection, and correction of errors are not tracked independently in traditional cognitive assessments. We believe that the research 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* is the final output on the task).

Digital cognitive assessments face similar challenges. Many digital cognitive assessments are digital copies of traditional pen-and-paper assessments that provide summative scores [12]. For example, like their pen-and-paper versions, 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 [4, 12, 29, 46, 53]. Some digital variants of traditional cognitive assessments, like the digital Trail Making Test and digital Whack-a-Mole, output multiple measures rather than a single summative score (Figure 2a) [20, 47]. For example, in the traditional (pen-and-paper) Trail Making Test, participants are asked to connect the dots labeled with either a letter or number (Figure 2b) [7]. Participants need 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 performance. Digital versions of the Trail



(a) (b)

Fig. 2. a) A design space summarizing different cognitive assessments along two dimensions: who administers the assessment (x-axis), and nature of measures tracked (y-axis). Some traditional assessments, like the Trail Making Test [7], produce a single summative score (time taken). Cognitive assessments are typically administered by experts. Others require substantial expert time (60 to 90 minutes) and produce detailed measures, like Wechsler Adult Intelligence Scale [48]. Our work contributes *Crossroads* which assesses cognitive performance 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 (also known as *task completion time*). Participants connect the dots while alternating between letters and numbers in an ascending order.

Making Test track additional measures like the number of pauses, pause duration, lifts, lift duration, number of errors, time inside each circle, and time between circles [20]. Such additional measures provided by digital variants are useful but overlook infrequent but useful occurrences–like self-corrections or exploratory movements–that reveal cognitive processes and individual strategies beyond what summative measures track. 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 cognitive performance [5].

Overall, we argue that developing digital cognitive assessments that reduce reliance on experts, take less time, and provide information about process-level measures of cognitive performance–like errors–can be both novel and useful.

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

One missing aspect across cognitive assessments is tracking error detection and correction. Errors can track the progression of cognitive performance over time [13, 34]. Errors demonstrate the potential to discriminate among groups with different cognitive performance [1]. Participants with different neurological disorders make different numbers of errors and take varying durations to correct their errors [1, 38]. Since motor performance–the ability to control and execute a movement task–reduces over time [45], older adults take longer on cognitive performance assessments like the Trail Making Test that rely on motor performance [7]. However, errors on such tasks can still be less susceptible to factors like age [1]. Focusing on errors–*e.g.* by dividing them into concrete stages–can provide richer information than summative measures over the entire task [5]. For instance, understanding the breakdown of time duration across multiple phases in an error (initiation, detection, fix, completion) provides useful information that a summative measure like median speed over the entire trajectory does not [5].

Even when errors could be tracked, doing so requires experts who need to observe participants' behavior during an assessment [5]. Relying on experts to track errors comes with multiple challenges: it requires expert time, and observers might miss subtle errors due to inherent limits of the human visual perception system. Digital assessments can track errors automatically without the need for an expert. For example, digital versions of the Trail Making Test automatically tracks the number of mistakes made [20]. However, such assessments do not track and analyze the different stages and process-level information of errors.

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 [17, 18]. 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 errors or rely on human perception. We characterize errors into slips and mistakes, and automatically analyze error detection and correction.

3 Context: Older Adults as the Target Population

Cognitive decline is among the most significant health challenges facing aging populations. An estimated 15-30% of adults over 65 experience mild cognitive impairment, and the proportion increases substantially with age [27, 37]. For many older adults, early cognitive changes–including reduced executive function and slower processing speed–are clinically meaningful but may go undetected between infrequent clinical visits [26, 32]. Tracking these changes over time, especially in everyday settings, is a recognized priority in both clinical and research contexts [44].

Older adults are disproportionately affected by the constraints of traditional cognitive assessment. Most standard assessments–including MoCA and similar screening instruments–rely on clinician availability and require commute/travel to a clinic (which might be difficult for older adults with mobility or transportation constraints) [12, 26, 32, 35]. As a result, cognitive performance is typically measured infrequently (often once a year), making it difficult to track ongoing changes that may be clinically relevant, such as those caused by aging and neurological decline [9, 44]. Older adults who are cognitively impaired also take longer to complete assessments, potentially causing tensions and delays in time-sensitive clinical workflows [52].

Errors during cognitive tasks are particularly informative for older adults. Error patterns–including increased frequency of errors and difficulty detecting and correcting errors–are established markers of cognitive decline and can help distinguish healthy aging from early impairment [1, 5]. Specifically, age-related declines in attention and executive function reduce the ability to monitor and correct one's own errors, a capacity known as error monitoring [13, 34]. Despite the importance of tracking errors, most traditional cognitive assessments focus entirely on final outcomes (summative scores or completion time) and discard the process-level information embedded in how errors unfold.

Older adults may have specific preferences regarding how cognitive performance is tracked. Designing for older adults requires attending to both the technical validity of measures and to the acceptability and perceived utility of the tool in everyday life [22]. Given these considerations–the importance of error monitoring in aging, the access barriers posed by traditional assessments, and the importance of acceptability–we focus on older adults as the primary population for Crossroads.

4 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, designed to take a few minutes, and provides quantitative details about errors and multiple measures from the overall trajectory. Crossroads' design derives from the traditional Trail Making Test (TMT) [7]. We chose Trail Making Test as the base task for Crossroads because TMT is a key part of many cognitive assessments [16, 24, 49]; and scores on TMT discriminate between differing cognitive performance [41].

4.1 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 (Figure 3, Step 2); 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.

4.1.1 Main Task. The main task comprises a directed finger dragging task. 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. The user needs to reach the final dot without lifting their finger. The main task is organized as eight blocks. 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) (Figure 3, Step 2).

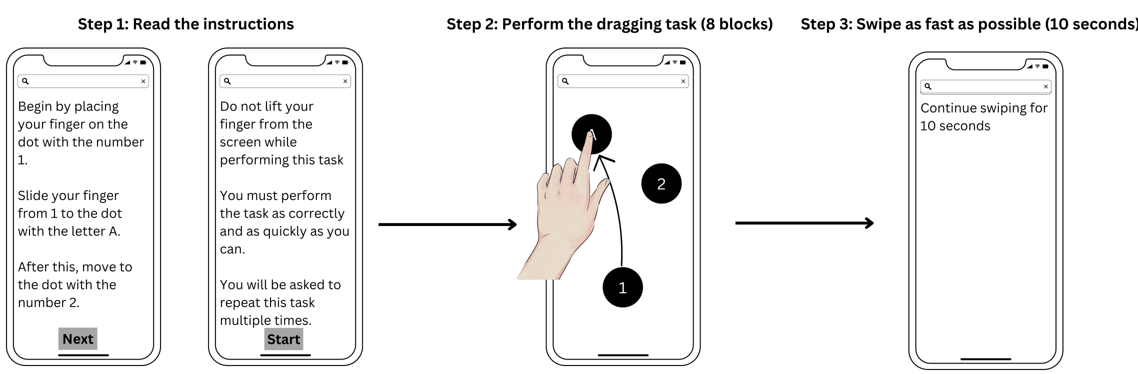


Fig. 3. Crossroads is a web-based touchscreen task that assesses cognitive performance without the need for expert administration. 1) Crossroads provides instructions to perform the task, 2) The main task is repeated eight times to measure cognitive performance, 3) The swiping task is used to calculate user's maximum speed when cognitive load is minimal.

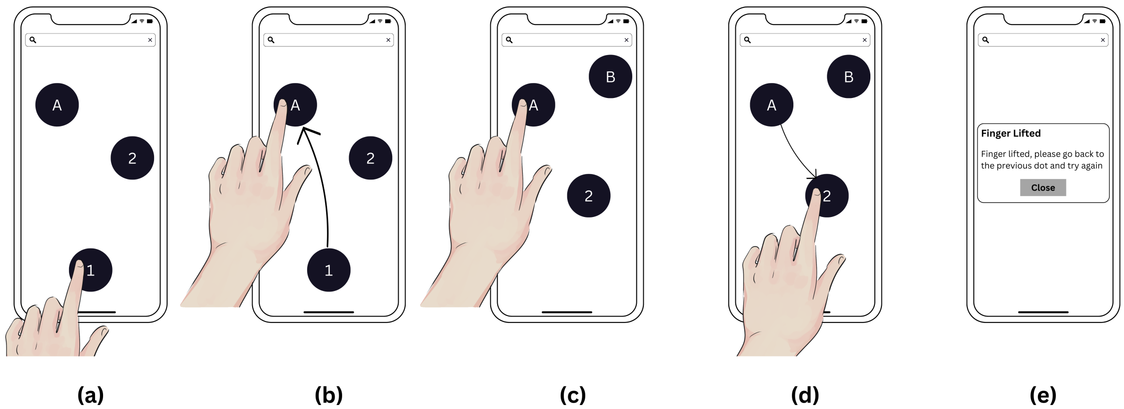


Fig. 4. During the main task, Crossroads displays three dots labeled with numbers or letters. Users need to drag their fingers from one dot to another while alternating between numbers and letters in ascending order. Users are asked to restart from the previous dot if they lift their finger or touch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

Crossroads first displays instructions about the task. Instructions include the order in which the dots should be connected, 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). The alternating 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 completed.

The position of the dots for a block is identical across all blocks and across all users. Identical positions enables easier comparisons between participants and between blocks without needing to account for varying distances between the dots. Each dot has a fixed diameter of 80 pixels. Crossroads uses relative positions to ensure that dots are identical across devices. The actual position of the dots scale in relation to the width and height of the device's screen. Details about the relative position of dots is provided in the supplementary document. 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.

4.1.2 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). No goal or direction is provided to the user for this rapid, swiping movement. The data from the swiping task is used to estimate each user's maximum speed when there is minimal cognitive load. The maximum speed (during swiping) is then used to normalize speed values during data analysis. Normalizing speed values helps Crossroads reduce the effects of individual-level motor performance during the main task.

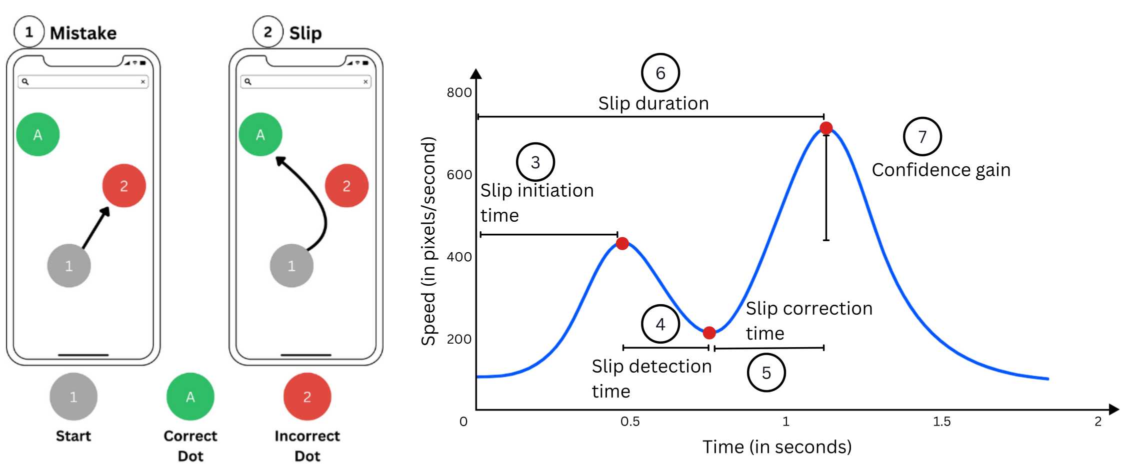
4.2 Episodic measures

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* and *pauses*. In Crossroads, errors are categorized as mistakes and slips. *Mistakes* occur when a user touches the incorrect dot (Figure 5a). For example, a mistake occurs when the user drags their finger from the dot labeled '1' to '2' instead of 'A'. *Slips* occur when a user initiates movement towards an incorrect dot but changes direction to reach the correct dot (Figure 5a). For each block, episodic 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, 7) confidence gain, 8) number of pauses, and 9) pause duration.

4.2.1 Computing slips and pauses. Slips are calculated using two angular measures - *angle with the correct dot* and *angle with the incorrect dot* (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. 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 incorrect dot* exceeds the *angle with the correct dot*. For each infinitesimal movement the user makes (tracked by the touchscreen), Crossroads calculates if the user is moving in the correct direction or not. A slip occurs when the user moving in the *incorrect* direction begins to move in the *correct* direction. Crossroads assumes that a user moves towards either the correct or incorrect dot and not in another random direction.

Similar to prior work, a pause occurs when the speed is 0 pixels/second for over 100 milliseconds during a block [36]. Crossroads measures the number of pauses and the duration of each pause. For the duration of pauses, we compute the minimum, maximum, mean, and median values.

4.2.2 Computing a slip profile. A slip has three segments: initiation, detection and correction (Figure 5b). The initiation segment starts at the beginning of the trial and ends at the point where the speed starts decreasing. The detection



(a) (b)

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 initiation time (3), slip detection time (4), slip correction time (5), slip duration (6), and confidence gain (7) (figure inspired by Figure 4 on page 7 in [36]).

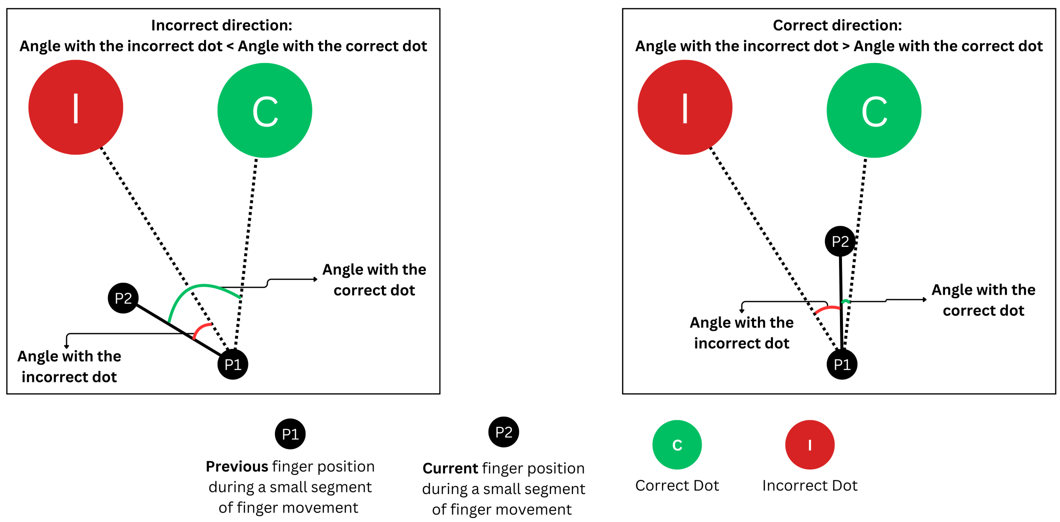


Fig. 6. Slips are identified using two angular measures: 1) *angle with the incorrect dot*, which is the angle between P2, P1 and I, and 2) *angle with the correct dot*, which is the angle between P2, P1 and C. A user is moving in the incorrect direction when the *angle with the correct dot* is greater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

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.

Initiation time (in sec.) is the time taken to the point when the speed starts decreasing from the start of the trial (Figure 5b-(3)). Initiation time estimates the duration when the user was moving in the incorrect direction.

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

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

Slip duration (in sec.) is the sum of initiation time, detection time and correction time (Figure 5b-(6)).

Confidence gain represents the difference between the maximum speed before and after a slip (Figure 5b-(7)). Decrease and increase in speeds have been empirically shown to correlate with a user's confidence in their movement [18]. Since different users can have different finger movement speeds, Crossroads normalizes the speed difference for each user by dividing it by the user's maximum speed during the swiping task.

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

4.3 Continuous measures

Continuous measures are calculated from the finger trajectory when the user's finger is in contact with the screen. Continuous measures are used to track the overall performance of a user on the task. Crossroads measures the following continuous measures for each trial: 1) time, 2) speed, 3) acceleration, and 4) jerk. For all continuous measures, we compute the minimum, maximum, mean, and median values.

5 Study

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

(1) *Reliability of measures*: Do some measures demonstrate high test-retest reliability across the blocks? Which ones?
(2) *Validity of measures*: Do some measures demonstrate a high correlation with scores on Montreal Cognitive Assessment? Which ones?
(3) *Value of errors*: What insights do error-based measures–like initiation time, detection time and correction time–provide?
(4) *Acceptability of the tool*: What did participants think of the Crossroads task and the Montreal Cognitive Assessment?

5.1 Methods

5.1.1 Participants. We collected data from two participant cohorts. We recruited older adults (aged 65+) participants. Older adults typically exhibit a broader range of cognitive performance than younger adults, allowing us to validate Crossroads across a wide range of cognitive scores. We excluded potential participants who reported any motor or vision impairments since the study involved the use of a touchscreen device. One participant's finger kept lifting from the screen due to a subtle tremor unobservable to the unaided eye; they were excluded from the data analysis since they could not perform the task. 12 individuals participated in the study (Table 1): 1 was aged 65-70, 2 were aged 70-75, 2 were aged 75-80, 4 were aged 80-85, and 3 were aged 85-90 years old. The cohort included 7 females and 5 males. All participants identified as White. 11 participants had a master's or higher education degree and 1 participant had a high school degree. Each participant received a $20 Amazon gift card for participating in the study. The older adult participants had MoCA scores ranging from 20 to 29 (on a scale of 0 to 30).

To test the feasibility of Crossroads across different age groups, we recruited students or staff members over the age of 18 at an educational institution (Table 1). 15 individuals participated: 10 were aged 18–25 and 5 were aged 25–30. The cohort included 6 females and 9 males. 8 participants identified as Asian or Asian American, 5 as White, 1 as Hispanic/Latino/a/Latinx, and 1 as belonging to more than one racial or ethnic group. 3 participants had an associate degree, 4 had a bachelor's degree, 4 had a high school degree, and 4 had a master's or higher education degree. None of the participants reported any cognitive or motor impairments. Each participant received a $20 Amazon gift card for participating in the study.

5.1.2 Material. MoCA Duo–a digital version of Montreal Cognitive Assessment (MoCA)–was used to generate a summative score on MoCA for each participant [4]. Participants used Crossroads on a 10.9 inch iPad (10th generation) provided by the research team; continuous measures were computed at an average sampling frequency of 41Hz.

5.1.3 Procedure. The study was conducted in a quiet room at a retirement center (for older adults) and at a university (for younger adults). Participants completed five steps during the study: 1) provided consent to participate and being video recorded, 2) completed a digital version of Montreal Cognitive Assessment (MoCA) on the iPad, 3) used Crossroads

ID	Age group	Gender	Education	Race or Ethnicity	MoCA score (0 to 30)
O1	75 - 80	Male	Master's degree or higher education	White	25
O2	85 - 90	Female	Master's degree or higher education	White	20
O3	80 - 85	Female	Master's degree or higher education	White	22
O4	85 - 90	Female	Master's degree or higher education	White	20
O5	80 - 85	Female	Master's degree or higher education	White	28
O6	85 - 90	Male	Master's degree or higher education	White	27
O7	70 - 75	Female	Master's degree or higher education	White	27
O8	75 - 80	Male	Master's degree or higher education	White	29
O9	80 - 85	Female	Master's degree or higher education	White	25
O10	80 - 85	Male	Master's degree or higher education	White	25
O11	65 - 70	Male	Master's degree or higher education	White	24
O12	70 - 75	Female	High school graduate	White	26
Y1	25 - 30	Female	Master's degree or higher education	Asian, Asian American	27
Y2	18 - 25	Male	Bachelor's degree	Asian, Asian American	26
Y3	18 - 25	Male	High school graduate	White	30
Y4	18 - 25	Male	Bachelor's degree	Asian, Asian American	29
Y5	18 - 25	Female	Master's degree or higher education	Asian, Asian American	28
Y6	18 - 25	Male	High school graduate	White	28
Y7	18 - 25	Female	Master's degree or higher education	Asian, Asian American	28
Y8	18 - 25	Male	Bachelor's degree	White	29
Y9	25 - 30	Male	Bachelor's degree	White	28
Y10	25 - 30	Female	Master's degree or higher education	Asian, Asian American	27
Y11	18 - 25	Male	High school graduate	Hispanic/Latino/a/Latinx	30
Y12	18 - 25	Female	Associate degree	Asian, Asian American	30
Y13	25 - 30	Male	Associate degree	White	26
Y14	25 - 30	Female	Associate degree	Mixed	28
Y15	18 - 25	Male	High school graduate	Asian, Asian American	26

Table 1. Participants O1 to O12 are older adults and Y1 to Y15 are younger adults. Older adults had MoCA scores ranging from 20 to 29 and younger adults had MoCA scores ranging from 26 to 30.

(main task across eight blocks and a swiping task), 4) took part in a short interview, and 5) filled a survey. We counterbalanced task order across steps 2 and 3: even-numbered participants completed MoCA first; odd-numbered participants completed Crossroads first. All other steps remained the same for all participants. Both cohorts (older and younger adults) followed identical study protocol. 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). The interview with older adult participants contained additional questions about their current ways of tracking cognitive performance. Each participant session was videotaped for future analysis. The study received approval from the Institutional Review Board (IRB) of the authors' institution.

5.1.4 Data Analysis. We analyzed data from older and younger adults separately. The older adult cohort included a wide range of cognitive performance (MoCA scores 20–29), while the younger adult cohort (age 18–30) received MoCA scores ≥ 26. Due to the difference in the spread of MoCA scores, we report separate analyses for each age group, then present between-cohort comparisons.

Reliability: Similar to prior work [36], 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 within each cohort [3, 30]. 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 tagged 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 [30]: 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.

Validity: Pearson correlation was computed between all measures and scores on Montreal Cognitive Assessment within each age group. We interpreted correlation values using commonly-accepted thresholds [43]: 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.

Slip profiles: Slip profiles created as plots of speed vs. time during a slip. These plots aided both visual inspection and subsequent data analysis to understand slips made by each participant. Kruskal-Wallis test was used to assess how slip features–like initiation time, detection time, etc. (Figure 5b)–differed among participants grouped by Montreal Cognitive Assessment scores.

For between-cohort comparisons, multiple Kruskal–Wallis tests were used to assess differences in slip features among participants grouped by age and MoCA score. Comparisons were made between three groups (younger adults, older adults with MoCA $\geq$ 26, and older adults with MoCA < 26). Dunn's post hoc tests were then applied to identify which specific group differences contributed to any significant overall effects, with Bonferroni correction to account for multiple comparisons.

We analyzed participant interviews using reflective thematic analysis to assess participants' experiences using Crossroads and MoCA [8].

6 Results

Overall, Crossroads provides a rich understanding of slips and associated cognitive performance. Crossroads produces reliable measures that also correlate well with scores from the Montreal Cognitive Assessment. Using Crossroads took a median time of 2 mins 19 sec (min: 1 min 28 sec, max: 6 mins 29 sec). The Montreal Cognitive Assessment took a median time of 11 mins 07 sec (min: 9 mins 05 sec, max: 14 mins 01 sec).

6.1 Do some Crossroads measures demonstrate high test-retest reliability across the blocks? Which ones?

6.1.1 Which continuous measures demonstrate reliability in older adults? When calculated one block at a time, one measure showed good reliability (0.90 > ICC > 0.75): median time taken (0.80) (Table 2).

When the median of two blocks was used, seven measures showed good reliability: minimum speed (0.78), median speed (0.76), median time taken (0.84), mean time taken (0.83), minimum time taken (0.77), number of pauses (0.80), and mean pause duration (0.75) (Table 2).

When the median of three blocks was used, five measures showed excellent reliability: mean time taken (0.96), median time taken (0.91), maximum time taken (0.95), mean speed (0.94), and mean acceleration (0.91). Eight other measures showed good reliability: median speed (0.78), maximum speed (0.78), minimum acceleration (0.89), maximum acceleration (0.85), median acceleration (0.80), number of pauses (0.88), mean pause duration (0.89), and median pause duration (0.81) (Table 2).

	Older adults				Younger adults			
Measure	**1 at a time**	**2 at a time**	**3 at a time**	**Correlation**	**1 at a time**	**2 at a time**	**3 at a time**	**Correlation**
Time								
Minimum time taken	0.67	0.77	0.74	-0.52	0.63	0.76	0.59	-0.22
Median time taken	0.80	0.84	0.91	-0.72	0.86	0.90	0.90	-0.38
Mean time taken	0.66	0.83	0.96	-0.56	0.83	0.88	0.84	-0.42
Maximum time taken	0.40	0.68	0.95	-0.30	0.72	0.75	0.78	-0.56
Speed								
Minimum speed	0.30	0.78	0.61	0.00	0.77	0.85	0.75	0.00
Median speed	0.61	0.76	0.78	-0.16	0.82	0.84	0.80	0.68
Mean speed	0.49	0.71	0.94	-0.05	0.90	0.93	0.92	0.64
Maximum speed	0.00	0.00	0.78	0.06	0.71	0.88	0.95	-0.09
Acceleration								
Minimum acceleration	0.00	0.00	0.89	0.14	0.21	0.48	0.78	-0.19
Median acceleration	0.00	0.02	0.80	0.00	0.33	0.57	0.77	0.44
Mean acceleration	0.00	0.00	0.91	0.18	0.40	0.59	0.77	-0.19
Maximum acceleration	0.00	0.00	0.85	-0.26	0.03	0.09	0.68	-0.10
Jerk								
Minimum jerk	0.00	0.00	0.48	-0.21	0.04	0.10	0.76	-0.10
Median jerk	0.42	0.63	0.53	0.29	0.47	0.59	0.68	0.41
Mean jerk	0.00	0.00	0.10	0.19	0.01	0.01	0.00	-0.19
Maximum jerk	0.00	0.00	0.23	0.04	0.01	0.02	0.69	-0.10
Pause								
Number of pauses	0.57	0.80	0.88	-0.72	0.74	0.78	0.72	-0.33
Mean pause duration	0.52	0.75	0.89	-0.32	0.39	0.66	0.73	-0.11
Median pause duration	0.42	0.72	0.81	-0.54	0.44	0.70	0.60	-0.32
Maximum pause duration	0.30	0.60	0.81	-0.10	0.35	0.64	0.73	-0.33
Error								
Number of errors	0.31	0.39	0.32	-0.32	0.85	0.91	0.92	-0.02
Number of slips	0.33	0.65	0.42	0.02	0.85	0.91	0.92	0.00
Number of mistakes	0.19	0.23	0.40	-0.49	0.15	0.32	0.11	-0.33

Table 2. Reliability and correlation scores for all measures. ICC scores were calculated for three different groups - "1 at a time", "2 at a time", and "3 at a time". For ICC values, a light green background represents good reliability and a dark green background represents excellent reliability. For correlation values, dark green backgrounds represent values that have p value greater than 0.05.

6.1.2 *Do some measures demonstrate reliability in both older adults and younger adults?*

When calculated one block at a time, median time taken showed good reliability in both older and younger adults (Table 2).

When the median of two blocks was used, five measures showed good reliability in both older and younger adults: median time taken, minimum speed, median speed, mean time taken, and number of pauses (Table 2).

When the median of three blocks was used, nine measures showed good-to-excellent reliability in both older and younger adults: mean speed, maximum speed, mean time taken, median time taken, maximum time taken, median speed, mean acceleration, median acceleration, and minimum acceleration (Table 2).

6.2 Do some measures demonstrate a high correlation with scores on Montreal Cognitive Assessment? Which ones?

Two measures are strongly correlated with MoCA scores (between 0.7 and 0.89) in older adults: median time taken (r=-0.72, p=0.008) and number of pauses (r=-0.72, p=0.008) (Figure 7). Four other measures are moderately correlated with MoCA scores (between 0.4 and 0.69) in older adults. These measures are mean time taken (r=-0.56, p=0.05), minimum time taken (r=-0.52, p=0.07), median pause duration (r=-0.54, p=0.06), and number of mistakes (r=-0.49, p=0.10) (Table 2).

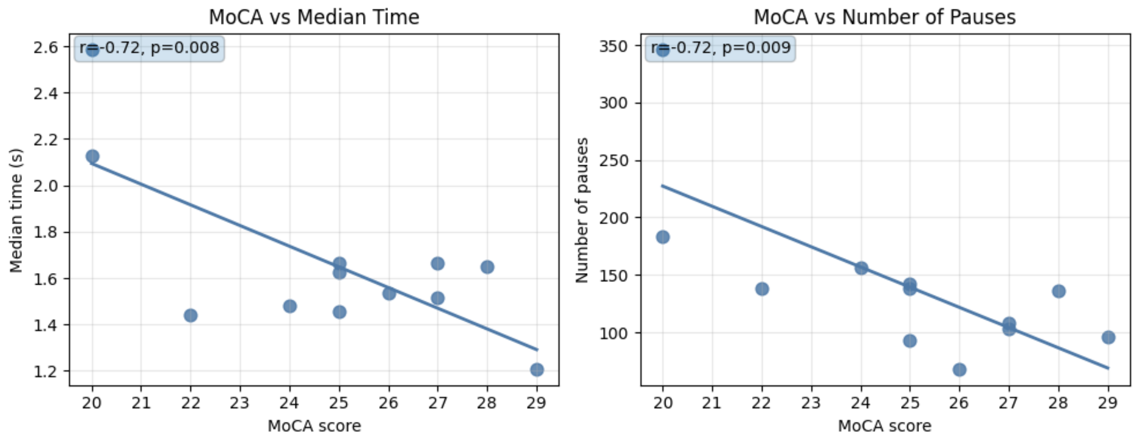


Fig. 7. Median time taken and number of pauses shows a strong negative correlation to scores on Montreal Cognitive Assessment

6.3 What insights do episodic measures provide?

6.3.1 How do slip profiles differ across MoCA scores in older adults? A Kruskal-Wallis test indicated that there was significant difference in slip related features across older adults with different MoCA scores: initiation time (χ^2(6, N=10) = 17.12, p = 0.008). No slip-related measure shows significant correlation with MoCA scores.

6.3.2 How do slips and mistakes differ between older and younger adults? Participants were grouped by age and MoCA scores to form three groups: older adults with MoCA score < 26, older adults with MoCA score ≥ 26, and younger adults with MoCA score ≥ 26 (MoCA score = 26 is the standard threshold to differentiate between normal and mild cognitive impairment).

Older adults make fewer errors (median = 7 and 6 for MoCA score < 26 and ≥ 26) than younger adults (median = 11). A Kruskal–Wallis test indicated there was a significant difference in slip related features across the three groups: slip initiation time (χ^2(3, N=23) = 38.46, p < 0.001), slip detection time (χ^2(3, N=23) = 20.18, p < 0.001), slip correction time (χ^2(3, N=23) = 58.45, p < 0.001), slip duration (χ^2(3, N=23) = 103.23, p < 0.001), confidence gain (χ^2(3, N=23) = 29.29, p < 0.001), and median speed during slip (χ^2(3, N=23) = 16.92, p < 0.001).

Post-hoc comparison using Dunn's method with a Bonferroni correction for multiple tests indicated that younger adults and older adults with MoCA ≥ 26 differ due to all six features (Table 3); younger adults and older adults with MoCA < 26 differ due to five features; and older adults with MoCA <26 and older adults with MoCA >=26 do not differ significantly due to any feature. Figure 8 plots how slips differ across the three groups.

	Younger vs Older (MoCA ≥ 26)			Younger vs Older (MoCA < 26)			Older (MoCA < 26 vs ≥ 26)		
Feature	z	p	p_{adj}	z	p	p_{adj}	z	p	p_{adj}
Initiation time	4.60	<.001	<.001	4.65	<.001	<.001	0.19	.847	1.00
Detection time	3.21	.001	.004	3.48	<.001	.001	0.31	.754	1.00
Correction time	5.02	<.001	<.001	6.30	<.001	<.001	1.15	.249	.748
Slip duration	7.56	<.001	<.001	7.59	<.001	<.001	0.28	.779	1.00
Confidence gain	3.90	<.001	<.001	4.16	<.001	<.001	0.32	0.741	1.00
Median speed during slip	-3.76	<.001	<.001	-2.09	.036	.109	1.15	.250	.751

Table 3. Dunn's post hoc tests (Bonferroni-corrected) for pairwise group comparisons. All older adults and younger differ due to time-based features while older adults (MoCA > 26 vs ≥ 26) do not differ due to any features.

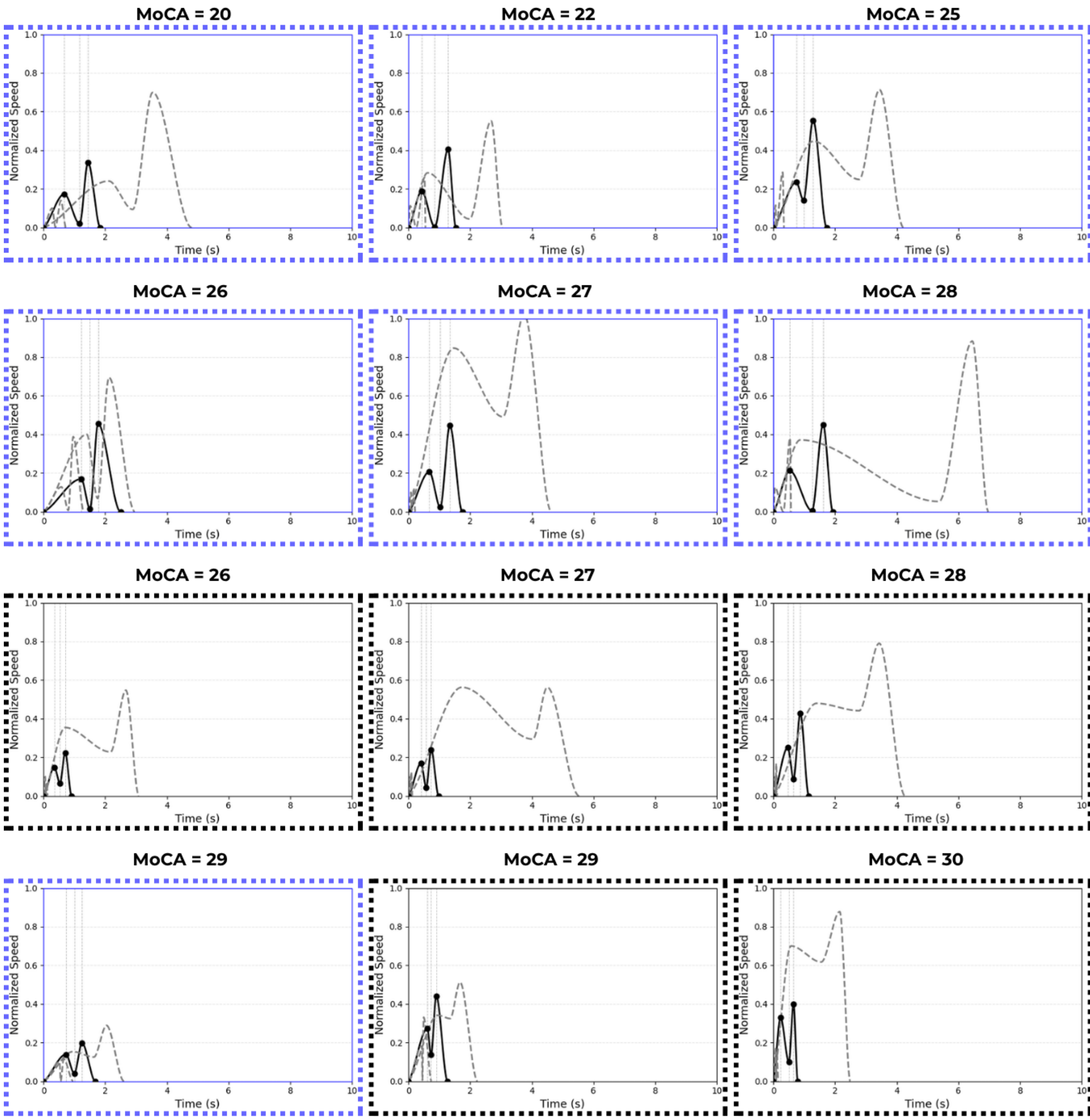


Fig. 8. Slip profiles differ between older and younger adults but not differ between older adults with MoCA <26 and older adults with MoCA ≥ 26. Minimum, maximum, and median slip features were used to define specific anchor points (when initiation, detection, and correction occurred), and cubic Hermite splines were used to interpolate between them, with acceleration controlling the curvature of the resulting profiles. The solid, black line represents the median slip profile and the dotted, gray lines represent the minimum and maximum slip profile. A blue border represents the slip profile for an older adult and a **black** border represents the slip profile for a younger adult. During detection and correction, most older adults (except users with MoCA = 25) reduce their speeds close to zero and all younger adults have speeds greater than zero.

6.4 Acceptability of Crossroads and Montreal Cognitive Assessment

The acceptability of Crossroads and the Montreal Cognitive Assessment are based on interview and video analysis. Older adults took a median of 2 mins 19 sec (min: 1 min 28 sec, max: 6 mins 29 sec) to perform eight blocks of the dragging task on Crossroads. Older adults took a median of 11 mins 07 sec (min: 9 mins 05 sec, max: 14 mins 01 sec) to complete all 12 sub-tasks of the Montreal Cognitive Assessment. One of the sub-tasks in MoCA is a version of the Trail Making Test where participants need to alternate between number and letters from 1 to E. Participants perform the task once (equivalent to one block of Crossroads). Older adults took a median of 25.5 sec (min: 14 sec, max: 42 sec) to perform the Trail Making Test in MoCA. The total time taken moderately correlated with MoCA scores for two measures: the total time taken to complete Crossroads (r=-0.60, p=0.037) and the total time taken to complete MoCA (r=-0.59, p=0.042) (Figure 9).

All participants used Crossroads without any additional instructions or research team's assistance.

> "It [Crossroads] was relatively simple, and the instructions were pretty straightforward." –Y14

> "They [instructions] were pretty clear" –Y8

6.4.1 Views on tracking cognitive performance. Most older adult participants reported getting their cognitive performance tracked by clinicians once a year when they visit their healthcare provider. Some older adult participants reported

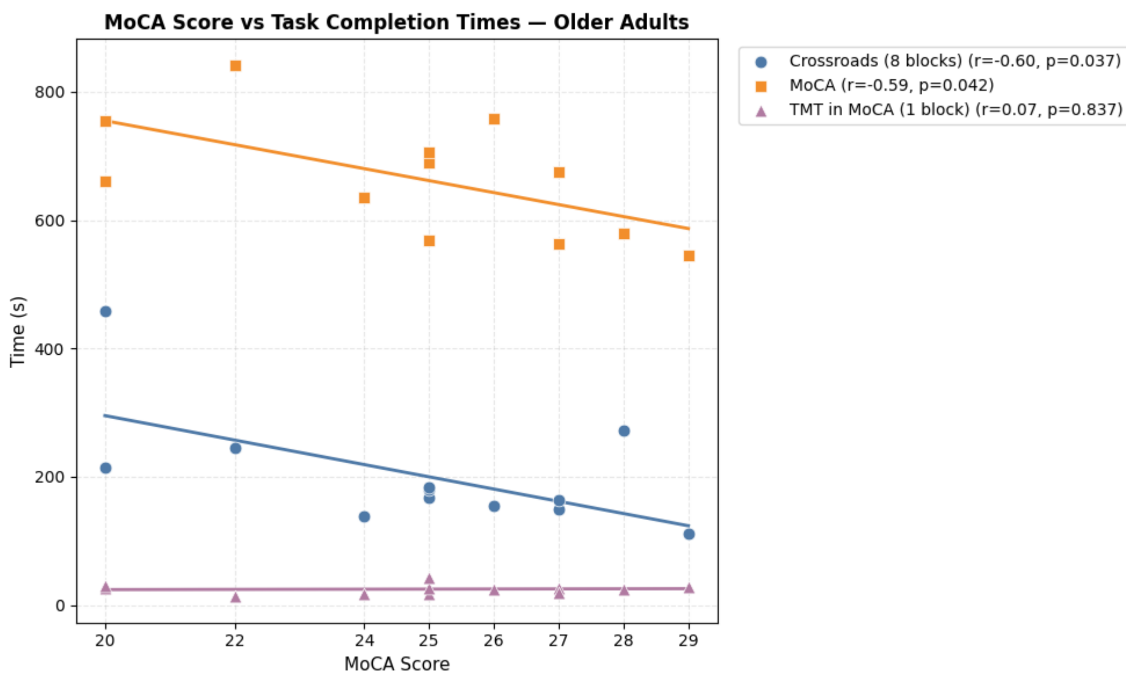


Fig. 9. Time taken to complete Crossroads and MoCA moderately correlate with MoCA scores. Time taken to complete the Trail Making Test in MoCA (one block) shows no correlation with MoCA scores; 8 blocks of TMT in Crossroads shows moderate correlation with MoCA scores.

wanting to track their cognitive performance more often. In particular, they were interested in seeing how the speed and clarity of their thinking changed over time.

Participants shared enthusiasm for cognitive assessments that are quick, easy to complete, and provides both summative scores and descriptive details about their cognitive performance.

> "If there's a quick and easy way to do it [track cognitive performance]" –O10

> "Both [summative and descriptive] because a number can be compared from one to the next [...] So if you can get both you get the best of both worlds." –O12

Some older adults reported a lack of enthusiasm in tracking their cognition frequently since it did not help them identify ways to improve their performance. They reported already being aware of their cognitive decline and said they did not need an assessment to confirm their decline. Instead they wanted tools that told them specifically what needs to be done to improve their cognitive performance. One participant mentioned that they would like to track their cognitive performance only after they notice decline.

> "We're getting older, we know that we're going to lose things [cognitive performance]. I just don't want it [cognitive assessment] in my face every day [...] if it also came with some things to try [...] but not you got a B-minus today. You're getting dementia. No, I didn't want to track losses unless there's some-oops, you had a little error. Here's something you could try" –O5

> "Because what's the point? I'm 84 years old, and I don't know of anybody who can really help me." –O3

6.4.2 Minor issues. Participants reported facing three minor issues while using Crossroads. First, some older adults participants kept lifting their finger off the screen due to subtle tremors. Second, some participants reported feeling anxious due to self-expectation to do well in the task. Third, the movement of dots on the screen after each trial caused initial confusion for some participants. However, participants reported that the task was fun to do after they got used to it. We account for the initial confusion by dropping the first block during data analysis.

> "It takes a minute to figure it [Crossroads' task design] out [...] It was challenging [...] if you're having problems [if someone had a cognitive impairment] that would be really frustrating [...] but it was a good test." –O4

6.4.3 What strategies do participants use while performing Crossroads? Seven older adult participants reported they prioritized accuracy, two prioritized speed, and three tried to prioritize both.

> "I wanted to accomplish the thing accurately and correctly. Accuracy first. If we got the wrong answer quick, it doesn't matter" –O1

> "My main goals to finish it quickly and to get through it" –O12

> "To move quickly and not make mistakes." –O10

Participants created heuristics by themselves to complete the task. One participant mentioned focusing only on the order of numbers (1, 2, 3, 4, etc.) and not keeping track of letters (A, B, C, D, etc.). Each time they got to a number, they moved to the letter they see on the screen. They did not actively keep track of which letter to move to since only one letter shows up on the screen.

> "I focused on the numbers because I knew that was going to be where I had to go. It just seemed to make sense just to focus on the numbers and then the letter was just there" –O7

Similarly, another participant mentioned not keeping track of the order of both numbers and letters. Instead, they alternated between any number or letter they saw on the screen while repeating “number-letter-number-letter” out loud.

> "Constantly repeating "letter, number, letter, number" was very helpful. The first few times, I was very conscious of what I’ve got to connect next. Towards the end, I realized I was just following whatever I was humming. I was dependent on what I was humming." –Y8

7 Discussion

Our research demonstrates that fine-finger tracking on touchscreen-based tasks yields slip profiles that offers insights into cognitive performance. Slip features–like initiation time and slip duration–show differences between older adults and younger adults. Additionally, continuous measures correlate well with scores on standard cognitive assessment (MoCA) and demonstrate good-to-excellent reliability. The results show that touchscreen-based tasks can offer insights into cognitive performance with a few minutes of use. In this section, we discuss how Crossroads-like efforts can complement traditional assessments by providing rich performance data beyond summative scores. We explore how the design of digital tools can meet varying assessment and research objectives.

7.1 Thoughtful interactive digital tasks can complement traditional assessments by automatically tracking and analyzing on-task errors

Our results show that appropriately designed interactive digital tasks can produce two outcomes that provide useful insights into cognitive performance: 1) process-level representation of cognitive performance and 2) reliable and valid features. Task design can also benefit from accounting for self-developed strategies and user preferences.

7.1.1 Motor primitives can help inform cognitive assessments. Traditional cognitive assessments largely overlook how errors unfold during task performance, including how they are initiated, detected, and corrected. Crossroads provides a representation of cognitive performance (slip profiles) which can be decomposed into highly interpretable features that are computed easily from finger trajectories. Our approach draws parallels to motor assessments that analyze movement primitives such as speed, acceleration, and pauses [36]. Similar to such assessments, Crossroads tracks motor primitives but it additionally embeds such tracking within a cognitively demanding task of choosing which target to move towards (number/letter). Making this decision for each trial provides the cognitive load while the tool tracks movement features via fine-finger movement. This cognitive load provides the necessary context to interpret movement features (like speed and time) as behavioral outcomes of underlying cognitive processes. Prior work has demonstrated that movement features can be used to infer higher-level cognitive processes–like real-time decision-making, change in mind, and change in confidence–through continuous input signals such as cursor or finger trajectories [18, 23]. Our work complements prior work by providing an easy to access and use tool that focuses on tracking errors.

7.1.2 Process-level representation of cognitive performance is useful. To our knowledge, Crossroads is the first approach in HCI literature to use an interactive task design that leverages motor primitives to construct a process-level representation of cognitive performance. Prior work–like digital versions of the Trail Making Test [20], digital Whack-a-Mole [47], and Drunk User Interface [33]–provide summative scores to track cognitive performance. Our results show that process-level features derived during slips–like slip initiation time and slip duration–can vary across participants with different cognitive scores. For example, we observed a significant difference in initiation time across MoCA groups in older adults, suggesting that the time spent moving in an incorrect direction might reflect underlying differences in cognitive processing during the task. Unlike measures–such as total time taken or median speed–that are aggregated over the entire task, process-level representation of cognitive performance–like slip features–isolate some error dynamics, tracking how participants detect and recover from incorrect actions in real time. Prior HCI and experimental psychology work–like Drunk User Interface [33] and Track it to Crack it [18]–have shown that user input patterns can enable inference of cognitive states. Similarly, our findings of slip profiles provide a window into error monitoring processes that are not observable in traditional assessments.

7.1.3 Reliable and valid features help understand how tasks are executed. Some low-level features are both reliable and strongly correlated with MoCA scores for older adults. In particular, measures like median time taken and number of pauses demonstrate high test–retest reliability and strong negative correlations with MoCA scores, indicating that participants with lower cognitive scores take longer and pause more frequently during task execution. These features provide additional detail into how cognitive performance impacts how people perform the same task. Prior work shows similar results: people with early symptoms of cognitive decline use pauses as compensatory mechanisms [39]. An increased number of pauses may reflect several underlying processes, including uncertainty during target selection [31] or intermittent planning [19, 23]. Pauses might also be *used* to avoid errors by slowing down decision-making. Our current analysis does not uniquely distinguish between these competing explanations. However, the structure of the task provides opportunities to potentially disambiguate these interpretations in future work. For instance, pauses that consistently occur near decision points (e.g., before transitions between targets) may reflect deliberation, whereas pauses distributed randomly along trajectories may indicate motor interruptions, uncertainty, or even distraction. Combining pause duration with spatial and trajectory-based features could help separate these possibilities. Participants vary in how they transition between decisions; future versions of cognitive assessment tools can offer a more fine-grained view of such performance.

7.1.4 Views on tracking cognitive performance and self-developed strategies can inform task design. Crossroads' acceptability by participants points to a broader design direction for self-administered cognitive assessment tools: useful assessment need not necessarily include long expert-administered sessions. Participants in our study completed the task in under three minutes without assistance, suggesting that lightweight, instruction-minimal interactions can support independent use while lowering access barriers. Participants' views on cognitive tracking were varied. Some valued quick checks with digital assessments to monitor changes over time; others were more resistant to frequent tracking unless it provided actionable insight. For tool designers, the varied response suggests that acceptability might be tightly coupled to perceived utility: tools must be fast, easy to use, and frame results in ways that support longitudinal reflection or intervention (depending on user preference), rather than solely reporting results (which might point to cognitive decline among older adults).

The design of digital assessments needs to account for participants' own heuristics to perform the task: users do not passively follow task instructions and instead actively optimize their behavior. Strategies such as focusing only on numbers or rhythmically alternating "number–letter" might reduce cognitive load while preserving task completion. Such adaptations may improve usability for a participant but it can also potentially mask certain impairments–an issue that likely extends to traditional assessments as well. Future tools should either account for different participant strategy or explicitly design tasks that are robust to such strategies.

7.2 Changes to digital tools' design can improve their use across different demographics and longitudinal deployment

Our efforts suggest that digital tools can be designed to meet the needs and requirements of different demographics like people with decline in motor performance and older adults with severe cognitive impairments.

7.2.1 Decline in motor performance can affect the quality of data collected. Digital assessment tools for older adults need to account for subtle decline in motor performance that comes with age. Older adults might have reduced overall task speed due to slower finger movements. Without accounting for motor decline, cognitive assessments risk confounding cognitive assessments with motor performance limitations. Calibration-based approaches that establish individual baselines before interpreting behavioral signals have been widely used in mobile computing systems. For example, SpiroCall calibrates raw sensor data against ground-truth measurements to account for user and device variability [25]. Similarly, including motor-only baseline tasks can 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 (such as maximum swiping speed) can be used to normalize cognitive performance metrics.

We faced one instance where a participant (O13) could not complete the Crossroads task because they kept lifting their fingers off the screen due to a subtle tremor which was unobservable to the naked eye. The current version of Crossroads shows an error message even when the finger is lifted briefly. During the interview, the participant also reported that they do not often use touchscreen devices. Since tools like Crossroads build on motor primitives, they might benefit from a quick task to assess subtle tremors and participants' ability to use touchscreens. In-task approaches can help too: digital assessments can benefit from distinguishing between subtle tremors and actual finger lifts by waiting for a specific duration before showing an error.

7.2.2 Older adults with severe impairments can benefit from simpler task design. For older adults with severe cognitive impairment, task complexity becomes a primary barrier to usability. The alternating structure of the Trail Making Test Part B (1-A-2-B-3-C...), which requires switching between letters and numbers, can place significant demands on executive function and working memory for some people [1]. As impairment increases, the complexity may lead to disengagement, frustration, or unusable data. A practical design adjustment is to shift toward simpler task variants–such as Trail Making Test Part A–that remove the alternation requirement and focus on sequential processing (1-2-3-4-5...). The reduced cognitive load can make the task more accessible while still enabling useful measurement through features like pauses, errors, and movement trajectories.

Tasks that are too simple risk suffering from ceiling effects, reducing sensitivity to mild impairment. Such ceiling effects might be a sign of *learning effect*–improvements in performance due to repeated exposure rather than a change in cognition [2]. To minimize such learning effects, 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 [36]. Additionally, learning effects themselves may not be entirely unfavorable when interpreted appropriately. 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 [21].

7.3 Limitations and future work

Crossroads' current limitations fall along three dimensions: 1) generating clinical scores, 2) tracking confounding variables; and 3) study population. First, Crossroads provides multiple measures that are reliable and correlate with standard cognitive scores; however, it does not currently *estimate* standard cognitive assessment scores from these measures. Prior work in motor performance assessments–like Hevelius [36]–estimate clinical scores using measures produced by the tool. The limitation stems from the scope of our study, which was designed as a feasibility evaluation focused on assessing reliability and validity of the measures rather than producing clinical estimates. Estimating clinical scores from tracked measures–as demonstrated by systems like Hevelius [36]–requires substantially larger datasets to support robust model training and validation. E.g., Hevelius used data from 138 participants with motor impairments and over 200,000 normative participants to account for age-related effects. Our work provides the building blocks for future research to develop interpretable regression models that estimate scores for standard cognitive assessments (e.g., MoCA) from touchscreen-based measures.

Second, factors such as time of day, caffeine intake, and quality of sleep can influence participants' performance [14, 42, 50]. These variables were not controlled in our study because our primary goal was to evaluate the feasibility and reliability of the task under typical usage conditions. Future longitudinal studies could explicitly track these contextual factors through self-reports or tracking tools. Another confounding variable is the finger used to perform the task. The research team ensured that all participants use the primary finger of their dominant hand to perform the task. When Crossroads is used at home, without the presence of an administrator, the tool needs to automatically prompt the user to use the primary finger of their dominant hand. Another approach could be to take the help of a family member/caregiver to visually ensure this expectation is met.

Third, our study involved a sample of 27 participants (12 older adults + 15 younger adults). All older adult participants identified as White. All younger adult participants were from a university and most older adults (N=11) had a master's degree or higher. MoCA scores were not evenly distributed, with six participants scoring 28 out of 30 and none scoring 21 or 23 out of 30. We do not believe that such challenges (that are typical in user studies for research prototypes) invalidate our findings. Multiple measures from the tool demonstrate useful correlations and group-level differences, suggesting that the approach is sensitive to variation in cognitive performance. Future studies will likely benefit from larger and more varied samples–in age, education levels, and cognitive scores–through broader deployment or targeted recruitment based on clinical screening.

8 Conclusion

This paper presents Crossroads, a self-administered, web-based tool for fine-finger tracking on touchscreen devices designed to assess cognitive performance in older adults without requiring expert presence. A feasibility study with 12 older adults spanning a range of cognitive performance levels evaluated two types of measures derived from finger trajectories: episodic measures focused on errors and continuous measures computed across the full trajectory. Our results demonstrate the feasibility of our approach across multiple dimensions. Slip profiles provide useful process-level insight into how participants detect and recover from errors. Traditional cognitive assessments do not track such information. Multiple continuous measures show high test-retest reliability, and two measures–median time taken and number of pauses–strongly correlate with scores on a standard cognitive assessment. Participants completed Crossroads in a median of 2 minutes 19 seconds, compared to over 11 minutes for MoCA, and all participants completed the task without additional assistance. These findings suggest that touchscreen-based fine-finger tracking is a promising direction for accessible cognitive performance monitoring: one that reduces reliance on clinical infrastructure, lowers the burden of assessment for older adults, and tracks how people perform. Our long-term objective is to deploy and refine such tools in longitudinal, real-world settings, where frequent and self-administered cognitive monitoring can benefit both older adults and the clinicians who support them.

References

[1] 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.

[2] 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.

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

[4] 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.

[5] 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.

[6] 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.

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

[8] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health* 11, 4 (2019), 589–597.

[9] Clarissa Bush, Jean Kozak, and Tom Elmslie. 1997. Screening for cognitive impairment in the elderly. *Canadian Family Physician* 43 (1997), 1763.

[10] Ștefan Sebastian Busnatu, Adelina-Gabriela Niculescu, Alexandra Bolocan, Octavian Andronic, Anca Mihaela Pantea Stoian, Alexandru Scafa-Udriște, Ana Maria Alexandra Stănescu, Dan Nicolae Păduraru, Mihnea Ioan Nicolescu, Alexandru Mihai Grumezescu, et al. 2022. A review of digital health and biotelemetry: modern approaches towards personalized medicine and remote health assessment. *Journal of Personalized Medicine* 12, 10 (2022), 1656.

[11] 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.

[12] 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.

[13] 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.

[14] Drew Dawson and Kathryn Reid. 1997. Fatigue, alcohol and performance impairment. *Nature* 388, 6639 (1997), 235–235.

[15] Ian J Deary, Janie Corley, Alan J Gow, Sarah E Harris, Lorna M Houlihan, Riccardo E Marioni, Lars Penke, Snorri B Rafnsson, and John M Starr. 2009. Age-associated cognitive decline. *British medical bulletin* 92, 1 (2009), 135–152.

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

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

[18] 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.

[19] Sara D'Ascanio, Fabrizio Piras, Caterina Spada, Clelia Pellicano, and Federica Piras. 2025. Pauses as a Quantitative Measure of Linguistic Planning Challenges in Parkinson's Disease. *Brain Sciences* 15, 11 (2025), 1131.

[20] 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.

[21] 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.

[22] Arthur D Fisk, Sara J Czaja, Wendy A Rogers, Neil Charness, and Joseph Sharit. 2020. *Designing for older adults: Principles and creative human factors approaches.* CRC press.

[23] Jonathan B Freeman, Rick Dale, and Thomas A Farmer. 2011. Hand in motion reveals mind in motion. *Frontiers in psychology* 2 (2011), 59.

[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] R Turner Goins, Kimberly A Williams, Mary W Carter, S Melinda Spencer, and Tatiana Solovieva. 2005. Perceived barriers to health care access among rural older adults: a qualitative study. *The Journal of Rural Health* 21, 3 (2005), 206–213.

[27] Jo Mhairi Hale, Daniel C Schneider, Neil K Mehta, and Mikko Myrskylä. 2020. Cognitive impairment in the US: Lifetime risk, age at onset, and years impaired. *SSM-population health* 11 (2020), 100577.

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

[29] 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.

[30] 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.

[31] Isabel Lacruz, Gregory M Shreve, and Erik Angelone. 2012. Average pause ratio as an indicator of cognitive effort in post-editing: A case study. In *Workshop on post-editing technology and practice*.

[32] Shumin Mai, Jingjing Cai, and Lu Li. 2022. Factors associated with access to healthcare services for older adults with limited activities of daily living. *Frontiers in public health* 10 (2022), 921980.

[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] 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.

[36] 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.

[37] Ronald C Petersen, Oscar Lopez, Melissa J Armstrong, Thomas SD Getchius, Mary Ganguli, David Gloss, Gary S Gronseth, Daniel Marson, Tamara Pringsheim, Gregory S Day, et al. 2018. Practice guideline update summary: Mild cognitive impairment: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology. *Neurology* 90, 3 (2018), 126.

[38] 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.

[39] Aurélie Pistono, Jeremie Pariente, Catherine Bézy, Béatrice Lemesle, Johanne Le Men, and Mélanie Jucla. 2019. What happens when nothing happens? An investigation of pauses as a compensatory mechanism in early Alzheimer's disease. *Neuropsychologia* 124 (2019), 133–143.

[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] Christina Schmidt, Fabienne Collette, Christian Cajochen, and Philippe Peigneux. 2007. A time to think: circadian rhythms in human cognition. *Cognitive neuropsychology* 24, 7 (2007), 755–789.

[43] Patrick Schober, Christa Boer, and Lothar A Schwarte. 2018. Correlation coefficients: appropriate use and interpretation. *Anesthesia & analgesia* 126, 5 (2018), 1763–1768.

[44] Brent J Small, Roger A Dixon, and John J McArdle. 2011. Tracking cognition–health changes from 55 to 95 years of age. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 66, suppl_1 (2011), i153–i161.

[45] 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.

[46] 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.

[47] 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.

[48] David Wechsler. 1955. Wechsler adult intelligence scale–. *Archives of Clinical Neuropsychology* (1955).

[49] 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.

[50] Ann M Williamson and Anne-Marie Feyer. 2000. Moderate sleep deprivation produces impairments in cognitive and motor performance equivalent to legally prescribed levels of alcohol intoxication. *Occupational and environmental medicine* 57, 10 (2000), 649–655.

[51] Jacob O Wobbrock and Julie A Kientz. 2016. Research contributions in human-computer interaction. *interactions* 23, 3 (2016), 38–44.

[52] Henry J Woodford and James George. 2007. Cognitive assessment in the elderly: a review of clinical methods. *QJM: An International Journal of Medicine* 100, 8 (2007), 469–484.

[53] 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.