Neuropsychology

Capturing Learning Curves With the Multiday Boston Remote Assessment of Neurocognitive Health (BRANCH): Feasibility, Reliability, and Validity

Emma L. Weizenbaum, Daniel Soberanes, Stephanie Hsieh, Cassidy P. Molinare, Rachel F. Buckley, Rebecca A. Betensky, Michael J. Properzi, Gad A. Marshall, Dorene M. Rentz, Keith A. Johnson, Reisa A. Sperling, Rebecca E. Amariglio, and Kathryn V. Papp

Online First Publication, November 16, 2023. https://dx.doi.org/10.1037/neu0000933

AMERICAN PSYCHOLOGICAL ASSOCIATION

Neuropsychology

ISSN: 0894-4105

https://doi.org/10.1037/neu0000933

Capturing Learning Curves With the Multiday Boston Remote Assessment of Neurocognitive Health (BRANCH): Feasibility, Reliability, and Validity

Emma L. Weizenbaum[1, 2], Daniel Soberanes[2], Stephanie Hsieh[2], Cassidy P. Molinare[2], Rachel F. Buckley[2, 3], Rebecca A. Betensky[4], Michael J. Properzi[3], Gad A. Marshall[2], Dorene M. Rentz[2, 3], Keith A. Johnson[3, 5], Reisa A. Sperling[2, 3], Rebecca E. Amariglio[2, 3], and Kathryn V. Papp[2, 3]

[1] Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School
[2] Department of Neurology, Brigham and Women's Hospital, Harvard Medical School
[3] Department of Neurology, Massachusetts General Hospital, Harvard Medical School
[4] Department of Biostatistics, School of Global Public Health, New York University
[5] Department of Radiology, Massachusetts General Hospital, Harvard Medical School

Objective: Unsupervised remote digital cognitive assessment makes frequent testing feasible and allows for measurement of learning over repeated evaluations on participants' own devices. This provides the opportunity to derive individual multiday learning curve scores over short intervals. Here, we report feasibility, reliability, and validity, of a 7-day cognitive battery from the Boston Remote Assessment for Neurocognitive Health (Multiday BRANCH), an unsupervised web-based assessment. ***Method:*** Multiday BRANCH was administered remotely to 181 cognitively unimpaired older adults using their own electronic devices. For 7 consecutive days, participants completed three tests with associative memory components (Face–Name, Groceries–Prices, Digit Signs), using the same stimuli, to capture multiday learning curves for each test. We assessed the feasibility of capturing learning curves across the 7 days. Additionally, we examined the reliability and associations of learning curves with demographics, and traditional cognitive and subjective report measures. ***Results:*** Multiday BRANCH was feasible with 96% of participants completing all study assessments; there were no differences dependent on type of device used ($t = 0.71$, $p = .48$) or time of day completed ($t = -0.08$, $p = .94$). Psychometric properties of the learning curves were sound including good test–retest reliability of individuals' curves (intraclass correlation = 0.94). Learning curves were positively correlated with in-person cognitive tests and subjective report of cognitive complaints. ***Conclusions:*** Multiday BRANCH is a feasible, reliable, and valid cognitive measure that may be useful for identifying subtle changes in learning and memory processes in older adults. In the future, we will determine whether Multiday BRANCH is predictive of the presence of preclinical Alzheimer's disease.

Key Points

Question: Is digital assessment of learning over 7 days feasible for older adults to complete in a remote, unsupervised context? and Is this multiday learning measure reliable and valid? ***Findings:*** Our sample of clinically unimpaired older adults found a 7-day unsupervised remote assessment to be feasible and enjoyable, and a multiday learning curve score was reliable across time and demonstrated good validity through its association with traditional cognitive and clinical measures. ***Importance:*** The feasibility, reliability, and validity of a 7-day remote assessment of memory signals the opportunity to measure cognition in a high-resolution manner that may be able to differentiate individuals with progressive disease at an earlier stage. ***Next Steps:*** Next, we will determine whether differences in multiday learning curves can be seen in those with risk factors for Alzheimer's disease at the preclinical stage.

Keywords: digital biomarkers, mobile testing, preclinical Alzheimer's disease

Supplemental materials: https://doi.org/10.1037/neu0000933.supp

Emma L. Weizenbaum https://orcid.org/0000-0002-3353-1428
Kathryn V. Papp https://orcid.org/0000-0003-1687-4370

Funding for this study was provided by the National Institute of Aging, National Institutes of Health Grants 2P01AG036694-11 (Reisa A. Sperling, Keith A. Johnson), R01AG053184 (Gad A. Marshall), and 1R01AG058825-01A (Rebecca E. Amariglio) and by the Davis Alzheimer Prevention Program and the Vettel Alzheimer Innovation Fund.

The authors thank all participants for their time and effort, as well as everyone involved in data collection.

Emma L. Weizenbaum played a lead role in writing–original draft, a supporting role in formal analysis and methodology, and an equal role in data curation. Daniel Soberanes played a lead role in project administration and visualization and an equal role in data curation and formal analysis. Stephanie Hsieh played a supporting role in data curation, formal analysis, project administration, and writing–review and editing. Cassidy

continued

Feasible, valid, and reliable tools capable of identifying adults at high risk for cognitive decline due to Alzheimer's disease (AD) remain elusive and pose a barrier to timely intervention (Harvey et al., 2017). The process of learning repeated information over multiple exposures appears to be uniquely impacted in early AD signaling a potentially sensitive cognitive marker at the preclinical stage (Samaroo et al., 2021). In cognitively unimpaired adults, this process can easily be taken for granted as one passively learns which grocery store in the area has the best prices for produce or the name of a new neighbor after several brief encounters. However, in AD, these types of associative learning processes become increasingly challenging (Bastin et al., 2014; El Haj & Antoine, 2018; Kormas et al., 2020). Reduced learning can be localized to the disruptions within the medial temporal lobe, which is one of the first brain regions affected by AD pathology (Braak & Braak, 1991; Jack et al., 2019). Thus, capturing learning, across several time points, may be a promising tool in the development of novel and, ideally, highly sensitive, cognitive assessments in preclinical AD.

A body of existing work has shown that decrements in learning, demonstrated by reduced practice effects after exposure to repeated in-clinic cognitive assessments, occur among cognitively impaired older adults in contrast to those who are unimpaired (Calamia et al., 2012; Hassenstab et al., 2015; Jutten et al., 2020; Papp et al., 2015). Additionally, diminished practice effects among cognitively unimpaired individuals can predict risk for clinical progression to mild cognitive impairment (MCI)/dementia (Duff et al., 2011, 2014; Oltra-Cucarella et al., 2022). While these studies provide evidence for the value of using decrements in learning as a risk factor for current or future impairment, instruments designed to capture decrements in learning over short-time intervals in cognitively unimpaired individuals are sparse (Duff et al., 2017; Lim et al., 2021; Schaefer & Duff, 2017).

Efforts to capture diminished learning over short time frames such as over the course of months or even days have become possible with digital devices. Recent work from our group has investigated multimodal associative learning over repeated evaluation within a span of subsequent months using a study-issued iPad (Samaroo et al., 2021). Within a sample of cognitively unimpaired older adults, monthly repeated face–name pairs revealed a diminished learning trajectory in those with high levels of amyloid deposition. Likewise, over the course of days, an experimental language/character learning paradigm differentiated those unimpaired older adults with evidence of AD biomarkers (Baker et al., 2020; Lim et al., 2020, 2021). The implication of the above work holds promise for optimizing a measurement of cognition that reflects the earliest signs of change in preclinical AD over shorter time frames than was previously thought possible.

This preliminary foundation of encouraging work motivates the development of a cognitive paradigm that measures associative learning over days with stimuli that are relevant to everyday life (e.g., meeting new people, shopping) and sensitive to AD-specific cognitive processes. Our group had previously demonstrated the feasibility and validity of a single time point assessment, the Boston Remote Assessment of Neurocognitive Health (BRANCH), a web-based, remote, and self-guided assessment of visual, verbal, and numeric associative memory processes (Papp, Samaroo, et al., 2021). Given the potential sensitivity gleaned from measuring learning across days, the present study seeks to identify the feasibility and psychometrics of administering Multiday BRANCH over 7 days both by examining adherence as well as usability and acceptability with a posttest questionnaire. We also assessed reliability, repeating the paradigm with new stimuli, as a means of determining the psychometric stability of a learning curve measure of cognition. Last, convergent validity was assessed with the understanding that this novel learning paradigm may demonstrate modest associations with traditional neuropsychological measures.

Method

Transparency and Openness

We report the details of our sample below, all data exclusions, all manipulations, and all measures in the study, and we follow Journal Article Reporting Standards (Kazak, 2018). All data, analysis code, and research materials are available upon request. Data were analyzed using R, Version v4.0.3. This study's design and its analysis were not preregistered.

Participants

The sample included 181 cognitively unimpaired, fluent English-speaking participants aged 60–90 from participants co-enrolled in three separate ongoing observational cohorts including 121 participants from the Harvard Aging Brain Study (HABS; 2P01AG036694-11-Sperling, Johnson), 34 from the Instrumental Activities of Daily Living study (IADL; R01AG053184-Marshall), and 26 from the Subjective Cognitive Decline study (SCD; 1R01AG058825-01A-Amariglio). Study procedures were conducted in accordance with human subjects' protections, and the study protocol was approved by the Mass General Brigham Institutional Review Board. All participants underwent informed consent. Participants were recruited for these co-enrolled observational studies through community-posted advertisements, clinic referrals, and through the efforts of the Outreach and Engagement Core of the Massachusetts Alzheimer's

P. Molinare played a supporting role in data curation, methodology, and project administration. Rachel F. Buckley played a supporting role in conceptualization, formal analysis, methodology, and supervision. Rebecca A. Betensky played a supporting role in methodology and supervision and an equal role in writing–review and editing. Michael J. Properzi played a lead role in data curation and a supporting role in methodology. Gad A. Marshall played a supporting role in conceptualization, data curation, investigation, and project administration and an equal role in funding acquisition. Dorene M. Rentz played a supporting role in conceptualization, funding acquisition, resources, and supervision and an equal role in writing–review and editing. Keith A. Johnson played a supporting role in conceptualization, funding acquisition, and resources. Reisa A. Sperling played a supporting role in conceptualization, funding acquisition, resources, and supervision and an equal role in writing–review and editing. Rebecca E. Amariglio played a supporting role in investigation and an equal role in conceptualization, funding acquisition, methodology, supervision, and writing–review and editing. Kathryn V. Papp played a supporting role in formal analysis and an equal role in conceptualization, investigation, methodology, supervision, and writing–review and editing.

Correspondence concerning this article should be addressed to Kathryn V. Papp, Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, United States. Email: kpapp@bwh.harvard.edu

Disease Research Center. Participants in these studies were required to be in general good health or have stable medical conditions with no significant cerebrovascular or psychiatric disorders. Participants needed a study partner who could answer questions pertaining to daily functioning to establish a Clinical Dementia Rating (CDR). They also underwent annual neuropsychological testing, separate and prior to their Multiday BRANCH administration, as part of their co-enrolled observational studies. They were considered in a cognitively unimpaired group of these studies if they had a CDR global score of zero, normal education-adjusted performance on the Mini-Mental State Examination (>25), and the Logical Memory Delayed Recall (≥16 years of education: ≥9; 8–15 years: ≥5). Furthermore, participants at the time of the study were diagnostically classified as cognitively unimpaired/cognitively normal by multidisciplinary clinical consensus (Papp, Samaroo, et al., 2021). No participants with a diagnosis of MCI were recruited or included in this study. Recruitment for the present study consisted of rolling enrollment through providing study information at in-lab visits related to other co-enrolled studies and emails to participants in these existing studies. For inclusion in the present Multiday BRANCH study, participants were required to have access to a device with a stable internet connection of Wi-Fi or cellular service. Participant characteristics are reported in Table 1.

Multiday BRANCH Paradigm

The study paradigm uses a modified version of BRANCH (Boston Remote Assessment Neurocognitive Health), a previously validated web-based program, which met hospital data privacy and security requirements (Papp, Samaroo, et al., 2021). The current assessment uses only one of the tests (i.e., Groceries) from the four original BRANCH tests. Multiday BRANCH is comprised of three tests that are various paired associative memories tests that are repeatable and sensitive to change in performance over 7 days. Together, the three tasks require roughly 10–15 min to complete and are presented in Figure 1 and described below along with the study procedure.

Table 1
Participant Characteristics

Participant characteristic	Overall (n = 181) N (%)
Age	
M (*SD*)	73.9 (8.42)
Mdn [min, max]	75.0 [53.0, 95.0]
Race	
Asian	3 (1.7%)
Black	17 (9.4%)
Native American	1 (0.6%)
White	160 (88.4%)
Sex	
Female	120 (66.3%)
Male	61 (33.7%)
Years of education	
M (*SD*)	16.5 (2.50)
Mdn [min, max]	17.0 [12.0, 20.0]

Note. min = minimum; max = maximum.

FNAME Test

This is a modified version of the Face–Name Associative Memory Exam (FNAME) incorporating face–name pairs (Rentz et al., 2011), a version of the task is also a part of the computerized cognitive composite (C3; Papp, Rentz, et al., 2021). Participants were shown 20 face–name pairs that were selected to roughly reflect the local community and sample demographics: 12/20 were over age 60, the faces were split evenly by gender (10/10), and 30% were of underrepresented racial and ethnic groups. With each face–name pairing participants are asked whether the name “fits” or “doesn’t fit” with the face to ensure adequate attentiveness to the stimuli. Following a delay, participants identify the previously learned faces, presented alongside two distractor faces (face recognition). In the second component of the FNAME test, the target face is presented, and the participant selects the first letter of the name paired with that face (first-letter name recall). Last, the target face is presented with three names (target name, a re-paired same-sex name, and an age and sex-matched foil name), and the participant must select the correct name (face–name memory). The position of the three response options alongside a given face was randomized to appear in a different order each day. The FNAME total outcome was an average of performance accuracy on the first-letter recall and the face–name recognition task, as these components of FNAME best capture the integrity of associative memory (Papp, Rentz, et al., 2021; Sperling et al., 2003).

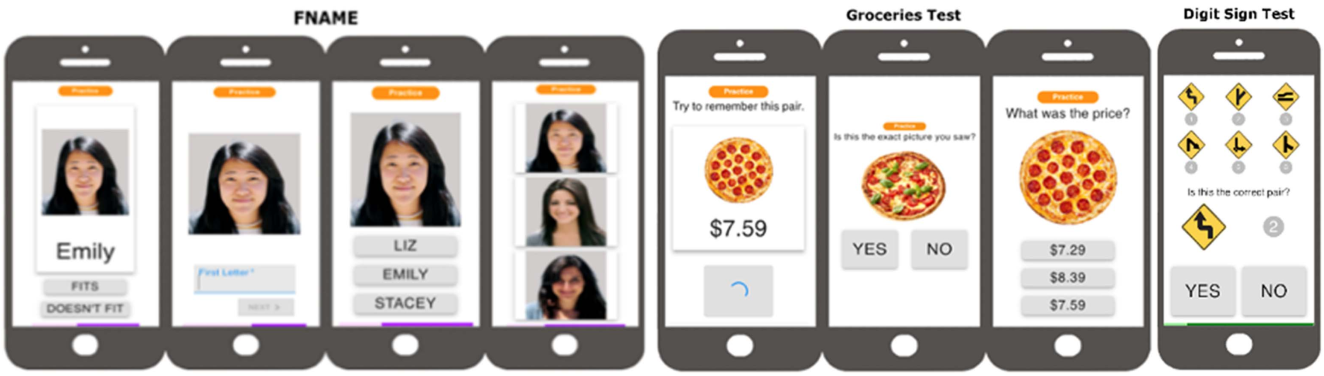
Groceries Test

This is an adapted paired associate learning test combining a visual and numerical element (Castel, 2005). Participants are asked to remember a price (ranging from $1.09 to $12.99) paired with a pictured grocery item. Items were selected to be universally recognized (e.g., bread, bottled water), and prices remain within approximately 15% of their estimated market value. Following a delay, participants complete an adapted pattern separation paradigm whereby they are required to indicate whether a pictured grocery item has been previously seen (equal number of targets, lures, same-category foils); this is the grocery item-recognition score. Just as in the FNAME test, the grocery item-recognition component was used as a task toward filling a delay given the high potential for ceiling effect and confound of increasingly familiar lure responses. The primary component of the Groceries test required participants to recognize the correct price among counterbalanced incorrectly paired and partially novel price/grocery distractor pairs; this is the grocery–price recognition score that was used in analyses.

Digit Signs Test

This measure is modeled on the Digit Symbol Substitution Test (DSST) from the Wechsler Adult Intelligence Scale (Wechsler, 1944). Stimuli include modified street traffic signs, in which participants were asked to indicate (yes–no) whether the street sign and number pairing at the bottom of the screen matched one of six pairings. The outcome score is the total number of correct pairs minus incorrect pairs completed within 90 s divided by the total (154) possible pairs. While the key is always shown and thus the Digit Signs test is not strictly a memory measure, performance is presumed to be more efficient among those who are able to hold the associations “in mind” without having to reference the key.

Figure 1
Three Tests of the Multiday BRANCH Assessment



Note. The Multiday BRANCH assessment includes three remote cognitive measures including a face–name memory test (FNAME), groceries and prices memory test (Groceries), and a speeded pattern separation test determining the accuracy of street sign–number pairs (Digit Sign). BRANCH = Boston Remote Assessment for Neurocognitive Health; FNAME = Face–Name Associative Memory Exam. See the online article for the color version of this figure.

Post-BRANCH Survey and Criteria for Validity

Each day, at the end of the three-test assessment, participants were asked to complete six survey questions regarding their experience while taking the test. First, due to previous experience in hearing the challenge associated with the FNAME test, we were interested in assessing whether participants found this task acceptable and potentially increasingly enjoyable over the course of 7 days. In the postassessment survey, they were asked to rate the enjoyability of the FNAME test on a scale of (1 = *very frustrating* to 10 = *very enjoyable*). From our interest with subjective cognitive appraisal, participants were also asked to rate their self-perceived performance on the FNAME and Groceries tests in a 0–100 point scale; Digit Signs was not included because the rate of completion would be challenging to self-assess. In regard to context, participants were asked where the assessment was completed (e.g., home, work), and whether there were any distractions (notifications, others' talking) or other factors (e.g., pain, nervousness, fatigue) that may have impacted their performance. See Supplemental Materials for the complete survey.

To promote study adherence, participants were told how far along they were within any day's given assessment (e.g., "You are 50% of the way there, keep going!"). At the end of each day's complete assessment, fun brain facts were provided alongside an indication of the percentage of the 7-day study that had been completed, thus far, by the participant.

Multiday BRANCH Procedure

Multiday BRANCH was completed remotely and unsupervised, although the study coordinators provided support for participants via phone and email as needed. The assessment was completed once per day, on participants' own devices—either a personal computer or mobile device such as a tablet or smartphone. Participants were compensated for each attempted assessment session. As shown in Figure 2, the assessment notifications were texted to participants at a predesignated time chosen by the participant to be convenient for the participants' schedule. If they did not respond to the initial notification within 2 hr, they were sent one additional text reminder to complete that day's assessment. On Study Days 1–3, participants were administered an abbreviated practice version of each of the assessment tasks to increase understanding and comfort with the measure. Participants were encouraged to freely reach out to study coordinators by phone, email, or video call for support if technical difficulties arose.

Before beginning the test each day, participants attested to completing tasks independently without recording stimuli or responses with the goal of advancing research. During practice trials, participants heard a chime when responses were correct and an alternate sound when responses were incorrect. No accuracy feedback was given outside of practice trials. During the tests, if participants did not respond in the given time frame, as seen with a timer bar at the bottom of the screen, a pop-up message would arise "Are you ready to proceed?" with a "Next" button that was necessary to press before the test would advance.

The order of tasks within the assessment remained consistent from day to day. The assessment always began with FNAME practice and learning trials, followed by the Groceries practice and learning trials, and then the Digit Sign test. After the Digit Sign test, participants were tested on the FNAME stimuli from the learning trials, roughly 5 min earlier, trials using face recognition, first letter of the name free recall, and face–name matching subtasks. Last, participants were tested on the Groceries test stimuli that had been shown in earlier learning trials and were tested on grocery item recognition followed by grocery prices matching subtasks. The assessment ended with the postassessment survey questionnaire.

Traditional Cognitive Testing and Subjective Report Measures

Participants completed the Preclinical Alzheimer's Cognitive Composite−5 (PACC-5) in the clinic as part of the observational study from which they were recruited (Donohue et al., 2014; Papp et al., 2017). The PACC-5 score used in this study includes two memory measures—the Wechsler Memory Scale-Revised Logical Memory Delayed Recall (Wechsler, 1987) and the Free and Cued Selective Reminding Task (FCSRT; Grober et al., 2009); a measure of global

Figure 2
Procedure for the Multiday BRANCH Assessment

Note. Participants were sent a daily text reminder with a link to the BRANCH web platform where they completed three tests with the same stimuli each day of the 7-day study period. The position of the correct item response shifted to avoid habitual responses. BRANCH = Boston Remote Assessment for Neurocognitive Health. See the online article for the color version of this figure.

cognition, Mini-Mental State Examination (Folstein et al., 1975); a measure of processing speed, DSST (Wechsler, 1981); and a language measure, category fluency that involved totaling the number of correct words produced in three 1-min category trials of animals, fruits, and vegetables. Given that the FCSRT is a multimodal (visual–verbal) associative memory test like the BRANCH FNAME and Groceries tests, we selected this measure from the PACC-5 to examine in association with the Multiday BRANCH subtests. We also were interested in how the PACC-5 DSST independently related to the Multiday BRANCH Digit Signs test, as it is an analog version of the speeded sign–numeric matching test. The Cognitive Function Index (CFI), used to assess subjective report of cognitive change in everyday life, was completed by participants and a study partner (e.g., spouse, family member; Li et al., 2017). The mean time difference between the traditional cognitive testing and self-report measures and the start of Multiday BRANCH was 4.0 months (SD = 8.0).

Data Quality Management

To ensure the completeness and usability of Multiday BRANCH data, we required that participants complete >90% of each day's test items for those scores to be included in analyses. If the 90% criterion was not met, participants were prompted the following day, at their scheduled prompting, to repeat the previous day's test session. So as not to remove any meaningful variance in learning performance over the study period, all data points were included in the Statistical Analyses section.

Statistical Analyses

Paradigm Feasibility and Acceptability

The feasibility of Multiday BRANCH was operationalized as acceptability of frequent testing as shown by >90% of the sample completing all 7 days of testing and a posttest rating indicating greater test enjoyability compared to frustration (>5/10). Participants were asked to complete the 7 study days consecutively, but days of participation were included if they fell within a 14-day period. We also collected data on the type of device used, the average duration of test completion, the time and location of test-taking, and any technical issues experienced.

Adherence. The degree to which participants adhered to the study protocol was measured by determining the percentage of the sample who completed all 7 days of the assessment and the number of days required to achieve 7 days of data. Additionally, we examined whether adherence rates were related to participant characteristics and performance on the assessment itself using *t* tests for sex and Pearson's correlations for age, education, and the composite learning curve score of Multiday BRANCH measures.

Enjoyability. The level and daily change in enjoyability of the FNAME test was found by identifying the mean postassessment rating (1 = *very frustrating* to 10 = *very enjoyable*) for participants by day. A linear mixed-effects model was run to determine whether there was a significant effect of time on enjoyability ratings.

Factors Related to Feasibility. We measured the time individuals spent completing the total assessment at each time point across participants and days and found the mean completion time and standard deviation. We also identified the percentage of assessments completed on a mobile device versus personal computer and the percentage of participants who used the same device throughout the study period, as requested in the study procedures. Time of test-taking was examined by identifying the percentage of tests taken in the morning (5:00 a.m.–11:59 a.m.), afternoon (12 p.m.–5:59 p.m.), evening (6 p.m.–11:59 p.m.), and late night (12:00 a.m.–4:59 a.m.) collapsed across participants and time points. Participants were asked to take the test at the same general time of day, and within-person consistency was examined by measuring the percentage of tests taken in the morning, afternoon, and evening across the 7 days of data. A similar approach was taken to examine the location in which participants took the assessments. The location of tests was examined, by identifying the percentage of tests taken: at home, work, someone else's home, a public place indoors, a public place outdoors, transportation, or elsewhere. Within-person consistency was examined by measuring the percentage of participants who completed tests in multiple locations across the 7 days. As part of feasibility analyses, external sources of interference such as technical problems (i.e., device not responding to taps, difficulty loading images, website freezing or crashing, difficulty seeing images or text) were counted across participants and assessment time points.

Measurement Construction and Reliability of the Learning Curve Scores. A cannon of work on learning theory has demonstrated that the acquisition of new information over time is made up of multiple parameters such as the intercept, or starting point, rate of acquisition, and point of saturation (Ebbinghaus, 1913; Yelle, 1979). Furthermore, early modeling of learning demonstrated that an acquisition model better fits learning data compared to negative exponential growth curve models (Mazur & Hastie, 1978). Thus, to best capture the accumulation and volume of information learned over study days, we chose to create a multiday learning curve metric from an adjusted area under the curve (AUC) score (see Supplemental Materials for details). Unlike an ordinary least squares slope value, the AUC-based learning curve score accounts for differences in participants' starting point and actual fluctuation of performance across days (Figure 3).

Development of a Composite Learning Curve

The Multiday BRANCH tests are multimodal associative memory tasks. The degree of collinearity between these measures was assessed with a Pearson's correlation, and if not overly correlated ($r > 0.8$; Kim, 2019), then each test may contribute similar but unique information to a more powerful and sensitive composite measure of multiday learning. Using Pearson's correlations and *t* tests, associations were examined for composite learning curve scores and the correlate variables described above (i.e., participant characteristics, traditional cognitive measures, and clinical measures).

Test–Retest Reliability of Learning Curves Over 5 Weeks

In order to measure test–retest reliability of the multiday learning curve, the initial participants recruited were asked to participate in a second study period 5–6 weeks later with test–retest enrollment capped at 95 participants. Reliability was assessed using an intraclass correlation (ICC) to measure the relation between individuals' learning curve scores when completed 5–6 weeks apart. Participants were shown a separate set of stimuli for the second Multiday BRANCH series to avoid a stimuli-driven practice effect from the first Multiday BRANCH series. The threshold for an acceptably good ICC value was 0.75 or greater (Koo & Li, 2016).

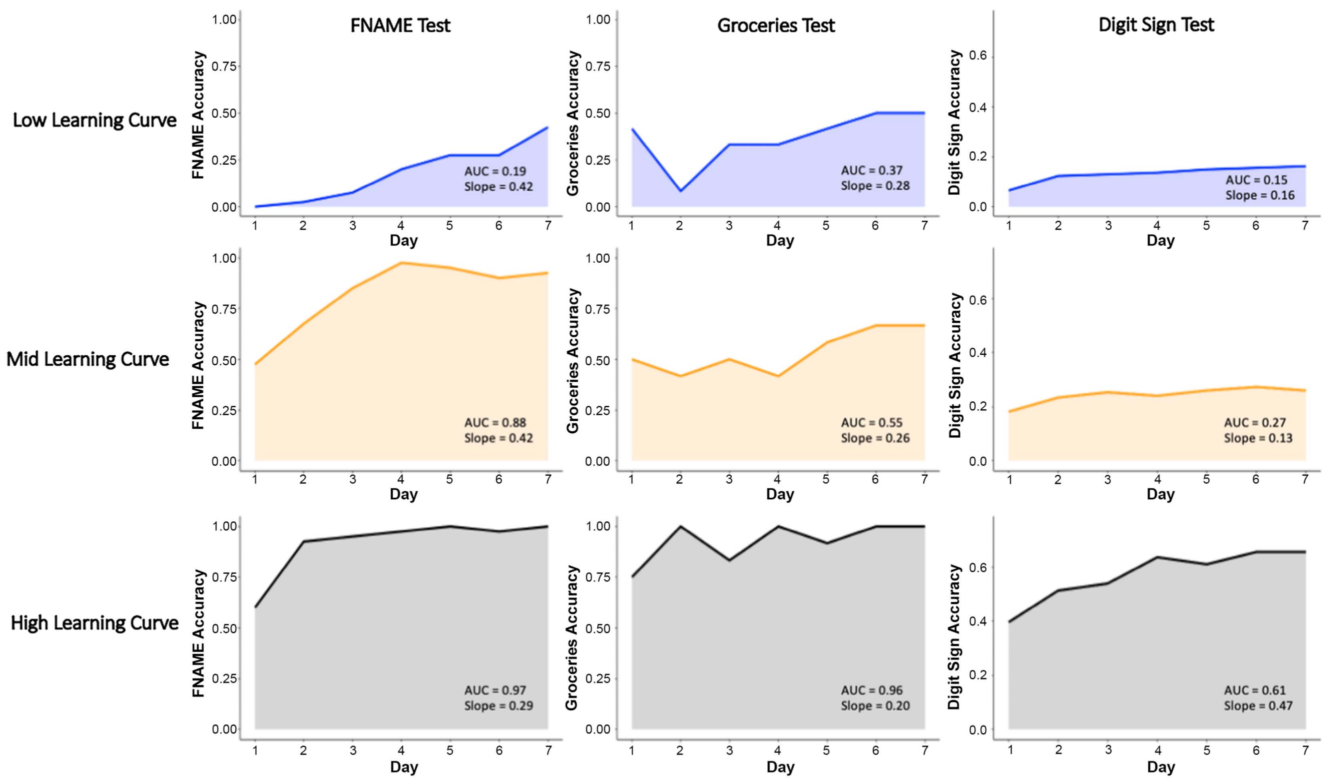
Validity of Multiday BRANCH

Associations With Multiday BRANCH Scores

Pearson's correlations were used to understand whether there was an association in the learning curves across each of the three Multiday BRANCH tests. Pearson's correlations and *t* tests were used to determine the association between learning curve scores and participant age, education, sex, and score differences between study cohorts. Convergent validity was assessed using correlations between learning curves and in-clinic cognitive measures including the PACC-5, which included measures such as the FCSRT, and the DSST. Included in the Supplemental Materials is a table of regression model results accounting for the time between in-clinic measures and Multiday BRANCH. Associations between the learning curves and CFI scores were also assessed. Last, *t* tests and Pearson's correlations were used to determine whether there was a relationship between Multiday BRANCH performance and (a) device type (i.e., predominant use of smartphone vs. personal computer), (b) time of day (e.g., predominant completion in the morning vs. afternoon), or (c) endorsement of environmental distractions (e.g., notifications, noises) or strong internal experiences (e.g., pain, fatigue,

Figure 3

Three Participants' Multiday BRANCH Performance



Note. Three actual participants' learning curves are shown for the FNAME, Groceries, and Digit Sign tests. The modified AUC derived from these performances is listed in contrast to an ordinary least squares slope value derived from their scores. This comparison is shown to demonstrate the poor, and potentially misleading, representation of learning from a linear slope value and motivation for using a modified AUC score. BRANCH = Boston Remote Assessment for Neurocognitive Health; FNAME = Face–Name Associative Memory Exam; AUC = area under the curve. See the online article for the color version of this figure.

nervousness, or other physical or emotional experiences). All statistical analyses were completed using R (v4.0.3).

Results

Paradigm Feasibility and Acceptability

Adherence

One hundred seventy-five participants, or 96.7% of the sample, completed all 7 days of fully remote Multiday BRANCH assessment. The mean number of days to complete the 7 days was 8.01 days (95% CI [7.50, 8.51]). There was no difference in performance on the composite learning curve between participants who completed the full 7 study days consecutively compared to participants who completed the 7 study days in 8–14 days ($n = 42$; $t = 0.64$, $p = .52$, 95% CI [−0.030, 0.058]). Of the six participants who did not complete all 7 days, four participants did not continue past Day 1, one participant completed 4 days, and one participant completed 6 days. There were no significant associations between days of study participation and sex ($t = 0.078$, $p = .94$, 95% CI [−0.26, 0.29]), age ($r = -0.04$, $p = .59$, 95% CI [−0.19, 0.11]), or years of education ($r = -0.01$, $p = .92$, 95% CI [−0.15, 0.14]).

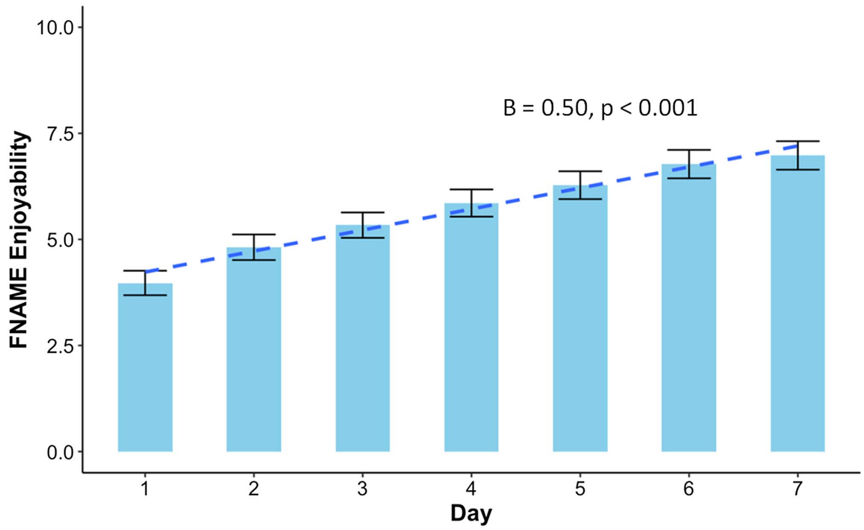
Enjoyability

Figure 4 shows the mean FNAME enjoyability across days was 5.76/10 (95% CI [5.63, 5.89]). FNAME enjoyability increased over days ($B = 0.50$, CI [0.47, 0.54], $p < .001$). For example, on Day 1, mean enjoyability was the lowest at 3.97 (95% CI [3.68, 4.27]), and ratings increased gradually with the highest mean enjoyability of 7.06 (95% CI [6.72, 7.40]) on Day 7.

Factors Related to Feasibility

The sample was relatively evenly split on which device type was used, with 46% of participants using smartphones or tablets and the remaining participants using a personal computer. Per the study request, the majority of participants (86%) were consistent in using the same device across the 7 days. Although not formally measured, research coordinators reviewed emails and voicemails and noted that only a handful of participants reached out for assistance during the study period. Reasons for contacting staff study primarily related to technical questions, such as how to access the study assessment on their web browser. If participants had not completed an assessment in several days, coordinators would reach out to participants to check in.

Figure 4
Average Self-Reported Enjoyability of the Face–Name Task



Note. Self-reported enjoyment (n = 181) on the FNAME ask over 7 days using a Likert scale from 1 (*not enjoyable*) to 10 (*very enjoyable*); B = 0.50, CI [0.46, 0.53], p < .001. FNAME = Face–Name Associative Memory Exam; CI = confidence interval. See the online article for the color version of this figure.

On average, participants took 12.2 min (SD = 2.3) to complete the three BRANCH tasks. Participants completed the assessment in the morning 39% of the time, 44% in the afternoon, and 17% in the evening. We found that 34.2% of participants typically completed the assessment in the morning (>50% of their study assessments), 44.6% typically completed the assessment in the afternoon, and 21.2% typically completed the assessment in the evening. Participants most often took the test in a familiar and predictable environment—typically at home (88.0%), followed by work (3.8%), and someone else's home (3.3%). The majority of the sample (69.9%) took the assessment in the same location over the course of the 7 days. A quarter of the sample completed the assessment in two locations (24.7%), and a few people (5.4%) took the assessment in three different locations.

Self-reported technical difficulties were endorsed during 23.3% of the assessments across the 7 days, with the most common concern being poor device response to finger taps (10.4%). There was a significantly higher number of technical problems reported in those using a mobile device (t = 2.71, p = .007, 95% CI [0.27, 1.72]). A quarter of participants (24.4%) did not endorse any distractions (e.g., loud noises, device notification), and over half of participants (52%) identified fewer than three distractions over the course of the 7-day study period. Nervousness was reported during 10.4% of assessments, fatigue was reported 8.4% of the time, pain was reported <1% of the time, and other physical or emotional distractions were reported during 7.3% of assessments. The reports of these instances were distributed relatively evenly across the 7 days.

Measurement Construction and Reliability of the Learning Curve Score

The average learning curve score for the FNAME test was 0.79 (95% CI [0.77, 0.82]). The average learning curve score for the Groceries test was 0.68 (95% CI [0.66, 0.70]), and the average learning curve score for the Digit Sign test was an adjusted AUC value of 0.14 (95% CI [0.13, 0.15]).

Test–Retest Reliability Over 5 Weeks

Of the 181 participants, the first 95 participants completed two study periods of Multiday BRANCH, with different stimuli, separated by 7 weeks. There was excellent reliability between participants' two composite learning curves (ICC estimate = 0.935) based on a single-rating, absolute-agreement, two-way mixed-effects model. The FNAME learning curve reliability was also excellent (ICC estimate = 0.914). The Groceries test learning curve showed good reliability (ICC estimate = 0.776), as did the Digit Sign test (ICC estimate = 0.828).

Validity of Multiday BRANCH

Associations Between the Three Multiday BRANCH Tests

The learning curve scores of the three individual tests, shown in Table 2, were all moderately correlated with each other. The strongest correlation was as expected between the two multimodal associative memory measures—FNAME and Groceries (r = 0.629, p < .001, 95% CI [0.531, 0.710]).

Multiday BRANCH Associations With Participant Characteristics

Greater age was significantly correlated with lower learning curve scores for all three tests. Men, compared to women, had lower learning curve scores on the FNAME and Groceries tests and composite measures, but there were no significant differences on

Table 2

Associations Between Learning Curves of the BRANCH Tests

Measure	Groceries	Digit Sign
FNAME	$r = 0.629^*$ $p \leq .001$ 95% CI [0.531, 0.710]	$r = 0.330^*$ $p \leq .001$ 95% CI [0.192, 0.456]
Digit Sign	$r = 0.434^*$ $p \leq .001$ 95% CI [0.306, 0.547]	

Note. BRANCH = Boston Remote Assessment for Neurocognitive Health; FNAME = Face–Name Associative Memory Exam; CI = confidence interval.
$^* p < .01$.

Digit Sign learning curve scores. Education level was not significantly associated with any of the three Multi-BRANCH tests. There were also no significant differences in learning curve scores between the study cohorts (i.e., HABS, IADL, and SCD; see Table 3).

Multiday BRANCH Associations With Traditional Cognitive and Subjective Report Measures

Participants' most recent annual score on the PACC-5 cognitive composite measure was significantly associated with all Multiday BRANCH test scores. There was no significant effect of time between traditional cognitive tests and Multiday BRANCH (see Supplemental Materials for table of model results when the time was included as a covariate). The in-person FCSRT total score from the PACC-5 was significantly associated with both the multimodal associative measures in Multiday BRANCH, the FNAME and Groceries learning curve score. The in-person DSST from the PACC-5 was also significantly associated with the similar Digit Sign test in Multiday BRANCH (see Table 4).

On the CFI self-report, higher scores, indicating higher levels of SCD, were associated with significantly lower learning curve scores across the three measures. The CFI study partner-report was also negatively related to Multiday BRANCH performance across the three measures. Of note, while statistically significant, correlation coefficients were small in magnitude (see Table 5).

Factors That May Influence Multiday BRANCH Performance

When comparing the mean composite learning curve scores for participants who primarily used a personal computer versus participants who primarily used a mobile device, there was no significant difference between the two groups ($t = -0.71$, $p = .48$, 95% CI [−0.055, 0.026]). The incidence of technical difficulties was not associated with participants' composite learning curves ($t = 0.657$, $p = .51$, 95% CI [−0.029, 0.058]). There was also no significant association between the number of external distractions reported and participants' composite learning curves ($r = 0.125$, $p = .099$, 95% CI [−0.023, 0.267]). Nor were there significant differences in composite learning scores between participants who consistently completed Multiday BRANCH in the morning versus the afternoon or evening ($t = -0.075$, $p = .94$, 95% CI [−0.015, 0.014]).

Discussion

This study demonstrated the feasibility, reliability, and validity of an unsupervised, remote multiday learning curve assessment. In a sample of cognitively unimpaired older adults, almost all participants were able to fully complete all 7 days of testing regardless of participant demographics or scores on cognitive and clinical measures. Study engagement was leveraged by using colorful stimuli, fun brain facts, and daily compensation. Participants were evenly split on their use of a mobile device versus computer for the test and there were no significant differences in multiday learning performance between those who used one device versus the other. Technical difficulties emerged for some participants, specifically poor device response to finger taps, which is consistent with previous studies showing skin conductance and lower touch screen manipulation abilities in older adults (Elboim-Gabyzon et al., 2021). In the future, as we broaden the sample to those from more diverse backgrounds, it

Table 3

Associations Between Learning Curves and Participant Characteristics

Demographics	FNAME	Groceries	Digit Sign	Composite
Age	$r = -0.323^*$ $p < .001$ CI [−0.449, −0.184]	$r = -0.396^*$ $p < .001$ CI [−0.514, −0.264]	$r = -0.443^*$ $p < .001$ CI [−0.553, −0.317]	$r = -0.449^*$ $p < .001$ CI [−0.560, −0.323]
Sex (male)	$t = -4.872^*$ $p < .001$ $M = 0.70$ CI [−0.191, −0.080]	$t = -4.784^*$ $p < .001$ $M = 0.609$ CI [−0.155, −0.064]	$t = -0.306$ $p = .7602$ $M = 0.132$ CI [−0.034, 0.025]	$t = -4.611^*$ $p < .001$ $M = 0.581$ CI [−0.143, −0.057]
Years of education	$r = 0.037$ $p = .620$ CI [−0.111, 0.184]	$r = -0.008$ $p = .916$ CI [−0.155, 0.140]	$r = -0.002$ $p = .978$ CI [−0.148, 0.144]	$r = 0.043$ $p = .572$ CI [−0.105, 0.189]
HABS versus IADL/SCD	$t = -0.429$ $p = .669$ $M = 0.789$ CI [−0.062, 0.040]	$t = -0.934$ $p = .352$ $M = 0.674$ CI [−0.066, 0.024]	$t = -0.676$ $p = .500$ $M = 0.132$ CI [−0.040, 0.019]	$t = -0.509$ $p = .612$ $M = 0.644$ CI [−0.050, 0.029]

Note. CI = 95% confidence interval; FNAME = Face–Name Associative Memory Exam; HABS = Harvard Aging Brain Study; IADL = Instrumental Activities of Daily Living study; SCD = Subjective Cognitive Decline study.
$^* p < .01$.

Table 4

Associations Between Learning Curves and Traditional Cognitive Measures

Measure	FNAME	Groceries	Digit Sign	Composite
PACC-5	$r = 0.509^{*}$ $p < .001$ CI [0.391, 0.611]	$r = 0.521^{*}$ $p < .001$ CI [0.403, 0.621]	$r = 0.415^{*}$ $p < .001$ CI [0.285, 0.529]	$r = 0.610^{*}$ $p < .001$ CI [0.508, 0.696]
FCSRT	$r = 0.278^{*}$ $p < .001$ CI [0.135, 0.409]	$r = 0.191$ $p = .011$ CI [0.044, 0.330]		
DSST			$r = 0.481^{*}$ $p < .001$ CI [0.360, 0.587]	

Note. CI = 95% confidence interval; FNAME = Face–Name Associative Memory Exam; PACC-5 = Preclinical Alzheimer's Cognitive Composite−5; FCSRT = Free and Cued Selective Reminding Task; DSST = Digit Symbol Substitution Test.
$^{*}p < .01$.

will be important to assess the degree of digital literacy of the sample through a measure such as the Mobile Device Proficiency Scale (Roque & Boot, 2018) to identify whether this impacts feasibility through adherence, acceptance, or technical difficulties. Reassuringly, there is accumulating evidence that today's older adults are quite capable of navigating digital platforms at levels similar to those of younger generations (Mace et al., 2022). Nervousness and fatigue were the most common individual contextual experiences reported, and environmental distractions were reported only a small percentage of the time. Of note, Multiday BRANCH performance was not related to the number of individual contextual experiences and distractions experienced over the 7 days indicating that for a multiday learning assessment, a remote environment does not appear to affect the integrity of the measurement.

For the most part, people tended to take the tests at home and in the afternoon. Further, there were no differences in learning curves between those who completed the test predominantly in the morning versus later in the day. This preference toward taking the test later in the day and the lack of time-of-day effect comes in contrast to recent work that showed older adults having poorer memory performance in evening hours on remote testing (Wilks et al., 2021). One hypothesis for the lack of time-of-day effect in our study was that, unlike other digital protocols where assessments are prompted at prespecified or quasi-random times, participants in our study were given a choice of when they would like to complete the assessment. The afternoon and evening test-taking preference may be an effect of high-functioning individuals having other engagements and activities earlier in the day.

Regarding participant burden, the test was relatively short to complete once a day (∼12 min), and the FNAME test was rated as increasingly enjoyable with each successive day. This finding was encouraging given that burden, test fatigue, and lack of enjoyment may all be factors that contribute to participants failing to maintain protocol adherence and participation across study time points (Wei et al., 2020).

In addition to evaluating feasibility, we sought to evaluate the psychometric properties of Multiday BRANCH. Using the test–retest method of assessing learning curves within the same individuals, 5 weeks apart, with different stimuli, we found that participants' learning curves were highly related (ICC = 0.94), signaling that the composite learning curve is a reliable metric of cognitive function. Given the stability of the measure, it has the potential to be repeated over longer intervals. For example, while traditional single time point measures may not show a significant decline over the course of 6 or 12 months, it is possible, given the density of learning data, that individuals' 7-day learning curves may show a reduction in performance over a period of months.

When comparing Multiday BRANCH performance with demographic factors, there was alignment with the existing literature on the expected decline in cognition with age (Boyle et al., 2013; Salthouse, 2009). We found that younger individuals had stronger performance on Multiday BRANCH measures than older participants. Similarly, prior work has shown that on some tests of memory, women tend to perform better than men (Rentz et al., 2017; Sundermann et al., 2017), and the same was seen in our study with higher learning curve scores in female participants, specific to the

Table 5

Associations Between Learning Curves and Subjective Report

Measure	FNAME	Groceries	Digit Sign	Composite
CFI self	$r = -0.180^{*}$ $p = .017$ CI [−0.320, −0.033]	$r = -0.168^{*}$ $p = .026$ CI [−0.309, −0.021]	$r = -0.119$ $p = .114$ CI [−0.262, 0.029]	$r = -0.205^{**}$ $p = .006$ CI [−0.343, −0.059]
CFI study partner	$r = -0.176^{*}$ $p = .026$ CI [−0.322, −0.022]	$r = -0.208^{**}$ $p = .008$ CI [−0.352, −0.055]	$r = -0.169^{*}$ $p = .031$ CI [−0.314, −0.016]	$r = -0.222^{**}$ $p = .005$ CI [−0.364, −0.069]

Note. CI = 95% confidence interval; FNAME = Face–Name Associative Memory Exam; CFI = Cognitive Function Index.
$^{*}p = .05$. $^{**}p < .01$.

memory-dominant subtests (FNAME and Groceries). Importantly, there were no differences in performance based on education, which differs from many existing traditional cognitive assessments (Ardila et al., 2000; Manly et al., 2002), even when considering the limitation of a highly educated sample. In fact, this finding aligns with previous studies of both short- and long-term practice effects that did not find significant associations with education contrary to expectations (Duff et al., 2012; Jutten et al., 2022). This lack of association may reflect the benefit of studying learning over repeated exposures in individuals of different backgrounds as it is a measure of process rather than accuracy at a single time point.

Regarding convergent validity with traditional cognitive assessments, learning curve scores were significantly correlated with several established measures of cognitive function. The PACC-5 is made up of measures particularly sensitive to cognitive change in those at risk for AD (Papp et al., 2022) and was significantly correlated with all three of the Multiday BRANCH tests. Further strengthening the construct validity of Multiday BRANCH, we found that the two associative memory measures, FNAME and Groceries, were significantly related to an associative memory measure from the PACC-5, the FCSRT. Similarly, the cross-modal associative processing speed task from the PACC-5, the DSST, was strongly associated with the related Multiday BRANCH Digit Sign test.

In addition to traditional cognitive assessments, clinical measures of cognition such as the CDR and CFI (self- and study partner-report) were all significantly associated with learning curve scores, but relatively small in the magnitude of correlation. Generally, this association is reassuring in that it reflects some degree of alignment between objective Multiday BRANCH performance and participants' and study partners' perspectives on everyday functioning despite being very different ways of measuring cognitive function. All three individual Multiday BRANCH scores (FNAME, Groceries, and Digit Sign) were significantly correlated with each other and yet had different learning curve patterns indicating they are related but unique contributions to a composite score.

Last, we were interested in contextual factors that may influence Multiday BRANCH scores and found that device type, time of day, total distractions, or an individual's internal experiences influenced the learning curve scores. The role of contextual factors on remote cognitive performance measures has been mixed (von Stumm, 2018; Weizenbaum et al., 2022; Wilks et al., 2021), and to date, these factors have not been assessed in relation to a multiday score of cognition but rather how they relate to cognitive performance contemporaneously. Given the expectation that scores would improve over the 7 days, we were not able to independently assess the role of contextual factors on an individual's score for any given day.

Taken together, this study shed light on the feasibility and psychometrics of Multiday BRANCH. This validation demonstrates promise in its potential as an adjunctive measure that may be easily deployable and useful to determine who is at risk for cognitive decline. However, the study included several limitations, one of which was a selection of bias toward highly motivated older adults who were already co-enrolled in existing longitudinal observational studies and self-selected to take part in a multiday remote digital assessment study. The feasibility and reliability of multiday BRANCH may decrease as we scale the validation in a more heterogeneous sample and can be compared to the known decrement in these psychometrics in the translation of traditional memory assessments from normative to clinical samples (Aldridge et al., 2017). Similarly, the current sample, while reflective of local demographics, was not racially representative of the greater United States. Moving forward, it will be important to assess Multiday BRANCH in a more racially diverse sample, and efforts are currently underway to use Multiday BRANCH in a sample of Latinx older adults. As a means of ensuring acceptability of the method, we are providing a mobile device and/or secure cellular connection for those without one; of note, a 2021 Pew Research Center survey showed similar levels of tablet ownership and smartphone use across White, Black, and Hispanic respondents. Furthermore, it will be important to see whether remote, unsupervised Multiday BRANCH will be feasible in individuals with cognitive impairment, including those with MCI and early AD. Individuals with impairment may be able to successfully complete Multiday BRANCH tests with additional support, such as daily reminders and technological guidance, from research coordinators and/or study partners at home.

The next steps include determining whether differences in the patterns of learning curves on Multiday BRANCH are related to AD risk factors including genotype and biomarkers such as amyloid and tau pathology. Previous work from our group has shown that a reduced learning curve on a 1-day version of FNAME repeated across months was associated with individuals with higher levels of amyloid deposition (Samaroo et al., 2021). The relatively short nature of the Multiday BRANCH testing period may allow one to potentially identify individuals at risk for decline in a shorter period of testing versus longitudinal cognitive testing, which traditionally occurs over the course of many years.

In conclusion, the present study found that Multiday BRANCH is a valid and meaningful measure of cognitive function that explores multiple metrics of learning and memory and can be feasibly completed unsupervised and within participants' own homes. Our current measure utilizes associative memory paradigms that are sensitive to changes in AD and include material (faces, names, grocery items) that are relevant to everyday life. By measuring cognition over the course of 7 days, we can capture the dynamic cognitive process of multiple-exposure learning in a manner that has traditionally been challenging due to the constraints of lab-based testing. With the fundamental psychometrics established, future work will determine the utility of Multiday BRANCH in the detection of preclinical AD.

References

Aldridge, V. K., Dovey, T. M., & Wade, A. (2017). Assessing test–retest reliability of psychological measures. *European Psychologist*, *22*(4), 207–218. https://doi.org/10.1027/1016-9040/a000298

Ardila, A., Ostrosky-Solis, F., Rosselli, M., & Gómez, C. (2000). Age-related cognitive decline during normal aging: The complex effect of education. *Archives of Clinical Neuropsychology*, *15*(6), 495–513. https://doi.org/10.1093/arclin/15.6.495

Baker, J. E., Bruns, L., Jr., Hassenstab, J., Masters, C. L., Maruff, P., & Lim, Y. Y. (2020). Use of an experimental language acquisition paradigm for standardized neuropsychological assessment of learning: A pilot study in young and older adults. *Journal of Clinical and Experimental Neuropsychology*, *42*(1), 55–65. https://doi.org/10.1080/13803395.2019.1665626

Bastin, C., Bahri, M. A., Miévis, F., Lemaire, C., Collette, F., Genon, S., Simon, J., Guillaume, B., Diana, R. A., Yonelinas, A. P., & Salmon, E. (2014). Associative memory and its cerebral correlates in Alzheimer's disease: Evidence for distinct deficits of relational and conjunctive memory. *Neuropsychologia*, *63*, 99–106. https://doi.org/10.1016/j.neuropsychologia.2014.08.023

Boyle, P. A., Yu, L., Wilson, R. S., Schneider, J. A., & Bennett, D. A. (2013). Relation of neuropathology with cognitive decline among older persons without dementia. *Frontiers in Aging Neuroscience*, *5*, Article 50. https://doi.org/10.3389/fnagi.2013.00050

Braak, H., & Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathologica*, *82*(4), 239–259. https://doi.org/10.1007/BF00308809

Calamia, M., Markon, K., & Tranel, D. (2012). Scoring higher the second time around: Meta-analyses of practice effects in neuropsychological assessment. *The Clinical Neuropsychologist*, *26*(4), 543–570. https://doi.org/10.1080/13854046.2012.680913

Castel, A. D. (2005). Memory for grocery prices in younger and older adults: The role of schematic support. *Psychology and Aging*, *20*(4), 718–721. https://doi.org/10.1037/0882-7974.20.4.718

Donohue, M. C., Sperling, R. A., Salmon, D. P., Rentz, D. M., Raman, R., Thomas, R. G., Weiner, M., Aisen, P. S., the Australian Imaging, Biomarkers, and Lifestyle Flagship Study of Ageing, the Alzheimer's Disease Neuroimaging Initiative, & the Alzheimer's Disease Cooperative Study. (2014). The preclinical Alzheimer cognitive composite: Measuring amyloid-related decline. *JAMA Neurology*, *71*(8), 961–970. https://doi.org/10.1001/jamaneurol.2014.803

Duff, K., Atkinson, T. J., Suhrie, K. R., Dalley, B. C., Schaefer, S. Y., & Hammers, D. B. (2017). Short-term practice effects in mild cognitive impairment: Evaluating different methods of change. *Journal of Clinical and Experimental Neuropsychology*, *39*(4), 396–407. https://doi.org/10.1080/13803395.2016.1230596

Duff, K., Callister, C., Dennett, K., & Tometich, D. (2012). Practice effects: A unique cognitive variable. *The Clinical Neuropsychologist*, *26*(7), 1117–1127. https://doi.org/10.1080/13854046.2012.722685

Duff, K., Foster, N. L., & Hoffman, J. M. (2014). Practice effects and amyloid deposition: Preliminary data on a method for enriching samples in clinical trials. *Alzheimer Disease and Associated Disorders*, *28*(3), 247–252. https://doi.org/10.1097/WAD.0000000000000021

Duff, K., Lyketsos, C. G., Beglinger, L. J., Chelune, G., Moser, D. J., Arndt, S., Schultz, S. K., Paulsen, J. S., Petersen, R. C., & McCaffrey, R. J. (2011). Practice effects predict cognitive outcome in amnestic mild cognitive impairment. *The American Journal of Geriatric Psychiatry: Official Journal of the American Association for Geriatric Psychiatry*, *19*(11), 932–939. https://doi.org/10.1097/JGP.0b013e318209dd3a

Ebbinghaus, H. (1913). Our knowledge concerning memory. In H. A. Ruger, C. E. Bussenius, H. A. Ruger, & C. E. Bussenius (Trans.), *Memory: A contribution to experimental psychology* (pp. 1–6). Teachers College Press. https://doi.org/10.1037/10011-001

El Haj, M., & Antoine, P. (2018). Context memory in Alzheimer's disease: The "who, where, and when". *Archives of Clinical Neuropsychology*, *33*(2), 158–167. https://doi.org/10.1093/arclin/acx062

Elboim-Gabyzon, M., Weiss, P. L., & Danial-Saad, A. (2021). Effect of age on the touchscreen manipulation ability of community-dwelling adults. *International Journal of Environmental Research and Public Health*, *18*(4), Article 2094. https://doi.org/10.3390/ijerph18042094

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Grober, E., Ocepek-Welikson, K., & Teresi, J. A. (2009). The free and cued selective reminding test: Evidence of psychometric adequacy. *Psychology Science Quarterly*, *51*(3), 266–281.

Harvey, P. D., Cosentino, S., Curiel, R., Goldberg, T. E., Kaye, J., Loewenstein, D., Marson, D., Salmon, D., Wesnes, K., & Posner, H. (2017). Performance-based and Observational Assessments in Clinical Trials Across the Alzheimer's Disease Spectrum. *Innovations in Clinical Neuroscience*, *14*(1–2), 30–39.

Hassenstab, J., Ruvolo, D., Jasielec, M., Xiong, C., Grant, E., & Morris, J. C. (2015). Absence of practice effects in preclinical Alzheimer's disease. *Neuropsychology*, *29*(6), 940–948. https://doi.org/10.1037/neu0000208

Jack, C. R., Jr., Wiste, H. J., Therneau, T. M., Weigand, S. D., Knopman, D. S., Mielke, M. M., Lowe, V. J., Vemuri, P., Machulda, M. M., Schwarz, C. G., Gunter, J. L., Senjem, M. L., Graff-Radford, J., Jones, D. T., Roberts, R. O., Rocca, W. A., & Petersen, R. C. (2019). Associations of amyloid, tau, and neurodegeneration biomarker profiles with rates of memory decline among individuals without dementia. *JAMA*, *321*(23), 2316–2325. https://doi.org/10.1001/jama.2019.7437

Jutten, R. J., Grandoit, E., Foldi, N. S., Sikkes, S. A. M., Jones, R. N., Choi, S. E., Lamar, M. L., Louden, D. K. N., Rich, J., Tommet, D., Crane, P. K., & Rabin, L. A. (2020). Lower practice effects as a marker of cognitive performance and dementia risk: A literature review. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *12*(1), Article e12055. https://doi.org/10.1002/dad2.12055

Jutten, R. J., Rentz, D. M., Fu, J. F., Mayblyum, D. V., Amariglio, R. E., Buckley, R. F., Properzi, M. J., Maruff, P., Stark, C. E., Yassa, M. A., Johnson, K. A., Sperling, R. A., & Papp, K. V. (2022). Monthly at-home computerized cognitive testing to detect diminished practice effects in preclinical Alzheimer's disease. *Frontiers in Aging Neuroscience*, *13*, Article 800126. https://doi.org/10.3389/fnagi.2021.800126

Kazak, A. E. (2018). Editorial: Journal article reporting standards. *American Psychologist*, *73*(1), 1–2. https://doi.org/10.1037/amp0000263

Kim, J. H. (2019). Multicollinearity and misleading statistical results. *Korean Journal of Anesthesiology*, *72*(6), 558–569. https://doi.org/10.4097/kja.19087

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kormas, C., Zalonis, I., Evdokimidis, I., Kapaki, E., & Potagas, C. (2020). Face-name associative memory performance among cognitively healthy individuals, individuals with subjective memory complaints, and patients with a diagnosis of aMCI. *Frontiers in Psychology*, *11*, Article 2173. https://doi.org/10.3389/fpsyg.2020.02173

Li, C., Neugroschl, J., Luo, X., Zhu, C., Aisen, P., Ferris, S., & Sano, M. (2017). The utility of the cognitive function instrument (CFI) to detect cognitive decline in non-demented older adults. *Journal of Alzheimer's Disease*, *60*(2), 427–437. https://doi.org/10.3233/JAD-161294

Lim, Y. Y., Baker, J. E., Bruns, L., Jr., Mills, A., Fowler, C., Fripp, J., Rainey-Smith, S. R., Ames, D., Masters, C. L., & Maruff, P. (2020). Association of deficits in short-term learning and Aβ and hippocampal volume in cognitively normal adults. *Neurology*, *95*(18), e2577–e2585. https://doi.org/10.1212/WNL.0000000000010728

Lim, Y. Y., Baker, J. E., Mills, A., Bruns, L., Jr., Fowler, C., Fripp, J., Rainey-Smith, S. R., Ames, D., Masters, C. L., & Maruff, P. (2021). Learning deficit in cognitively normal APOE ε4 carriers with LOW β-amyloid. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *13*(1), Article e12136. https://doi.org/10.1002/dad2.12136

Mace, R. A., Mattos, M. K., & Vranceanu, A.-M. (2022). Older adults can use technology: Why healthcare professionals must overcome ageism in digital health. *Translational Behavioral Medicine*, *12*(12), 1102–1105. https://doi.org/10.1093/tbm/ibac070

Manly, J. J., Jacobs, D. M., Touradji, P., Small, S. A., & Stern, Y. (2002). Reading level attenuates differences in neuropsychological test performance between African American and White elders. *Journal of the International Neuropsychological Society*, *8*(3), 341–348. https://doi.org/10.1017/S1355617702813157

Mazur, J. E., & Hastie, R. (1978). Learning as accumulation: A reexamination of the learning curve. *Psychological Bulletin*, *85*(6), 1256–1274. https://doi.org/10.1037/0033-2909.85.6.1256

Oltra-Cucarella, J., Sánchez-SanSegundo, M., Ferrer-Cascales, R., & the Alzheimer Disease Neuroimaging Initiative. (2022). Predicting

Alzheimer's disease with practice effects, APOE genotype and brain metabolism. *Neurobiology of Aging*, *112*, 111–121. https://doi.org/10.1016/j.neurobiolaging.2021.12.011

Papp, K. V., Amariglio, R. E., Mormino, E. C., Hedden, T., Dekhytar, M., Johnson, K. A., Sperling, R. A., & Rentz, D. M. (2015). Free and cued memory in relation to biomarker-defined abnormalities in clinically normal older adults and those at risk for Alzheimer's disease. *Neuropsychologia*, *73*, 169–175. https://doi.org/10.1016/j.neuropsychologia.2015.04.034

Papp, K. V., Rentz, D. M., Maruff, P., Sun, C.-K., Raman, R., Donohue, M. C., Schembri, A., Stark, C., Yassa, M. A., Wessels, A. M., Yaari, R., Holdridge, K. C., Aisen, P. S., & Sperling, R. A. (2021). The computerized cognitive composite (C3) in an Alzheimer's disease secondary prevention trial. *The Journal of Prevention of Alzheimer's Disease*, *8*(1), 59–67. https://doi.org/10.14283/jpad.2020.38

Papp, K. V., Rentz, D. M., Orlovsky, I., Sperling, R. A., & Mormino, E. C. (2017). Optimizing the preclinical Alzheimer's cognitive composite with semantic processing: The PACC5. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *3*(4), 668–677. https://doi.org/10.1016/j.trci.2017.10.004

Papp, K. V., Rofael, H., Veroff, A. E., Donohue, M. C., Wang, S., Randolph, C., Grober, E., Brashear, H. R., Novak, G., Ernstrom, K., Raman, R., Aisen, P. S., Sperling, R., Romano, G., & Henley, D. (2022). Sensitivity of the preclinical Alzheimer's cognitive composite (PACC), PACC5, and repeatable battery for neuropsychological status (RBANS) to Amyloid status in preclinical Alzheimer's disease -Atabecestat Phase 2b/3 EARLY clinical trial. *The Journal of Prevention of Alzheimer's Disease*, *9*(2), 255–261. https://doi.org/10.14283/jpad.2022.17

Papp, K. V., Samaroo, A., Chou, H. C., Buckley, R., Schneider, O. R., Hsieh, S., Soberanes, D., Quiroz, Y., Properzi, M., Schultz, A., García-Magariño, I., Marshall, G. A., Burke, J. G., Kumar, R., Snyder, N., Johnson, K., Rentz, D. M., Sperling, R. A., & Amariglio, R. E. (2021). Unsupervised mobile cognitive testing for use in preclinical Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *13*(1), Article e12243. https://doi.org/10.1002/dad2.12243

Rentz, D. M., Amariglio, R. E., Becker, J. A., Frey, M., Olson, L. E., Frishe, K., Carmasin, J., Maye, J. E., Johnson, K. A., & Sperling, R. A. (2011). Face-name associative memory performance is related to amyloid burden in normal elderly. *Neuropsychologia*, *49*(9), 2776–2783. https://doi.org/10.1016/j.neuropsychologia.2011.06.006

Rentz, D. M., Weiss, B. K., Jacobs, E. G., Cherkerzian, S., Klibanski, A., Remington, A., Aizley, H., & Goldstein, J. M. (2017). Sex differences in episodic memory in early midlife: Impact of reproductive aging. *Menopause*, *24*(4), 400–408. https://doi.org/10.1097/GME.0000000000000771

Roque, N. A., & Boot, W. R. (2018). A new tool for assessing mobile device proficiency in older adults: The mobile device proficiency questionnaire. *Journal of Applied Gerontology: The Official Journal of the Southern Gerontological Society*, *37*(2), 131–156. https://doi.org/10.1177/0733464816642582

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging*, *30*(4), 507–514. https://doi.org/10.1016/j.neurobiolaging.2008.09.023

Samaroo, A., Amariglio, R. E., Burnham, S., Sparks, P., Properzi, M., Schultz, A. P., Buckley, R., Johnson, K. A., Sperling, R. A., Rentz, D. M., & Papp, K. V. (2021). Diminished Learning Over Repeated Exposures (LORE) in preclinical Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *12*(1), Article e12132. https://doi.org/10.1002/dad2.12132

Schaefer, S. Y., & Duff, K. (2017). Within-session and one-week practice effects on a motor task in amnestic mild cognitive impairment. *Journal of Clinical and Experimental Neuropsychology*, *39*(5), 473–484. https://doi.org/10.1080/13803395.2016.1236905

Sperling, R., Chua, E., Cocchiarella, A., Rand-Giovannetti, E., Poldrack, R., Schacter, D. L., & Albert, M. (2003). Putting names to faces: Successful encoding of associative memories activates the anterior hippocampal formation. *NeuroImage*, *20*(2), 1400–1410. https://doi.org/10.1016/S1053-8119(03)00391-4

Sundermann, E. E., Biegon, A., Rubin, L. H., Lipton, R. B., Landau, S., Maki, P. M., & the Alzheimer's Disease Neuroimaging Initiative. (2017). Does the female advantage in verbal memory contribute to underestimating AD pathology in women versus men? *Journal of Alzheimer's Disease*, *56*(3), 947–957. https://doi.org/10.3233/JAD-160716

von Stumm, S. (2018). Feeling low, thinking slow? Associations between situational cues, mood and cognitive function. *Cognition and Emotion*, *32*(8), 1545–1558. https://doi.org/10.1080/02699931.2017.1420632

Wechsler, D. (1944). *The measurement of adult intelligence* (3rd ed.). Williams & Wilkins. https://psycnet.apa.org/record/2006-12104-000

Wechsler, D. (1981). *WAIS-R: Manual: Wechsler adult intelligence scale-revised*. Harcourt Brace Jovanovich [for] Psychological Corp.

Wechsler, D. (1987). *Wechsler memory scale-revised*. Harcourt Brace Jovanovich. https://cir.nii.ac.jp/crid/1570854175088800512

Wei, Y., Zheng, P., Deng, H., Wang, X., Li, X., & Fu, H. (2020). Design Features for Improving Mobile Health Intervention User Engagement: Systematic Review and Thematic Analysis. *Journal of Medical Internet Research*, *22*(12), Article e21687. https://doi.org/10.2196/21687

Weizenbaum, E. L., Fulford, D., Torous, J., Pinsky, E., Kolachalama, V. B., & Cronin-Golomb, A. (2022). Smartphone-based neuropsychological assessment in Parkinson's disease: Feasibility, validity, and contextually driven variability in cognition. *Journal of the International Neuropsychological Society*, *28*(4), 401–413. https://doi.org/10.1017/S1355617721000503

Wilks, H., Aschenbrenner, A. J., Gordon, B. A., Balota, D. A., Fagan, A. M., Musiek, E., Balls-Berry, J., Benzinger, T. L. S., Cruchaga, C., Morris, J. C., & Hassenstab, J. (2021). Sharper in the morning: Cognitive time of day effects revealed with high-frequency smartphone testing. *Journal of Clinical and Experimental Neuropsychology*, *43*(8), 825–837. https://doi.org/10.1080/13803395.2021.2009447

Yelle, L. E. (1979). The learning curve: Historical review and comprehensive survey. *Decision Sciences*, *10*(2), 302–328. https://doi.org/10.1111/j.1540-5915.1979.tb00026.x

Received March 28, 2023
Revision received July 19, 2023
Accepted August 8, 2023 ■