Alzheimer's & Dementia
Diagnosis, Assessment
& Disease Monitoring

RESEARCH ARTICLE

Remote and in-clinic digital cognitive screening tools outperform the MoCA to distinguish cerebral amyloid status among cognitively healthy older adults

Louisa I. Thompson PhD[1] | Zachary J. Kunicki PhD, MS, MPH[1] | Sheina Emrani PhD[1] | Jennifer Strenger MA[1] | Alyssa N. De Vito PhD[1] | Karysa J. Britton MA, MS[2] | Catherine Dion MS[3] | Karra D. Harrington PhD[5] | Nelson Roque PhD[4] | Stephen Salloway MD[1] | Martin J. Sliwinski PhD[5] | Stephen Correia PhD[6] | Richard N. Jones ScD[1]

[1]Department of Psychiatry & Human Behavior, Alpert Medical School, Brown University, Providence, Rhode Island, USA

[2]Memory & Aging Program, Butler Hospital, Providence, Rhode Island, USA

[3]Department of Clinical and Health Psychology, University of Florida, Gainesville, Florida, USA

[4]Department of Psychology, University of Central Florida, Orlando, Florida, USA

[5]Department of Human Development & Family Studies, The Pennsylvania State University, University Park, Pennsylvania, USA

[6]Department of Health Promotion and Behavior, School of Public Health, University of Georgia, Florida, USA

Correspondence
Louisa I. Thompson, Department of Psychiatry, Alpert Medical School, Brown University, 345 Blackstone Blvd., Providence, RI 02906, USA.
Email: louisa_thompson@brown.edu

Abstract

INTRODUCTION: We evaluated the accuracy of remote and in-person digital tests to distinguish between older adults with and without AD pathological change and used the Montreal Cognitive Assessment (MoCA) as a comparison test.

METHODS: Participants were 69 cognitively normal older adults with known beta-amyloid (Aβ) PET status. Participants completed smartphone-based assessments 3×/day for 8 days, followed by TabCAT tasks, DCTclock™, and MoCA at an in-person study visit. We calculated the area under the curve (AUC) to compare task accuracies to distinguish Aβ status.

RESULTS: Average performance on the episodic memory (Prices) smartphone task showed the highest accuracy (AUC = 0.77) to distinguish Aβ status. On in-person measures, accuracy to distinguish Aβ status was greatest for the TabCAT Favorites task (AUC = 0.76), relative to the DCTclock™ (AUC = 0.73) and MoCA (AUC = 0.74).

DISCUSSION: Although further validation is needed, our results suggest that several digital assessments may be suitable for more widespread cognitive screening application.

KEYWORDS
cognitive screening, digital technology, early diagnosis, preclinical Alzheimer's disease

1 | INTRODUCTION

With record levels of older adults currently living with Alzheimer's disease (AD), there is an urgent need for more sensitive and scalable screening tools to facilitate earlier diagnosis and intervention for AD and related dementias.[1] In part due to multiple studies showing that cognitive impairment goes unrecognized or misdiagnosed in 27% to 81% of patients in primary care settings,[2,3] the Patient Protection and Affordable Care Act now requires an Annual Wellness Visit for older adults, including assessment to detect cognitive impairment. The use of brief, easily administered cognitive screening tools may help to identify patients with dementia; however, existing screening tools have many limitations and are typically not sensitive to mild cognitive impairment or preclinical AD.[4,5] Additionally, primary care providers may not have time to administer cognitive tests, thereby limiting the number of patients that are screened.[3] Therefore, there is a need to find alternative screening measures that are able to circumvent these challenges.

Multiple tablet-based cognitive assessments have been shown to be accurate and reliable for in-person and remote administration, and can be more readily adjusted for different ethnicities, cultures, and languages.[6–9] More recently, smartphone app- and web-based cognitive assessments for remote screening and monitoring have also been developing rapidly.[10–12] Older adults are increasingly using both smartphones and tablets,[13] and the adoption of this technology for cognitive screening in various settings is feasible, including in older adults with no prior experience with touchscreen technology.[14–16] There are several potential advantages to using digital cognitive assessments in remote and primary care settings, including reduced practitioner burden, improved ease of repeat testing, and better cost and time effectiveness of care.[17,18] Digital cognitive assessments may also allow clinicians to reach underserved and under-resourced populations and adapt digital assessments to numerous languages, thus addressing healthcare disparities.[19] Visual working memory and episodic memory measures are shown to be sensitive to early-stage AD and can be easily adapted for digital presentation with minimal need for written or auditory language.[20,21] Smartphone- and web-based cognitive screening tools have the additional advantage of improving accessibility for those in rural areas or with physical or financial limitations that prevent them from visiting the clinic.[17,22,23] Remote assessment approaches also have psychometric benefits, including increased ecological validity (ie, being at home may reduce test anxiety and better capture day-to-day cognitive abilities), easy collection of multiple data points at different times of day, and automated scoring mechanisms that reduce scoring errors, potentially leading to more reliable and sensitive detection of cognitive change over time.[18,24] Despite these benefits, there are also potential costs to be aware of, including poor adherence, and a lack of examiner control over environmental distractors or interference from others.[25] For this reason, remote digital assessments are likely best utilized as a method for screening and monitoring, but not diagnosis, of cognitive disorders.

There are numerous examples of successful in-clinic digital assessment validation work in the literature. The TabCAT Brain Health

---

RESEARCH IN CONTEXT

1. **Systematic review**: The authors reviewed the literature using Google Scholar and meeting abstracts and presentations. Existing cognitive screening measures fall short in capturing preclinical Alzheimer's disease. Digital assessment technology has the potential to deliver more efficient and sensitive screening, but requires rigorous validation.

2. **Interpretation**: Our findings show that several brief digital screening approaches (memory-specific tasks on the Mobile Monitoring of Cognitive Change [M2C2] and Tab-CAT) are noninferior to the Montreal Cognitive Assessment for distinguishing cerebral beta-amyloid status.

3. **Future directions**: Further validation of these digital screening tools in community and clinic-based samples is needed. Additional studies should also include the following: (1) examination of M2C2 task convergent validity with a battery of gold-standard neuropsychological tests; (b) evaluation of digital screening measure psychometrics in diverse underrepresented and underserved populations; and (c) investigation of repeated remote assessment metrics, including practice effects and intraindividual variability, for screening purposes.

Assessment (BHA) is a 15-min, tablet-based cognitive screening tool being used globally for research in a variety of settings, languages, and cultures.[6,7,26] BHA subtests have shown moderate to high correlations with reference standard neuropsychological tests, as well as medial temporal, frontal, parietal, and basal ganglia volumes.[26] The BHA has also demonstrated sensitivity to predict cerebral beta-amyloid (Aβ) status in cognitively normal individuals and associations with regional levels of tau positron emission tomography (PET) signal in the medial temporal lobe.[27,28] Harvard Aging Brain Study (HABS) researchers have also developed and/or validated several digital assessments with cognitively healthy adults, including tablet and digital pen-based measures. The HABS demonstrated that performance on a digitized clock drawing measure, the DCTclock™, is able to distinguish between normal cognition and MCI and is associated with cerebral amyloid and tau pathology, suggesting utility for early detection.[29] This task has also recently been shown to effectively distinguish between MCI subtypes (eg, amnestic versus dysexecutive) in a memory clinic sample.[30]

Finally, a number of studies are leveraging smartphones to measure cognitive performance remotely, with the aim of identifying early markers of decline.[31] Smartphone-based digital assessments are reliable and feasible for repeated assessments of cognitive function in naturalistic settings.[24,31–33] In one study, an ecological momentary assessment protocol was employed among a diverse adult lifespan sample (aged 24 to 65 years) where participants completed smartphone-based cognitive tests up to five times a day for 14

consecutive days.[24] These measures were highly reliable and associated with in-lab assessments, and adherence to the remote assessment protocol was good. We have similarly demonstrated high daily adherence and within-person reliabilities of average scores across sessions ranging from 0.89 to 0.97 in a sample of cognitively unimpaired older adults who completed brief app-based testing sessions three times per day for several days.[32] Additionally, older adults with MCI have recently been found to have greater within-day variability on smartphone-based measures of processing speed and memory compared to their cognitively healthy counterparts.[31] Within-day variability versus day-to-day variability was found to be more sensitive to MCI, providing important insight to how repeated remote digital assessments can be utilized in a clinical setting.[31] Together, these findings suggest that using mobile app-based cognitive assessments in short bursts is feasible and reliable in older adult samples.

The overarching goal of this pilot study was to conduct an initial validation of novel smartphone-based remote assessments and clinic-based digital assessments in a convenience sample of cognitively healthy older adults. We compared the accuracy of these digital measures to distinguish between participants with and without AD pathological change (preclinical AD) determined via Aβ PET scan. We also compared digital task accuracies with the Montreal Cognitive Assessment (MoCA), a clinical cognitive screening tool sensitive to MCI, as a reference standard. We hypothesized that remote digital assessments would demonstrate noninferior accuracy to detect preclinical AD relative to clinic-based digital screening tools and the MoCA. Further, we hypothesized that participants with AD pathological change would show reduced learning curves on smartphone-based visual working memory and episodic memory subtests over an 8-day assessment period relative to those without AD pathological change, revealing subtle cognitive impairment.

2 | METHODS

2.1 | Procedures

Participants were cognitively unimpaired older adults (ages 60 to 80) from the Butler Hospital Alzheimer's Prevention Registry, a local database of older adults interested in AD research at the Butler Hospital Memory and Aging program. We used a targeted recruitment to enroll up to 100 individuals with prior amyloid PET data (elevated [Aβ+] or non-elevated [Aβ−] determined via clinical read) as well as individuals without PET data. A total of 256 individuals were invited to the study via email or phone call, and 146 consented and completed online screening. Of those, 23 were excluded during screening. Three participants withdrew from the study after enrollment. An enrollment diagram and inclusion and exclusion table are included in the supplementary material (Supplemental 1 and Supplemental 2, respectively). Screening was conducted via online survey and the modified Telephone Interview for Cognitive Status (TICSm).[34] Unimpaired cognition was defined as a TICSm cutoff score of ≥34.[35] Participants completed an exit survey online to provide feedback at the end of the study. A $20

gift card compensation was provided. The project was approved by the Butler Hospital Institutional Review Board and all participants gave consent.

2.2 | Remote assessment

Android smartphones preloaded with the cognitive assessment app were shipped to participants with a detailed use guide. All phones were locked down to prevent use of other features (eg, web browsing, camera). Cognitive tasks were completed for 8 consecutive days in the Mobile Monitoring of Cognitive Change (M2C2) app, a cognitive testing platform developed as part of the National Institute of Aging's Mobile Toolbox initiative (Figure 1).[24] On each of the 8 days, participants completed brief (ie, 3 to 4 minutes) M2C2 sessions within morning, afternoon, and evening time windows. Extra sessions (optional or makeup) could be completed on day 9. Support for the M2C2 assessments was logged during the remote phase only (n = 52). A majority (30/52) of participants required a support check-in, but only two required more than one. Top reasons for support included the following: could not hear phone beep, logged out of the app, start date mismatch/error, and date adjustment needed due to participant reasons or shipping issues.

2.3 | M2C2 tasks

We selected established M2C2 cognitive measures of visual working memory (Color Shapes), processing speed (Symbol Match), and episodic memory (Prices) with prior evidence of sensitivity to age and/or age-related neuropathology as described previously (Figure 1).[24,32] Each task took approximately 60 seconds to complete. The Prices task is a forced-choice recognition task, where participants incidentally encode grocery item-price pairs for later recall while judging whether or not the item's price is "good."[36,37] Recall trials begin immediately after the learning trials. Performance is summarized as the proportion of correct responses on 10 recall trials. The Color Shapes task is a visual array change detection task, measuring intra-item feature binding, where participants determine if shapes change color across two sequential presentations in which the shape locations change.[38,39] Performance is summarized with the hit rate (proportion of correct identifications) and false-alarm rate (proportion of misidentified stimuli).[40] The Symbol Match task is a speeded continuous performance task of conjunctive feature search, where participants are asked to identify matching symbol pairs.[32,41] Performance is summarized as the median reaction time to complete the task across all trials (in milliseconds).

2.4 | In-person assessment

After completion of the remote assessments, participants were scheduled for a single study visit to complete in-person neuropsychological
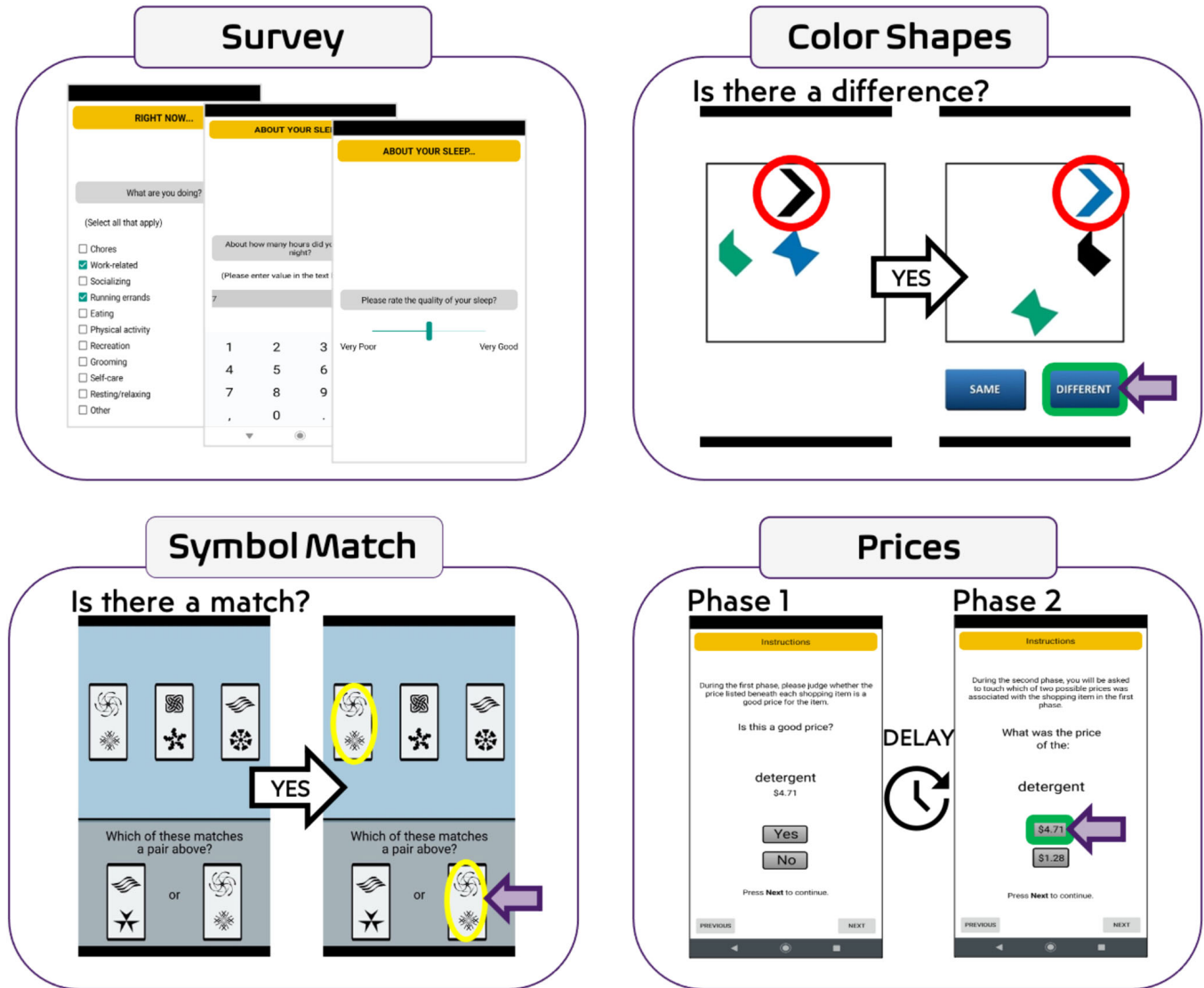
FIGURE 1 Mobile assessment interface for Mobile Monitoring of Cognitive Change tasks. Participants complete approximately 1 minute of survey questions at the start of each study session, including questions about sleep quality at the start of the morning session. The Prices task is completed in two phases: the learning trials (phase 1), followed by immediate delay forced choice recognition trials (phase 2).

assessments, including the MoCA, TabCAT BHA, and DCTclock™ (Linus Health). The MoCA is a 10-minute, paper-and-pencil-based measure routinely used in both clinical practice and research to screen for MCI and dementia in older adults.[42] The TabCAT BHA is a 15-minute, tablet-based cognitive screening test developed at the University of California San Francisco Memory and Aging Center.[26] The BHA consists of subtests of paired associates learning, executive functions and processing speed, visuospatial skills, and language, and has alternative forms developed for repeat assessment. The DCTclock™ is similar to the standard paper-and-pencil clock drawing test with command and copy conditions, but uses a digital pen to capture clock images drawn by the participant. A machine learning-based scoring algorithm is used to evaluate the drawing process and features and calculates a total score as well as other composite performance metrics.[43]

2.5 | Protocol adjustments for the SARS-CoV-2 pandemic

Due to SARS-CoV-2 pandemic restrictions, the first 52 participants enrolled in the study completed only the remote portion of the study and have been characterized previously.[32] These participants were later recontacted and invited to participate in the in-person study visit as described above. A total of 40/52 participants completed this follow-up in-person visit, which took place approximately 18 months ($M = 513.6$ days, SD = 86.7 days) after they completed the remote assessment (see Supplemental 1). The remainder of the sample ($n = 70$) enrolled in the study after restrictions were lifted and completed the remote M2C2 sessions followed by the in-person study visit within approximately 1 month of each other ($M = 12.5$ days, SD = 9.4 days). Because participants from across all protocol phases were included

in this analysis, a variable for protocol type (remote only, short delay follow-up, and long delay follow-up) was included in our statistical models to group participants and control for potential effects related to this variation.

2.6 | Analysis plan

We conducted a receiver operating characteristic (ROC) curve analysis and computed the area under the ROC curve (AUC) analyses to characterize the accuracies of each digital test and the MoCA to distinguish Aβ status, after controlling for age and protocol type. To determine whether accuracy to distinguish Aβ status differed significantly between the digital cognitive tests and the MoCA, we used a q-test to compare standardized regression coefficients, taking a noninferiority approach.[44,45] Noninferiority testing is a form of equivalence testing, where researchers test if there is an absence of the smallest effect size of clinical interest.[44] An equivalence test would test if the relationship between a given digital task and Aβ status is equivalent (ie, not higher or not lower than some pre-specified caliper) to the relationship between MoCA results and Aβ status. This test is two-sided, by definition, and approached using two one-sided tests. Noninferiority testing is analogous to a one-sided test, where the purpose is to determine if one effect is "not worse" than another effect (eg, if the relationship between a digital task and Aβ status is not weaker than the relationship between MoCA results and Aβ status). If the results show noninferiority between a digital task and the MoCA, then it could be considered reasonable to use the digital task in place of the MoCA when conducting research on Aβ and AD. We obtained the standardized regression coefficients by fitting a probit regression model with Aβ status as the outcome, cognitive test as the predictor, and adjusting for age and protocol type. We then converted the regression coefficient to a polyserial regression coefficient, which can be Fisher z-transformed to conduct the q-test.[45] The value of q is defined as the difference between two Fisher z-transformed correlations, where $q = 0.10$ is a small effect, 0.30 is a moderate effect, and 0.50 is a large effect. For the noninferiority analysis, we set a lower bound of 0.30 based on the moderate effect size. Thus, if the lower 90% confidence interval on the difference between standardized regression coefficients did not exceed −0.30, we could conclude the digital task was not inferior to the MoCA.[44] The 90% confidence interval is used for this analysis, because the noninferiority test is analogous to a one-sided statistical test.

3 | RESULTS

3.1 | Participant characteristics

Of the 122 participants who completed remote assessments, 73 had Aβ PET status data (Aβ+ $n = 25$, Aβ− $n = 48$). To reduce the possibility of a false negative in the Aβ− group, we restricted the allowable time window between the PET date and study screening date for this group to no more than 3 years unless more recent biomarker confirmation of

status was available. This excluded four Aβ− participants, and thus our final analytic sample consisted of 69 participants (see Supplemental 3). There were no significant demographic differences between participants by Aβ PET status, with the exception of age, which was slightly higher in the group Aβ+ group (see Table 1 demographics). There were also no significant differences in Aβ status by protocol type used to adapt to the SARS-CoV-2 pandemic. Finally, there were no significant differences in baseline characteristics by protocol type, with the exception of age. The short delay group was somewhat younger (mean age 67.5) than the long delay (mean age 70.5) and remote only (mean age 70.1) groups ($p = 0.03$). The final sample ($n = 69$) had a mean age of 69 years (SD = 4.6), a mean education of 16.4 years (SD = 2.6), was 72% female, and 91% White (Table 1).

Participants completed an average of 22.2 (SD = 2.9) out of 24 assigned M2C2 sessions (3 sessions per day for 8 days). Adherence ranged from 100% on days 1 to 2, to 88% on day 7, and 65% on day 8. On average, the Aβ+ group completed 22.9 sessions (SD = 1.4) and the Aβ− group completed 21.8 sessions (SD = 3.3). Performance on all M2C2 tasks was significantly lower among participants with older age. Women performed significantly better on working memory and episodic memory tasks relative to men. There were no detectable significant differences in task performance on the basis of race or education. See Table 2 for correlations between demographic variables and cognitive assessments. Age, sex, and protocol type were controlled for in subsequent analyses.

3.2 | Task accuracies to detect Aβ status

To obtain summary-level area under the ROC curve (AUC) estimates, we used logistic regression models with Aβ status as the outcome, and cognitive tasks (ie, M2C2 tasks, MoCA, DCTclock™, or TabCAT) as predictors. Different models were run for each cognitive task. Table 3 displays the AUC results for each cognitive test. Standardized regression coefficients are reported for each task with and without adjustment for age, sex, and protocol type. The adjusted baseline AUC was 0.73. The Prices task showed the highest AUC among M2C2 tasks (AUC = 0.77) (Figure 2). The remaining M2C2 tests had AUCs of 0.73 (Figure 2). The TabCAT Favorites task showed the highest AUC among in-person tests (AUC = 0.76), followed by TabCAT BHA Composite, DCTclock™ (AUC = 0.73), and the MoCA (AUC = 0.74) (Figure 3). The AUC analyses for M2C2 tasks were run with and without the inclusion of the remote only participants (Supplemental 4) to allow for more direct comparison between the M2C2 and in-person tasks. No qualitative difference was observed between the analyses with and without these participants.

3.3 | Comparison with the MoCA

The righthand columns of Table 3 show the results of comparing the M2C2 tasks, TabCAT, and DCTclock™ to the MoCA including q and the q-test results. The q and the q-test results were calculated based on the

TABLE 1 Sample characteristics by amyloid status (N = 69).

Characteristic	Overall N = 69[a]	Aβ− n = 44[a]	Aβ+ n = 25[a]	p-value[b]
Race	–	–	–	0.44
American Indian or Alaska Native	1 (1.5%)	0 (0%)	1 (4.3%)	–
Asian	0 (0%)	0 (0%)	0 (0%)	–
Black or African American	2 (3.0%)	2 (4.5%)	0 (0%)	–
More than one race	3 (4.5%)	2 (4.5%)	1 (4.3%)	–
White	61 (91%)	40 (91%)	21 (91%)	–
Unknown	2	0	2	–
Ethnicity	–	–	–	0.55
Hispanic or Latino	2 (3.2%)	2 (4.8%)	0 (0%)	–
Non-Hispanic or Latino	61 (97%)	40 (95%)	21 (100%)	–
Unknown	6	2	4	–
Gender	–	–	–	0.95
Female	50 (72%)	32 (73%)	18 (72%)	–
Male	19 (28%)	12 (27%)	7 (28%)	–
Age (years)	69.1 (4.6)	68.1 (4.0)	70.8 (5.0)	0.02
Education (years)	16.35 (2.60)	16.32 (2.63)	16.40 (2.61)	0.90
MoCA (total score)	27.10 (1.91)	27.32 (1.86)	26.67 (1.96)	0.21
TICSm (total score)	39.01 (3.09)	39.43 (3.34)	38.28 (2.48)	0.14
Protocol type	–	–	–	0.38
Delayed	31 (45%)	21 (48%)	10 (40%)	–
Moderately delayed	30 (43%)	20 (45%)	10 (40%)	–
Remote only	8 (17%)	3 (6.8%)	5 (20%)	–

Abbreviations: Aβ, beta-amyloid; MoCA, Montreal Cognitive Assessment; TICSm, modified Telephone Interview for Cognitive Status.
[a]n (%); Mean (SD).
[b]Fisher's exact test; Pearson's Chi-squared test; One-way ANOVA.

age- and protocol type-adjusted standardized regression coefficients. TabCAT Favorites and M2C2 Prices accuracy were both superior to the MoCA (Table 3). TabCAT BHA composite was the only other metric that achieved noninferiority to the MoCA. The M2C2 Color Shapes hit rate and Symbol match reaction time were both clearly noninferior. DCTclock™ and Color Shapes False alarms were borderline inferior (Table 3).

4 | DISCUSSION

Digital cognitive assessments are increasingly utilized in clinical research as a novel method to detect AD-related pathological and cognitive changes that occur years before clinical diagnosis.[14] The present study sought to evaluate two primary aims: first—to examine whether brief, smartphone-based digital cognitive assessments could be used to distinguish cognitively normal individuals based on Aβ PET status; and second—to compare several novel digital assessments (both remote and in-person) to the MoCA, a standard paper-and-pencil-based cognitive screening measure, in terms of their ability to distinguish cerebral Aβ PET status.

Consistent with expectations and prior work, two digital tasks assessing memory, the TabCAT Favorites task (administered by an examiner via tablet) and the M2C2 Prices task (self-administered remotely via smartphone), demonstrated the best ability to distinguish Aβ PET status of all administered tasks. Both of these measures utilize a paired associates learning paradigm consistently shown to be sensitive to subtle AD-related cognitive impairment and progression of cerebral Aβ burden on PET imaging.[46,47] Our findings also show that both the TabCAT Favorites task and M2C2 Prices task were superior to the MoCA for distinguishing Aβ status. These findings suggest that TabCAT Favorites and M2C2 Prices (or similar paired associates learning paradigms) may improve upon standard paper-and-pencil-based screening measures for detecting early AD-related cognitive changes, while providing advantages of brevity and easy administration with digital testing. The TabCAT BHA composite score also demonstrated noninferiority to the MoCA, suggesting that this may be an additional measure to consider for further investigation.

In contrast, the DCTclock™ and the M2C2 Symbol Match and Color Shapes tasks failed to demonstrate noninferiority to the MoCA and had little to no association with Aβ status, findings that are somewhat inconsistent with a small but growing literature suggest-

TABLE 2 Summary of correlations between cognitive assessments (N = 69).

Variable	Age	Education	MoCA	TabCAT (BHA Composite)	TabCAT (Favorites)	Color Shapes (hit rate)	Color Shapes (false alarm)	Symbol Match	Prices	DCTclock™ (total score)	Female
Age	–	0.01 (–0.22, 0.25)	–0.14 (–0.38, 0.11)	–0.02 (–0.44, 0.41)	–0.22 (–0.55, 0.16)	–0.27* (–0.48, –0.04)	0.35*** (0.13, 0.55)	0.43*** (0.21, 0.60)	–0.38** (–0.56, –0.16)	–0.08 (–0.32, 0.16)	–0.28* (–0.48, –0.04)
Education		–	–0.03 (–0.38, 0.22)	0.05 (–0.32, 0.40)	–0.09 (–0.37, 0.22)	–.04 (–0.28, 0.20)	0.15 (–0.09, 0.37)	–0.13 (–0.36, 0.11)	–0.09 (–0.32, 0.15)	0.07 (–0.19, 0.32)	–0.03 (–0.26, 0.21)
MoCA			–	0.46** (0.11, .71)	0.37* (0.07, 0.61)	0.26* (0.01, 0.48)	–0.37** (–0.55, –0.14)	–0.18 (–0.41, 0.07)	0.37** (0.13, 0.57)	0.35** (0.11, 0.56)	0.12 (–0.13, 0.36)
TabCAT (Composite)				–	0.62*** (0.39, 0.78)	0.22 (–0.12, 0.51)	–0.19 (–0.46, 0.10)	–0.09 (–0.37, 0.21)	0.23 (–0.09, 0.46)	0.29 (–0.10, 0.60)	0.15 (–0.15, 0.43)
TabCAT (Favorites)					–	0.19 (–0.08, 0.44)	–0.23 (–0.51, 0.08)	0.09 (–0.25, 0.40)	0.40*** (0.17, 0.59)	0.30 (–0.19, 0.66)	0.24 (–0.12, 0.55)
Color Shapes (Hit rate)						–	–0.50*** (–0.66, –0.30)	–0.19 (–0.41, 0.05)	0.31** (0.08, 0.51)	0.23 (–0.03, 0.45)	0.37*** (0.14, 0.56)
Color Shapes (False alarm)							–	0.39*** (0.17, 0.57)	–0.37*** (–0.56, –0.15)	–0.18 (–0.41, 0.07)	–0.17 (–0.39, 0.07)
Symbol Match								–	–0.14 (–0.37, 0.10)	–0.07 (–0.31, 0.18)	–0.08 (–0.31, 0.16)
Prices									–	0.29* (0.02, 0.52)	0.21 (–0.03, 0.42)

Note: Symbol match is the median response time, and Prices is the proportion correct. The 95% confidence intervals are shown in parentheses.

Abbreviations: MoCA, Montreal cognitive assessment.

*p < 0.05, **p < 0.01, ***p < 0.001.

TABLE 3 Area under the curve and regression coefficients for each cognitive test, comparison of regression coefficients (*q*) for each test relative to the MoCA.

Cognitive Test	AUC[a]	Unadjusted standardized regression coefficients	Adjusted standardized regression coefficients[a]	q	90% CI	q-test	p-value	Noninferiority
In-person tests								
MoCA (total score)	0.74	−0.12	−0.11	–	–	–	–	NA
TabCAT (favorites)	0.76	−0.40	−0.42	0.33	0.13, 0.54	2.72	0.01	Noninferior (superior)
TabCAT (composite)	0.73	−0.11	−0.15	0.04	−0.16, 0.24	0.33	0.75	Noninferior
DCTclock™ (total score)	0.73	−0.01	−0.01	−0.10	−0.303, 0.10	−0.82	0.41	Possibly inferior, uncorrelated with Aβ+
M2C2 tests								
Prices (accuracy)	0.77	−0.43	−0.41	0.33	0.12, 0.53	2.65	0.01	Noninferior (superior)
Color shapes (false alarm)	0.73	−0.17[a]	−0.07[b]	−0.18	−0.39, 0.02	−1.50	0.13	Possibly inferior
Color shapes (hit rate)	0.73	−0.09	−0.02	−0.09	−0.29, 0.11	−0.72	0.47	Noninferior, uncorrelated with Aβ+
Symbol match (reaction time)	0.73	−0.01[b]	0.18[b]	0.07	−0.13, 0.28	0.59	0.56	Noninferior

Note: AUC is discriminating between amyloid positive (*n* = 25) and amyloid negative (*n* = 44) participants. Standardized regression coefficients refer to an increase in the log odds of being amyloid positive per standard deviation difference in the test score. The statistic *q* refers to the difference between two Fisher's *z* transformed standardized regression coefficients, where 0.1 is a small effect, 0.3 is a moderate effect, and 0.5 is a large effect. The value of *q* is defined as the difference between adjusted standardized regression coefficients (MoCA − other test), with *q* calculated based on the age- and protocol-adjusted standardized regression coefficients. Noninferiority can be claimed if the 90% CI does not exceed −0.30.

Abbreviations: AUC, area under the curve; CI, confidence interval; M2C2, Mobile Monitoring of Cognitive Change; MoCA, Montreal Cognitive Assessment.

[a]Models adjusted for age and protocol type (short delay follow-up, long delay follow-up, remote only). The adjusted coefficients are adjusted for age and protocol type, whereas the unadjusted coefficients are not.

[b]Coefficient multiplied by −1 so higher scores indicate better performance instead of worse performance.
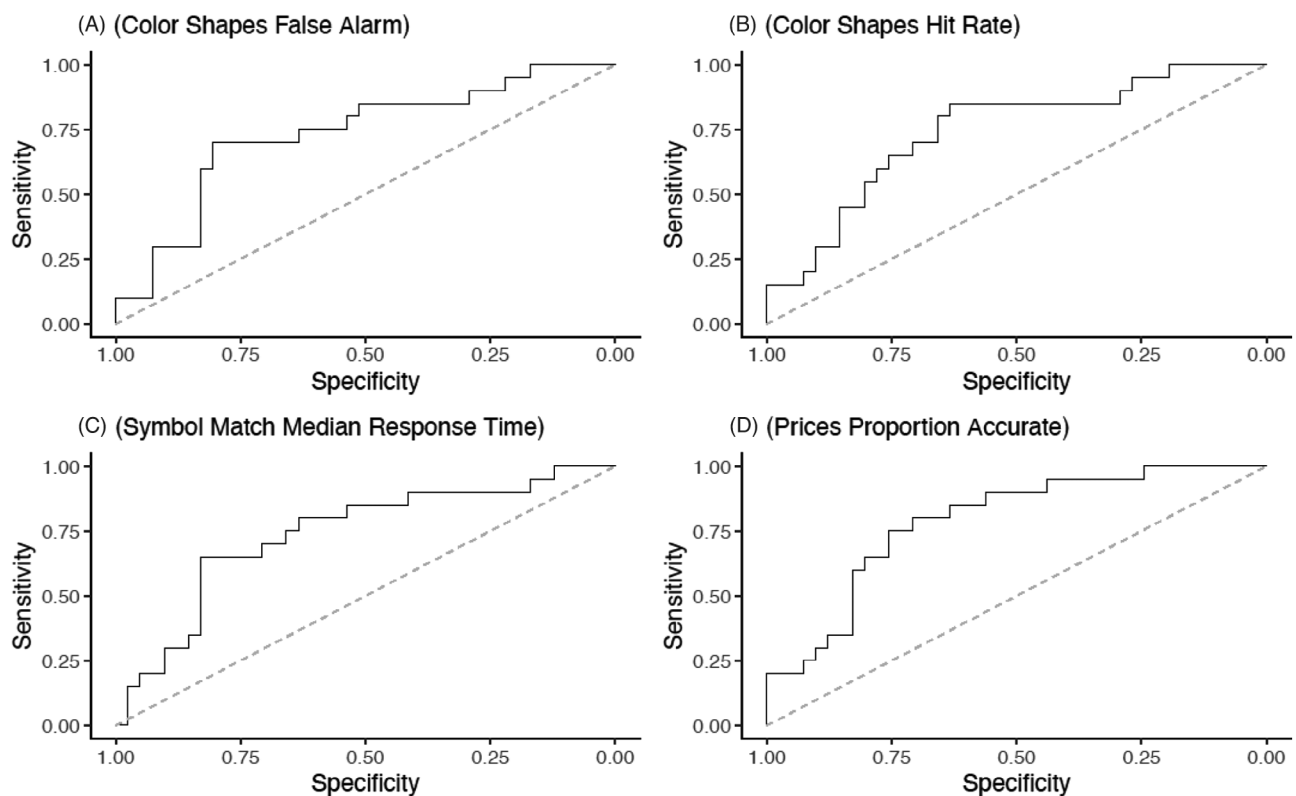
FIGURE 2 Receiver operating characteristic curves for Mobile Monitoring of Cognitive Change metrics distinguishing amyloid positron emission tomography status.
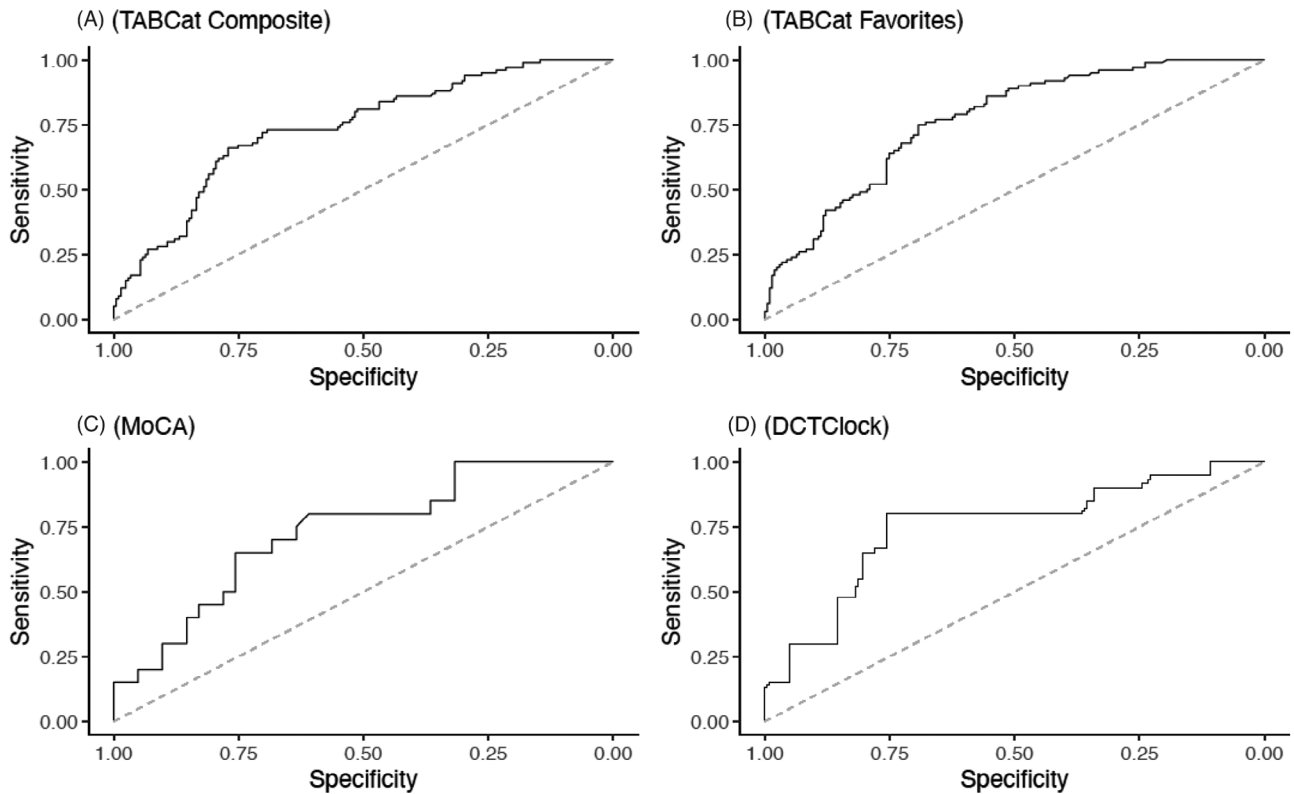
FIGURE 3 Receiver operating characteristic curves for in-person digital assessment measures distinguishing amyloid PET status. MoCA, Montreal Cognitive Assessment; PET, positron emission tomography.

ing that these metrics may be sensitive to preclinical and prodromal AD.[11,29] Possible reasons for the discrepancy include the small sample size or fact that we were limited to dichotomous PET Aβ status data instead of standardized uptake value ratio (SUVr) values. Previous studies using the DCTclock[TM] had continuous biomarker variables and therefore may have been better suited to detect more subtle associations.[29] The absence of a relationship between Color Shapes performance and Aβ status in this study is partially aligned with a recent meta-analysis showing that group differences in working memory performance between individuals who were A$\beta+$ versus A$\beta-$ were much smaller than differences seen for global cognition and other domains.[48] Additionally, recent work that has examined color-shape binding paradigms and their relationship with AD biomarkers found stronger associations with tau than with Aβ.[49]

4.1 | Limitations and future directions

The present study has several strengths including the head-to-head comparison of mutliple digital tests, inclusion of Aβ PET scan data, and use of a fully remote, app-based cognitive assessment protocol, which allowed us to continue data collection despite research disruptions during the SARS-CoV-2 pandemic. Despite these strengths, this study is not without limitations that should be considered and addressed in future work. In particular, more work is needed before these digital measures may be ready for use as screening tools in general medical settings. First, given that the study was conducted in a cognitively normal sample enriched for AD preclinical pathologic change, additional work is needed to determine if task performance predicts not only Aβ PET status, but additional clinical correlates such as conversion to cognitively impaired states over time. Relatedly, the present study was cross-sectional, and further research comparing digital assessments to standard screening measures over time is necessary to assess the reliability of the novel digital assessments for routine screening and monitoring in real-world clinical settings. Another consideration is the sociodemographic homogeneity of our sample. The majority of participants were White individuals and had high levels of education. These individuals also all had prior experience using smartphones. Finally, our participants were recruited from an AD research registry and therefore were self-selected and motivated to contribute to AD research. Together, these factors suggest that our findings may be limited in generalizability to the broader U.S. older adult population. Nonetheless, this pilot study supports further protocol development and validation efforts for larger studies trialing digital tools in broader settings, such as primary care.

Our future studies plan to examine M2C2 task convergent validity with a battery of gold-standard neuropsychological tests that measure the same cognitive domains, and evaluate the utility of digital screening measures in more racially and ethnically diverse populations, including Spanish-speaking individuals. We also seek to develop protocols for the use of these measures in primary care practices, with attention to maximizing adherence and convenience for patients, and

minimizing burden for healthcare providers. A unique advantage of in-person digital assessments is their potential to reduce burden for healthcare providers by automating much of the administration and scoring process. Finally, there are advantages to conducting remote, self-administered cognitive screenings via smartphone app, including the ability to characterize subtle performance changes over time via brief repeated testing sessions. More sensitive and ecologically valid metrics of cognition are obtainable through high frequency remote assessment, including examination of practice effects and intraindividual variability, which have previously demonstrated sensitivity to cerebral Aβ burden and MCI. These types of analyses will be a focus of future investigation in this study.[31,50,51] Important feasibility and scalability logistics to consider for future research include: (1) the need for participants to use their own smartphones and other electronic devices to complete assessments, and (2) the need to reduce the number of assessment days to minimize participant burden.

5 | CONCLUSIONS

This study demonstrated that select brief digital cognitive assessments have the ability to distinguish individuals with elevated cerebral Aβ from those without elevated cerebral Aβ. Further, our results suggest that two measures, specifically the memory-based tasks in M2C2 and TabCAT, are superior to the MoCA, a gold standard paper-and-pencil cognitive screening measure, in the prediction of Aβ PET status. Although further validation in community and clinic-based samples is needed, these early positive results suggest that remote digital cognitive assessments offer opportunities for increased sensitivity to detect early pathological changes in neurodegenerative disorders and may be suitable for more widespread cognitive screening.

CONFLICT OF INTEREST STATEMENT
Dr. Stephen Salloway has been a paid consultant for Lilly, Biogen, Roche, Genentech, Eisai, Bolden, Amylyx, Novo Nordisk, Prothena, Ono, and Alnylam. Dr. Thompson has been a paid consultant for the Davos Alzheimer's Collaborative. All other authors have nothing to disclose.

CONSENT STATEMENT
All participants in the study gave consent.

ORCID
Louisa I. Thompson PhD https://orcid.org/0000-0002-7231-870X
Zachary J. Kunicki PhD, MS, MPH https://orcid.org/0000-0002-3466-7589
Sheina Emrani PhD https://orcid.org/0000-0002-1235-5285
Alyssa N. De Vito PhD https://orcid.org/0000-0002-4078-3869
Catherine Dion MS https://orcid.org/0000-0003-3153-7858
Karra D. Harrington PhD https://orcid.org/0000-0001-9230-2978
Stephen Correia PhD https://orcid.org/0000-0003-3416-3660
Richard N. Jones ScD https://orcid.org/0000-0002-1049-218X

REFERENCES
1. Association As. 2021 Alzheimer's Disease Facts and Figures. *Alzheimers Dement.* 2021;1(1):327-406.
2. Cordell CB, Borson S, Boustani M, et al. Alzheimer's Association recommendations for operationalizing the detection of cognitive impairment during the Medicare Annual Wellness Visit in a primary care setting. *Alzheimers Dement.* 2013;9(2):141-150.
3. Bradford A, Kunik ME, Schulz P, Williams SP, Singh H. Missed and delayed diagnosis of dementia in primary care. *Alzheimer Dis Assoc Disord.* 2009;23(4):306-314.
4. Borson S, Scanlan J, Hummel J, Gibbs K, Lessig M, Zuhr E. Implementing routine cognitive screening of older adults in primary care: process and impact on physician behavior. *J Gen Intern Med.* 2007;22(6):811-817.
5. Spencer R, Wendell C, Giggey P, et al. Psychometric limitations of the mini-mental state examination among nondemented older adults: an evaluation of neurocognitive and magnetic resonance imaging correlates. *Exp Aging Res.* 2013;39(4):382-397.
6. Tsoy E, Brugulat-Serrat A, Vandevrede L, et al. Distinct patters of associations between cognitive performance and Alzheimer's disease biomarkers in a multinational cohort of clinically normal older adults. *Alzheimers Dement.* 2021;17(S6):e055076.
7. Rodriguez-Salgado AM, Llibre-Guerra JJ, Peñalver AI, et al. Brief digital cognitive assessment for detection of cognitive impairment in low- and middle-income countries. *Alzheimers Dement.* 2021;17(S10):85-94.
8. Stricker NH, Lundt ES, Alden EC, et al. Longitudinal comparison of in clinic and at home administration of the cogstate brief battery and demonstrated practice effects in the Mayo clinic study of aging. *J Prev Alzheimer's Dis.* 2020;7:21-28.
9. Barnett JH, Blackwell AD, Sahakian BJ, Robbins TW. The paired associates learning (PAL) test: 30 years of CANTAB translational neuroscience from laboratory to bedside in dementia research. *Curr Top Behav Neurosci.* 2016;28:449-474.
10. Staffaroni AM, Taylor JC, Clark AL, et al. A remote smartphone cognitive testing battery for frontotemporal dementia: completion rate, reliability, and validity. *Alzheimers Dement.* 2021;17:e056136.
11. Papp KV, Samaroo A, Chou HC, et al. Unsupervised mobile cognitive testing for use in preclinical Alzheimer's disease. *Alzheimers Dement.* 2021;13(1):e12243.
12. Sabbagh MN, Boada M, Borson S, et al. Early detection of Mild Cognitive Impairment (MCI) in an at-home setting. *J Prev Alzheimer's Dis.* 2020;7(3):171-178.
13. Vogels E. Millennials stand out for their technology use, but older generations also embrace digital life. Pew Research Center Fact Tank. 2019. Published September 9, 2019. Accessed December 15, 2021. https://www.pewresearch.org/fact-tank/2019/09/09/us-generations-technology-use/
14. Öhman F, Hassenstab J, Berron D, Schöll M, Papp KV. Current advances in digital cognitive assessment for preclinical Alzheimer's disease. *Alzheimers Dement.* 2021;13(1):e12217.
15. Adler K, Apple S, Friedlander A, et al. Computerized cognitive performance assessments in the brooklyn cognitive impairments in health disparities pilot study. *Alzheimers Dement.* 2019;15(11):1420-1426.
16. Jacobs DM, Peavy GM, Banks SJ, Gigliotti C, Little EA, Salmon DP. A survey of smartphone and interactive video technology use by participants in Alzheimer's disease research: implications for remote cognitive assessment. *Alzheimers Dement.* 2021;13(1):e12188.

17. Staffaroni AM, Tsoy E, Taylor J, Boxer AL, Possin KL. Digital cognitive assessments for dementia: digital assessments may enhance the efficiency of evaluations in neurology and other clinics. *Pract Neurol (Fort Wash Pa)*. 2020;2020:24-45.

18. Sahoo S, Grover S. Technology-based neurocognitive assessment of the elderly: a mini review. *Consortium Psychiatricum*. 2022;3(1):37-44.

19. Gamaldo AA, Tan SC, Sardina AL, et al. Older black adults' satisfaction and anxiety levels after completing alternative versus traditional cognitive batteries. *J Gerontol Series B, Psychol Sci Soc Sci*. 2020;75(7):1462-1474.

20. Loewenstein DA, Curiel RE, Duara R, Buschke H. Novel cognitive paradigms for the detection of memory impairment in preclinical Alzheimer's disease. *Assessment*. 2018;25(3):348-359.

21. Jia Y, Woltering S, Deutz N, et al. Working memory precision and associative binding in mild cognitive impairment. *Exp Aging Res*. 2023:1-19.

22. Paddick S-M, Yoseph M, Gray WK, et al. Effectiveness of app-based cognitive screening for dementia by lay health workers in low resource settings. A validation and feasibility study in Rural Tanzania. *J Geriatr Psychiatry Neurol*. 2021;34(6):613-621.

23. Owens AP, Ballard C, Beigi M, et al. Implementing remote memory clinics to enhance clinical care during and after COVID-19. *Front Psychiatry*. 2020;11:579934.

24. Sliwinski MJ, Mogle JA, Hyun J, Munoz E, Smyth JM, Lipton RB. Reliability and validity of ambulatory cognitive assessments. *Assessment*. 2018;25(1):14-30.

25. Madero EN, Anderson J, Bott NT, et al. Environmental distractions during unsupervised remote digital cognitive assessment. *J Prev Alzheimer's Dis*. 2021;8(3):263-266.

26. Possin KL, Moskowitz T, Erlhoff SJ, et al. The brain health assessment for detecting and diagnosing neurocognitive disorders. *J Am Geriatr Soc*. 2018;66(1):150-156.

27. Tsoy E, Strom A, Iaccarino L, et al. Detecting Alzheimer's disease biomarkers with a brief tablet-based cognitive battery: sensitivity to Aβ and tau PET. *Alzheimers Res Ther*. 2021;13(1).

28. Tsoy E, Erlhoff SJ, Goode CA, et al. BHA-CS: a novel cognitive composite for Alzheimer's disease and related disorders. *Alzheimers Dement*. 2020;12(1):e12042.

29. Rentz DM, Papp KV, Mayblyum DV, et al. Association of digital clock drawing with PET amyloid and tau pathology in normal older adults. *Neurology*. 2021;96(14):e1844.

30. Matusz EF, Price CC, Lamar M, et al. Dissociating statistically determined normal cognitive abilities and mild cognitive impairment subtypes with DCTclock. *J Int Neuropsychol Soc*. 2023;29(2):148-158.

31. Cerino ES, Katz MJ, Wang C, et al. Variability in Cognitive performance on mobile devices is sensitive to mild cognitive impairment: results from the Einstein Aging Study. *Front Digit Health*. 2021;3:758031.

32. Thompson LI, Harrington KD, Roque N, et al. A highly feasible, reliable, and fully remote protocol for mobile app-based cognitive assessment in cognitively healthy older adults. *Alzheimers Dement*. 2022;14(1):e12283.

33. Wilks H, Aschenbrenner AJ, Gordon BA, et al. Sharper in the morning: cognitive time of day effects revealed with high-frequency smartphone testing. *J Clin Exp Neuropsychol*. 2021;43(8):825-837.

34. Brandt JSM, Folstein M. The telephone interview for cognitive status. *Neuropsychiatry Neuropsychol Behav Neurol*. 1988(1):111-117.

35. Cook SE, Marsiske M, McCoy KJM. The use of the Modified Telephone Interview for Cognitive Status (TICS-M) in the detection of amnestic mild cognitive impairment. *J Geriatr Psychiatry Neurol*. 2009;22(2):103-109.

36. Gallo DA, Shahid KR, Olson MA, Solomon TM, Schacter DL, Budson AE. Overdependence on degraded gist memory in Alzheimer's disease. *Neuropsychology*. 2006;20(6):625-632.

37. Naveh-Benjamin M. Adult age differences in memory performance: tests of an associative deficit hypothesis. *J Exp Psychol Learn Mem Cogn*. 2000;26(5):1170-1187.

38. Parra MA, Sala SD, Abrahams S, Logie RH, Guillermo Méndez L, Lopera F. Specific deficit of colour-colour short-term memory binding in sporadic and familial Alzheimer's disease. *Neuropsychologia*. 2011;49:1943-1952.

39. Parra MA, Abrahams S, Logie RH, Méndez LG, Lopera F, Della Sala S. Visual short-term memory binding deficits in familial Alzheimer's disease. *Brain*. 2010;133(9):2702-2713.

40. Stanislaw H, Todorov N. Calculation of signal detection theory measures. *Behav Res Methods Instrum Comput*. 1999;31(1):137-149.

41. Deary IJ, Johnson W, Starr JM. Are processing speed tasks biomarkers of cognitive aging? *Psychol Aging*. 2010;25(1):219-228.

42. Nasreddine ZS, Phillips NA, Bã©Dirian VR, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005;53(4):695-699.

43. Souillard-Mandar W, Davis R, Rudin C, et al. Learning classification models of cognitive conditions from subtle behaviors in the digital clock drawing test. *Machine Learning*. 2016;102(3):393-441.

44. Lakens D, Scheel AM, Isager PM. Equivalence testing for psychological research: a tutorial. *Adv Methods and Pract Psychol Sci*. 2018;1(2):259-269.

45. Cohen J. Differences between correlation coefficients. *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates; 1988:109-144.

46. Grober E, Lipton RB, Sperling RA, et al. Associations of stages of objective memory impairment with amyloid PET and structural MRI: the A4 study. *Neurology*. 2022;98(13):e1327-e1336.

47. Baker JE, Pietrzak RH, Laws SM, et al. Visual paired associate learning deficits associated with elevated beta-amyloid in cognitively normal older adults. *Neuropsychology*. 2019;33(7):964-974.

48. Duke Han S, Nguyen CP, Stricker NH, Nation DA. Detectable neuropsychological differences in early preclinical Alzheimer's disease: a meta-analysis. *Neuropsychol Rev*. 2017;27:305-325.

49. Norton DJ, Parra MA, Sperling RA, et al. Visual short-term memory relates to tau and amyloid burdens in preclinical autosomal dominant Alzheimer's disease. *Alzheimers Res Ther*. 2020;12(1):1-11.

50. Jutten RJ, Rentz DM, Amariglio RE, et al. Fluctuations in reaction time performance as a marker of incipient amyloid-related cognitive decline in clinically unimpaired older adults. *Alzheimers Dement*. 2022;18(S7):e066578.

51. Jutten RJ, Grandoit E, Foldi NS, et al. Lower practice effects as a marker of cognitive performance and dementia risk: a literature review. *Alzheimers Dement*. 2020;12(1):e12055.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.