SHORT REPORT

A highly feasible, reliable, and fully remote protocol for mobile app-based cognitive assessment in cognitively healthy older adults

Louisa I. Thompson[1]　│　Karra D. Harrington[2]　│　Nelson Roque[3]　│　Jennifer Strenger[1]　│　Stephen Correia[1]　│　Richard N. Jones[1]　│　Stephen Salloway[1]　│　Martin J. Sliwinski[2,4]

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

[2] Center for Healthy Aging, Penn State University, University Park, Pennsylvania, USA

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

[4] Department of Human Development and Family Studies, Penn State University, University Park, Pennsylvania, USA

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

Funding information
Brown University and National Institute on Aging, Grant/Award Number: T32 AG049676; Alzheimer's Association Grant AACSF-20-685786 to Brown University (Thompson, PI)

Abstract

Introduction: The early detection of cognitive impairment is one of the most important challenges in Alzheimer's disease (AD) research. The use of brief, short-term repeated test sessions via mobile app has demonstrated similar or better reliability and validity compared to standard in-clinic assessments in adult samples. The present study examined adherence, acceptability, and reliability for a remote, app-based cognitive screening protocol in healthy older adults.

Methods: Cognitively unimpaired older adults ($N = 52$, ages 60–80) completed three brief cognitive testing sessions per day within morning, afternoon, and evening time windows, for 8 consecutive days using a mobile app-based cognitive testing platform. Cognitive tasks assessed visual working memory, processing speed, and episodic memory.

Results: Participants completed an average of 93% (M = 22.3 sessions, standard deviation = 10.2) of the 24 assigned sessions within 8 to 9 days. Average daily adherence ranged from 95% of sessions completed on day 2 to 88% of sessions completed on day 8. There was a statistically significant effect of session time on adherence between the morning and afternoon sessions only $F(1, 51) = 9.15, P = .004, \eta_p^2 = 0.152$, with fewer afternoon sessions completed on average. The within-person reliabilities of average scores, aggregated across all 24 sessions, were exceptionally high, ranging from 0.89 to 0.97. Performance on the episodic memory task was positively and significantly associated with total score and word list recall score on the Telephone Interview for Cognitive Status. In an exit survey, 65% of participants reported that they "definitely" would complete the sessions again.

Discussion: These findings suggests that remote, mobile app–based cognitive testing in short bursts is both highly feasible and reliable in a motivated sample of cognitively normal older adults. Limitations include the limited diversity and generalizability of the

sample; this was a largely White, highly educated, and motivated sample self-selected for AD research.

1 | INTRODUCTION

The early detection of cognitive impairment is one of the most important challenges in Alzheimer's disease (AD) research.[1] The SARS-CoV-2 (COVID-19) pandemic has highlighted potential advantages of remote methods of cognitive assessment.[2] App-based cognitive testing on mobile devices is rapidly becoming more feasible. The use of brief, short-term repeated test sessions via mobile app has demonstrated similar or better reliability and validity compared to standard in-clinic assessments in adult samples.[3,4] Conducted during the 2020–2021 COVID-19 pandemic, the present study examined adherence, acceptability, and reliability for a remote, app-based cognitive screening protocol in healthy older adults. Our approach involved three brief assessments administered multiple times per day across several days.

2 | METHODS

2.1 | Procedures

Participants were the initial 52 cognitively unimpaired adults ages 61 to 80 (71% female, 86% White) recruited for an ongoing study assessing novel cognitive screening methods. We recruited from the Butler Alzheimer's Prevention Registry (BAPR); a local database of older adults interested in AD research. A majority of BAPR registrants in this study had previously participated in AD research at our site. One hundred forty individuals were invited to the study via e-mail or phone call, and 59 consented and completed online screening. Seven were excluded during screening. There were no dropouts (see enrollment diagram in supporting information).

Screening was conducted by online survey and the modified Telephone Interview for Cognitive Status (TICSm).[5] Unimpaired cognition was defined as a TICSm cutoff score of $\geq$34.[6] Prior smartphone experience was required for enrollment. Android smartphones were shipped to participants with detailed use instructions.

Cognitive tasks were completed for 8 consecutive days using the Mobile Monitoring of Cognitive Change (M2C2), a mobile app–based cognitive testing platform developed as part of the National Institute on Aging's Mobile Toolbox initiative.[3] On each of the 8 days, participants completed brief (i.e., 3–4 minute) M2C2 sessions within morning, afternoon, and evening time windows. Participants chose among three possible time windows (set by staff) for the morning and afternoon sessions at the start of the study. Extra sessions (optional or make-up) could be completed on day 9. Staff provided subsequent support by phone or e-mail as needed. Participants returned the study phones in a prepaid envelope and completed an exit survey online to provide feedback about the study. A $20 gift card compensation was provided on the condition that the study phone was returned. The project was approved by the Butler Hospital Institutional Review Board and all participants gave consent.

2.2 | Cognitive tasks

We selected M2C2 cognitive measures of visual working memory (WM), processing speed (PS), and episodic memory (EM) with prior theoretical and empirical support, including evidence of sensitivity to age and/or age-related neuropathology. Each task took approximately 60 seconds to complete. The Shopping-List task (EM) is a delayed forced-choice recognition task, in which participants incidentally encode grocery item price pairs for later recall while judging whether the item's price is "good."[7,8] Performance is summarized as the proportion of correct responses on recall trials. The Color Shapes task (WM) is a visual array change detection task, measuring intra-item feature binding, in which participants determine whether shapes change color across two sequential presentations in which the shape locations change.[9] Performance is summarized with the hit rate (proportion of correct identifications) and signal detection calculations. Specifically, the discriminability index (d') is the difference in the inverse cumulative standard normal distribution value for the hit rate (proportion of correct identifications) and false-alarm rate (proportion of misidentified stimuli).[10] The Symbol Match (PS) task is a speeded continuous performance task of conjunctive feature search, in which participants are asked to identify matching symbol pairs.[3,11] Performance is summarized with reaction time to complete the task (milliseconds).

3 | RESULTS

3.1 | Adherence

Participants completed an average of 93% (M = 22.3 sessions, standard deviation = 10.2) of the 24 assigned sessions within 8 to 9 days. A minority (23/52) completed all 24 sessions, while a majority (46/52) completed at least 20. We detected no differences in adherence based on age, education, sex, or race/ethnicity. Average daily adherence ranged from 95% of sessions completed on day 2 to 88% of sessions completed on day 8 (Figure 1A). Twenty-one participants completed at

least one optional additional session on day 9. On average, participants completed fewer afternoon than morning sessions ($F[1, 51] = 9.15$, $P = .004$, $\eta_p^2 = 0.152$).

3.2 | Acceptability

In an exit survey, 65% of participants reported that they "definitely" would complete the sessions again, while only 4% said that they "definitely would not." Interestingly, 80% reported that they would at least "somewhat" consider using this app as part of an annual primary care cognitive screen. About 80% reported wanting to know their performance results. About 17% reported having to effortfully motivate themselves to complete the sessions. Only 4% said that it was difficult to complete the sessions during the time windows. Most participants (89%) reported that it was "very easy" to use the app and navigate the tasks. On average, participants ranked the Shopping List task as the most challenging and least favorite task.

3.3 | Reliability

To evaluate test stability, we examined between- and within-person variance in scores. Intraclass correlations (ICCs) were computed by fitting unconditional multilevel mixed models using restricted maximum likelihood to each of the M2C2 tasks. The ICCs were 0.26 for the Shopping List task, 0.34 for the Color Shapes task, and 0.54 for the Symbol Match task (Table 1) indicating that between 26% and 54% of the across-task performance variance was between-persons. The within-person reliabilities of average scores, aggregated across all 24 sessions, were exceptionally high: 0.89 for Shopping List, 0.93 for Color Shapes, and 0.97 for Symbol Match. Reliabilities of average M2C2 scores based on 9, 15, and 21 observations are also reported in Table 1.

3.4 | Validity

To assess convergent validity with an existing cognitive screening tool, we looked at associations between the TICSm and performance on each mobile cognitive task averaged across sessions. Proportion of correct responses on the Shopping List task recall was positively and significantly correlated with TICSm total score (TICSm-TS; r[50] = 0.42, $P = .002$) and TICSm total list recall (TICSm-LR; r[50] = 0.39, $P = .005$). Color shapes beta metric (probability of a hit) was positively and significantly associated with TICSm-LR (r[50] = 0.29, $P = .036$), but not with TICSm-TS. Neither TICSm-TS nor TICSm-LR were significantly correlated with Color Shapes d' prime metric (r[50] = 0.25 and 0.18, respectively, $P > .07$) or any of the Symbol Search metrics (e.g., median response time, (r[50] = −0.15 and −0.09, respectively, $P > .30$). Average correct response rate on the Color Shapes task improved gradually over time, while average correct response rate on the Shopping List task showed minimal variability, with 70% to 80% mean accuracy across all days (Figures 1B and 1C). On the Symbol Match task, reac-

RESEARCH IN CONTEXT

1. **Systematic Review**: Prior work suggests that using a mobile app–based approach to collect repeated data on cognitive functioning in real-world settings is both reliable and valid compared to traditional cognitive screening approaches. However, the feasibility and reliability of this approach in older adults, including those with cerebral amyloidosis, is not yet widely studied.

2. **Interpretation**: Our findings suggests that remote, mobile app–based cognitive testing in short bursts is both highly feasible and reliable in a motivated sample of cognitively normal older adults. Our findings are consistent with prior literature using similar mobile cognitive assessment approaches in younger samples.

3. **Future Directions**: The article provides insights about a novel direction for cognitive assessment with older adults in the context of the COVID-19 pandemic, which has posed challenges for routine in-person care for older adults. Next steps include further understanding: (1) the effectiveness of remote cognitive screening approaches in diverse older adult samples, (2) associations between app-based cognitive tasks and standard in-person tests, (3) how to optimize the sensitivity of mobile cognitive screening tasks to predict neuropathological changes and the emergence of clinical symptoms of AD.

tion times improved successively over the first 4 to 5 days before leveling off (Figure 1D).

4 | DISCUSSION

Adherence to this 8- to 9-day, fully remote, mobile app–based cognitive testing protocol was very good, with 93% of assigned sessions completed. Overall, adherence was lower for afternoon sessions relative to morning and evening sessions. The reason for this is unclear, but participants' afternoon schedules and routines may have been more variable, increasing the likelihood of a time conflict or a forgotten session. The protocol was largely well tolerated by participants, with a majority reporting that they would consider completing the tasks again for research or as part of their own clinical care. The Shopping List task was rated as the least enjoyable measure, and on average, performance on this task did not improve over time. In contrast, performance on the Color Shapes and Symbol Match tasks improved gradually over the first 5 days of the study. Each task showed excellent within-person reliability. Moreover, measures of episodic memory (Shopping List) and visual working memory (Color Shapes) were associated with the TICSm, an established telephone-based screening measure. This result likely reflects shared variance in episodic memory between the TICSm
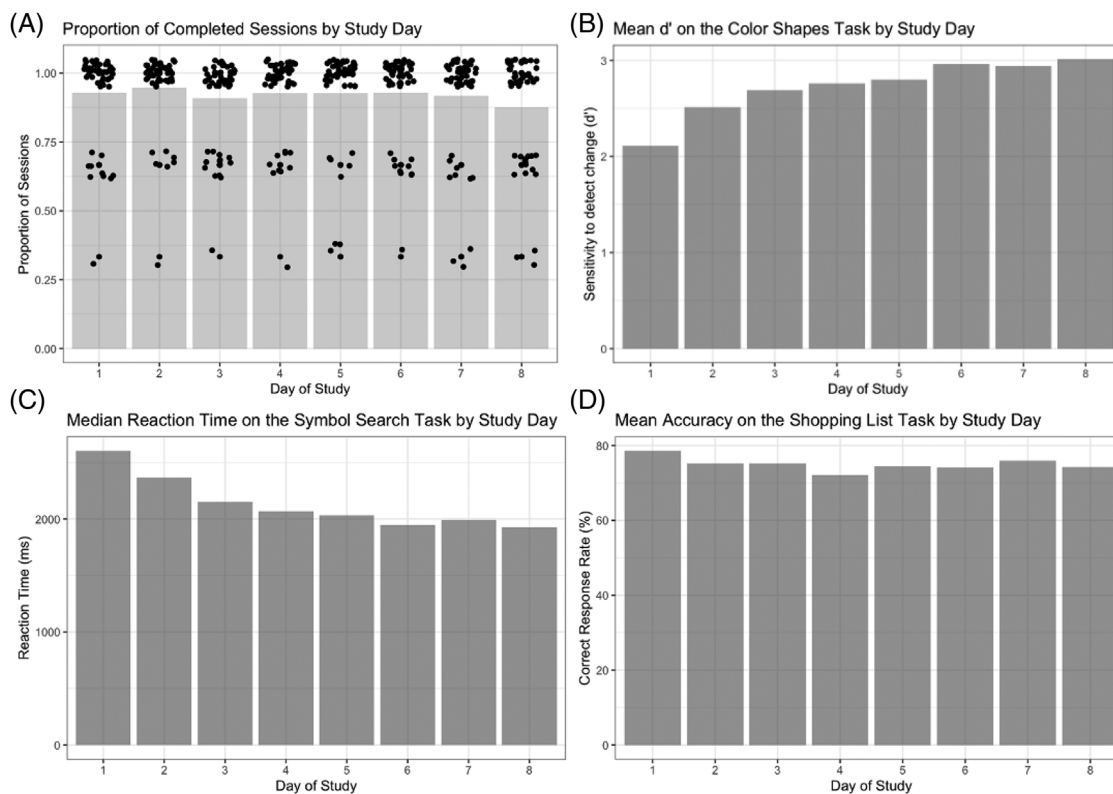
FIGURE 1 Overall session adherence rate and average performance on Mobile Monitoring of Cognitive Change (M2C2) tasks by study day. A, Average proportion of completed M2C2 sessions by study day. B, Average discriminability (d' = d prime) accuracy on the Color Shapes task. C, Average reaction time in milliseconds on the Symbol Search task by study day. D, Mean accuracy on the Shopping List task by study day

TABLE 1 Reliabilities for individual and aggregated cognitive test scores

	Symbol Search	Color Shapes	Shopping List
Between-person variance	166490.89	0.480	0.007
Within-person variance	143865.70	0.923	0.021
Reliability of one occasion (ICC)	0.534	0.342	0.261
Reliability of average (24 sessions)	0.965	0.926	0.894
Reliability of average (21 sessions)	0.881	0.916	0.961
Reliability of average (15 session)	0.841	0.886	0.946
Reliability of average (9 sessions)	0.761	0.824	0.912

Abbreviation: ICC, intraclass correlation.

and the Shopping List and Color Shapes tests. The Shopping List task demonstrated the least practice effects of the three M2C2 tasks, which might reflect the use of different stimuli in each session to avoid practice effects. Future work will retain some item consistency across sessions to permit examination of practice effects as a possible marker of pre-clinical AD.

These results support the feasibility of remote app–based cognitive testing. It is notable that this study showed considerably higher assessment adherence than rates typically reported in the extant digital health assessment literature.[12] The reasons for our high adherence rate may be multifactorial: (1) our BAPR sample was self-selected and motivated to contribute to AD research; (2) many had previously and adherently participated in research at our site; (3) we provided detailed verbal, written, and visual aids for the M2C2 app, and (4) session time windows were tailored to participants' schedules as much as possible.

The limited diversity or our largely White and highly educated sample might limit generalizability of the findings. Additionally, it is unclear how the contextual/historical considerations related to the COVID-19 pandemic enhanced or hampered the generalizability of our method and results.

CONFLICTS OF INTEREST

Stephen Salloway has been study site investigator for the aducanumab trials; received consulting fees from Bolden Therapeutics, Ono, Genentech, Biogen, Prothena, Alnylam, ATRI, Roche, and Mayo; received payment/honoraria for Grand Rounds lecture at VA Providence, CME Talk for GME, CME Talk for PlatformQ, CME Talk for WebMD Medscape, CME Talk for PER, Biogen Master Class Series, from Biogen for RSNA Symposium, AD Comm Adv Board, Webinar for SNMMI, Genentech - Adv Board, ATRI-IMPACT AD seminar, and CTAD. He received travel payment from Acumen - Sci Board, AMGEN, AAIC, GEMVAX, and AVID. All other authors no conflicts of interest.

REFERENCES

1. National plan to address Alzheimer's disease: 2019 update, ASPE. 2019: https://aspe.hhs.gov/report/national-plan-address-alzheimers-disease-2019-update. [Accessed October 15 2021].
2. Owens AP, Ballard C, Beigi M, et al. Implementing remote memory clinics to enhance clinical care during and after COVID-19. *Front Psychiatry*. 2020;11:990.
3. 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.
4. Hassenstab J, Aschenbrenner AJ, Balota DA, Mcdade E, Lim YY, Fagan AM, et al. Remote cognitive assessment approaches in the Dominantly Inherited Alzheimer Network (DIAN). Alzheimer's & Dementia 2020;16. doi:10.1002/alz.038144.
5. Brandt J, Spencer M, Folstein M. The telephone interview for cognitive status. *Neuropsychiatry, Neuropsychology and Behavioral Neurology*. 1988;1 (2):111-117.
6. 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.
7. 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.
8. 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.
9. Parra MA, Abrahams S, Logie RH, Méndez LG, Lopera F, Sala SD. Visual short-term memory binding deficits in familial Alzheimer's disease. *Brain*. 2010;133(9):2702-2713.
10. Stanislaw H, Todorov N. Calculation of signal detection theory measures. *Behav Res Methods Instrum Comput*. 1999;31(1):137-149.
11. Deary IJ, Johnson W, Starr JM. Are processing speed tasks biomarkers of cognitive aging? *Psychol Aging*. 2010;25(1):219-228.
12. Pratap A, Neto EC, Snyder P, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. *NPJ Digit Med*. 2020;3(1):21.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.