










## Research Article

# Psychometric and adherence considerations for high-frequency, smartphone-based cognitive screening protocols in older adults

Louisa I. Thompson<sup>1</sup> , Alyssa N. De Vito<sup>1</sup> , Zachary J. Kunicki<sup>1</sup> , Sheina Emrani<sup>1</sup> , Jennifer Strenger<sup>1,2</sup>,  
Caroline Nester<sup>3</sup>, Karra D. Harrington<sup>4</sup>, Nelson Roque<sup>5</sup>, Masood Manoocheri<sup>1</sup> , Stephen Salloway<sup>1</sup>,  
Stephen Correia<sup>7</sup> , Richard N. Jones<sup>1</sup>  and Martin J. Sliwinski<sup>4,6</sup>

<sup>1</sup>Department of Psychiatry & Human Behavior, Alpert Medical School, Brown University, Providence, RI, USA, <sup>2</sup>Memory & Aging Program, Butler Hospital, Providence, RI, USA, <sup>3</sup>Department of Clinical and Health Psychology, University of Florida, Gainesville, FL, USA, <sup>4</sup>Center for Healthy Aging, Penn State University, University Park, PA, USA, <sup>5</sup>Department of Psychology, University of Central Florida, Orlando, FL, USA, <sup>6</sup>Department of Human Development & Family Studies, Penn State University, University Park, PA, USA and <sup>7</sup>Department of Health Promotion and Behavior, School of Public Health, University of Georgia, FL, USA

### Abstract

**Objective:** The psychometric rigor of unsupervised, smartphone-based assessments and factors that impact remote protocol engagement is critical to evaluate prior to the use of such methods in clinical contexts. We evaluated the validity of a high-frequency, smartphone-based cognitive assessment protocol, including examining convergence and divergence with standard cognitive tests, and investigating factors that may impact adherence and performance (i.e., time of day and anticipated receipt of feedback vs. no feedback). **Methods:** Cognitively unimpaired participants (N = 120, M<sub>age</sub> = 68.8, 68.3% female, 87% White, M<sub>education</sub> = 16.5 years) completed 8 consecutive days of the Mobile Monitoring of Cognitive Change (M2C2), a mobile app-based testing platform, with brief morning, afternoon, and evening sessions. Tasks included measures of working memory, processing speed, and episodic memory. Traditional neuropsychological assessments included measures from the Preclinical Alzheimer's Cognitive Composite battery. **Results:** Findings showed overall high compliance (89.3%) across M2C2 sessions. Average compliance by time of day ranged from 90.2% for morning sessions, to 77.9% for afternoon sessions, and 84.4% for evening sessions. There was evidence of faster reaction time and among participants who expected to receive performance feedback. We observed excellent convergent and divergent validity in our comparison of M2C2 tasks and traditional neuropsychological assessments. **Conclusions:** This study supports the validity and reliability of self-administered, high-frequency cognitive assessment via smartphones in older adults. Insights into factors affecting adherence, performance, and protocol implementation are discussed.

**Keywords:** Cognitive screening; repeat assessment; Alzheimer's disease; smartphone; mobile app; feasibility

(Received 13 February 2024; final revision 30 June 2024; accepted 24 July 2024; First Published online 20 September 2024)

### Introduction

Alzheimer's disease and Alzheimer's disease-related dementias (AD/ADRDs) currently affect over 55 million people globally with numbers projected to increase to 139 million by 2050. Widespread early detection and diagnosis during the early stages of the disease is seen as increasingly critical to maximize the effectiveness of intervention and treatment. Despite this, most individuals are currently diagnosed after the onset of dementia and only half ever receive a diagnosis from a clinician (Amjad et al., 2018; Hampel et al., 2022). Challenges in assessment and diagnosis may stem from several sources, such as a limited availability of qualified AD/ADRD specialists, time-, cost-, and distance-related barriers for patients, and limitations of existing traditional neuropsychological

tests to detect subtle cognitive decline. These factors present significant obstacles to early detection and disease monitoring of AD/ADRD in community settings and clinical trials investigating novel therapeutic agents.

New methods of cognitive assessment are urgently needed to keep pace with rapidly evolving biomarkers that are now being used for early detection of AD. The use of digital technologies to assess cognition may help to overcome many of the limitations of traditional paper-and-pencil testing and can also help to support the transition toward partially remote or "decentralized" AD/ADRD clinical trials which have potential to increase accessibility and reduce participant burden (Leroy et al., 2023). Remote cognitive assessment using smartphones has several potential

**Corresponding author:** Louisa I. Thompson; Email: [louisa\\_thompson@brown.edu](mailto:louisa_thompson@brown.edu)

**Cite this article:** Thompson L.I., De Vito A.N., Kunicki Z.J., Emrani S., Strenger J., Nester C., Harrington K.D., Roque N., Manoocheri M., Salloway S., Correia S., Jones R.N., & Sliwinski M.J. (2024) Psychometric and adherence considerations for high-frequency, smartphone-based cognitive screening protocols in older adults. *Journal of the International Neuropsychological Society*, 30: 785–793, <https://doi.org/10.1017/S1355617724000328>

© The Author(s), 2024. Published by Cambridge University Press on behalf of International Neuropsychological Society. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

advantages over traditional assessment methods including increasing accessibility, improving ecological validity, and enabling high-frequency assessment (HFA) (Öhman *et al.*, 2021). Smartphone-based cognitive assessment platforms have the ability to track environmental, health, and behavioral factors that may impact the validity of cognitive performance (Emert *et al.*, 2023; Scott *et al.*, 2015; Wilks *et al.*, 2021). HFA can improve the reliability and validity of cognitive assessment through averaging performance across multiple trials (Sliwinski *et al.*, 2018). HFA also allows researchers to track person-specific changes and patterns that may signal decline (e.g., increased cognitive variability, poor learning curves); an approach consistent with personalized medicine and one that may also reduce the time needed to detect therapeutic efficacy and/or reach clinical trial endpoints (Dodge *et al.*, 2015; Weizenbaum *et al.*, 2023).

Recent research suggests that mobile cognitive assessments are acceptable and feasible to use in older adult populations (Koo & Vizer, 2019; Nicosia *et al.*, 2023; Papp *et al.*, 2021; Thompson *et al.*, 2022). However, research on the reliability and validity of unsupervised mobile cognitive assessments is still evolving. Preliminary work from Thompson *et al.*, and others suggests that Mobile Monitoring of Cognitive Change (M2C2), a set of smartphone assessments developed for the National Institute on Aging's Mobile Toolbox initiative, demonstrates comparable or better reliability (.89 or higher) than traditional neuropsychological assessments (Sliwinski *et al.*, 2018; Thompson *et al.*, 2022). For example, reliabilities for the Montreal Cognitive Assessment (MoCA) can range from .50 to .71, while Mini Mental Status reliability can be as low as .35 (Bernstein *et al.*, 2011; Spencer *et al.*, 2013). Regarding validity, prior work by Thompson *et al.*, demonstrated that smartphone-based cognitive tests were superior to the MoCA in the detection of cerebral amyloid status (Thompson *et al.*, 2023). Additionally, Nicosia *et al.* (2023) demonstrated associations between smartphone-based cognitive tests with 1) traditional neuropsychological testing, 2) amyloid and tau positron emission tomography, 3) cerebrospinal fluid markers of amyloid beta (A $\beta$ ) 40, A $\beta$ 42, total tau, and tau phosphorylated at threonine 181, and 4) cortical thickness in AD-related regions of interest (Nicosia *et al.*, 2023).

Despite these promising initial findings, remaining challenges need to be addressed before smartphone-based cognitive assessments can be widely implemented in clinical and research settings. Successful implementation of unsupervised, smartphone-based cognitive assessment requires adequate protocol adherence and engagement. Attrition can be a common problem seen in remote assessment studies, which may have important implications if smartphone-based testing protocols will be used for longitudinal disease monitoring (Pratap *et al.*, 2020). Additionally, it is critical to demonstrate convergence with existing reference standard neuropsychological tests, particularly measures sensitive to AD/DRD neuropathology, such as the Free and Cued Selective Reminding task and the Preclinical Alzheimer's Composite Index, which are commonly used for early-stage AD detection and monitoring (Papp *et al.*, 2017; Schindler *et al.*, 2017). Therefore, it is critical to investigate validity considerations that may influence the implementation of smartphone-testing protocols for both clinical care and research.

The present analysis examined two novel hypotheses to address the validity of smartphone-based cognitive assessments in a sample of 120 cognitively unimpaired older adults. First, we hypothesized that the smartphone tasks would demonstrate convergence with same-domain standard neuropsychological tests, as well as

divergence with different-domain tests. Second, we evaluated two factors hypothesized to affect adherence and performance on remote HFA, time of day and anticipation of feedback.

Specifically, we hypothesized that better adherence and performance would be seen for morning vs. evening sessions. This was based on prior evidence of cognitive "sundowning" captured on remote assessments, with better performance seen in the morning (Wilks *et al.*, 2021), as well as the broader aging literature, which indicates an increased likelihood of time-specific cognitive variability with both age and neurodegenerative pathology (Anderson *et al.*, 2014; Musiek *et al.*, 2018).

Additionally, we hypothesized that anticipation of feedback on cognitive test performance may increase engagement and performance on remote tests. Research has shown that feedback impacts task performance and motivation, and is therefore typically not provided during controlled testing in neuropsychological practice (Clark *et al.*, 2024; Di Rosa *et al.*, 2015). However, providing patient feedback after testing is routine, as is providing feedback in neurorehabilitation programs and commercial brain game apps to maintain engagement and motivation ((Burgers *et al.*, 2015; van Dokkum *et al.*, 2015). One way to potentially impart these benefits of feedback without disrupting controlled, real-time testing conditions is to tell participants what they can expect to gain from the assessment (i.e., to let them know they will receive feedback after testing is complete). To our knowledge, ours is the first study to examine how the anticipation of receiving feedback about one's performance might relate to subsequent engagement and performance on self-administered cognitive tasks by randomly assigning participants to either anticipated or surprise feedback conditions.

Finally, we provide a full sample update on the within-subject reliabilities for M2C2 HFA, originally reported in the study's preliminary ( $n = 52$ ) reliability findings published in 2022.

## Methods

### Participants and recruitment

Participants were cognitively unimpaired older adults, between the ages of 60 and 80, recruited from the Butler Alzheimer's Prevention Registry, a local database of older adults interested in AD research at the Butler Hospital Memory and Aging Program. A total of 256 individuals were invited to the study through email or phone call, and 146 consented and completed an online screening. Twenty-three participants were excluded during the screening process, and three participants withdrew after enrollment, for a final sample size of 120. Please refer to Thompson *et al.*, 2023 for detailed enrollment data and inclusion and exclusion criteria. Familiarity with smartphones (defined by a minimum 1-year history of use) was required for enrollment. Screening was conducted using an online survey and the modified Telephone Interview for Cognitive Status (TICS<sub>m</sub>) (Brandt *et al.*, 1988; Cook *et al.*, 2009). Unimpaired cognition was defined as a TICS<sub>m</sub> cutoff score of  $\geq 34$  (Cook *et al.*, 2009). Participants completed an exit survey online to provide feedback at the end of the study, and a \$20 gift card compensation was provided. All participants were made aware of this compensation during the consent process. The project received approval from the Butler Hospital Institutional Review Board, and all participants provided consent. The research was completed in accordance with Helsinki Declaration.

### Remote cognitive assessment

Remote cognitive tasks were completed using 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 and described previously (Thompson et al., 2022; Thompson et al., 2023). Android smartphones preloaded with the cognitive assessment app were mailed to participants along with a detailed use guide. The phones were locked down to prevent the use of other features such as web browsing and the camera. Participants completed brief (i.e., 3–4 minutes) M2C2 sessions each day for eight consecutive days, within morning, afternoon, and evening time blocks. Participants received a push notification on their phone when it was time to complete a session and every 30 minutes thereafter for the duration of the session (or until completed). Session start times were set at random, but always began within fixed 1-hour windows which were chosen at the start of the study with input from the participant to fit their schedules (Thompson et al., 2022, 2023). If participants received a notification when they were busy, they could complete the session later - anytime until the close of the 2-hour session window. Additional sessions could be completed on day 9 as optional or make-up sessions. Staff provided support through phone or email as needed, as described in previously (Thompson et al., 2023). During each M2C2 session participants completed three previously characterized cognitive tasks assessing episodic memory (Prices), visual working memory (Color Shapes), and processing speed (Symbol Match) (Sliwinski et al., 2018; Thompson et al., 2023). Each task took approximately 60 s to complete. The Prices task is a delayed forced-choice recognition paradigm (Gallo et al., 2006; Naveh-Benjamin, 2000). Ten grocery item-price pairs are encoded followed by immediate recall recognition trials. Performance was measured by the proportion of correct responses on ten recall trials. The Color Shapes task is a visual change detection task measuring intra-item feature binding (Parra et al., 2010; Parra et al., 2011). Performance was measured by discriminability ( $d'$ ) performance calculated from the proportion of correct identifications and proportion of misidentified stimuli (Stanislaw & Todorov, 1999). The Symbol Match task is a speeded continuous performance task in which participants are asked to identify matching symbol pairs (Hassenstab et al., 2020; Sliwinski et al., 2018). Performance was measured by the median reaction time to complete the task across all trials in milliseconds (Sliwinski et al., 2018). There were no performance validity measures included in the M2C2 protocol.

### Feedback manipulation

We investigated whether a simple experimental manipulation – telling participants that they would or would not learn about their test results – was associated with adherence or performance on the remote cognitive assessment protocol. Participants were randomly assigned to one of two conditions, feedback or no feedback, and were told that they either would or would not receive performance feedback upon study completion, respectively. At the end of the study, all participants were given the option to receive feedback, regardless of condition. Feedback was given in the form of a pdf summary of raw scores on all three digital assessments plotted over the course of the 8-day test period. No benchmarks (e.g., age specific normative data) were provided. Feedback was presented as for the participants' curiosity only, and was labeled as not for non-clinical use. Participants who obtained low scores on the MoCA during the in-person study visit and had cognitive concerns were

advised to follow-up with their primary care provider and given referral information for neuropsychological evaluation.

### Protocol phases

Our study was launched during the COVID-19 pandemic and all study procedures were completed remotely (i.e., screening, consent, and M2C2 orientation). After COVID-19 restrictions were lifted, we were able to add an in-person study visit, but the main study protocol remained fully remote and unchanged. Prior participants and newly enrolled participants were all invited to complete the in-person study visit. Forty of the 52 prior participants came in for the optional in-person visit approximately 17 months after finishing the fully remote study. The remainder of the sample enrolled after pandemic restrictions were lifted and completed the in-person study visit within approximately one month of the remote protocol. To account for this variation, a variable for protocol type (short versus long delay) was included in the statistical models. A dichotomous variable was used given the bimodal distribution of the time delay (clustered around 1 month and 17 months) as well as for ease of interpretation.

### In-person cognitive assessment

Participants were scheduled for a single in-person study visit to complete standard paper-and-pencil neuropsychological assessments. The assessment battery consisted of widely used and thoroughly validated measures previously used in the Preclinical Alzheimer's Cognitive Composite battery and shown to be sensitive to detect prodromal AD (Donohue et al., 2014). Individual tests included the Wechsler Adult Intelligence Scale Revised (WAIS-R) Digit Symbol Substitution (DSST) and Digit Span subtests (David, 1981), the Wechsler Memory Scale Revised (WMS-R) Logical Memory Immediate and Delayed Recall subtest (Wechsler, 1987), Trail Making Test A & B (Reitan & Wolfson, 1985), Category Fluency Test (animals) (Martin & Fedio, 1983), and the Free and Cued Selective Reminding Test – Immediate Recall (FCSRT-IR) (Grober & Buschke, 1987; Grober et al., 2018). The Wechsler Test of Adult Reading (WTAR) word reading subtest was used as an estimate of premorbid verbal intellectual functioning (Wechsler, 2001). The testing took approximately 60–90 minutes to complete.

### Analysis

M2C2 adherence was operationalized by a compliance cutoff of 80%, which was examined overall and by time of day (morning, afternoon, evening). Compliance was defined by the number of sessions a participant initiated. Incomplete sessions were counted toward compliance, as long as participants had completed at least one of the three tasks. A Cochran's Q-test was used to test for group differences in completion rates between time of day (Cochran, 1950). McNemar's test was used for post-hoc analyses. Pearson's correlations were used to examine associations between M2C2 task performance and demographic variables. Linear regression analyses adjusted for age, sex, and education were used to evaluate significant differences between the feedback and control conditions on protocol adherence and performance on M2C2 tasks. Linear regression models were constructed to evaluate convergent and divergent validity between M2C2 tasks and the standard neuropsychological assessments. Models were adjusted for feedback condition and protocol type. We utilized the Benjamini-Yekutieli method to adjust for multiple comparisons (Benjamini &

**Table 1.** Sample characteristics ( $N = 120$ )

Characteristic	Mean or $n$	SD or %
<b>Race</b>		
American Indian or Alaska Native	1	0.8%
Asian	1	0.8%
Black or African American	4	3.3%
More than one race	6	4.9%
White	106	87%
Unknown	2	1.6%
<b>Ethnicity</b>		
Hispanic or Latino	4	3.3%
Non-Hispanic or Latino	107	89.2%
Unknown	9	7.5%
<b>Sex</b>		
Female	82	68.3%
Male	38	31.6%
Age (years)	68.9	4.9
Education (years)	16.5	2.4
TICS <sub>m</sub> (total score)	39.3	3.3

Note: TICS<sub>m</sub> = Modified Telephone Interview for Cognitive Status, MoCA = Montreal Cognitive Assessment.

Yekutieli, 2005). To evaluate test stability, we examined between- and within-person variance (standard deviations) in scores. Intraclass correlations (ICCs) were computed by fitting unconditional multilevel mixed models using restricted maximum likelihood to each of the M2C2 tasks, as previously reported (Sliwinski *et al.*, 2018). With this approach, the ICCs indicate the expected correlation between two randomly sampled measurements from the same person. The ICCs we report are the reliabilities for the aggregate scores based on the number of sessions completed. We have presented reliabilities for a range of sessions including the median number of sessions, as not all participants complete 100% of the sessions (24 sessions).

## Results

### Sample characteristics

The sample consisted of 120 participants with a mean age of 68.9 years ( $SD = 4.9$ ) and a mean education of 16.5 years ( $SD = 2.4$ ). The sample was 68.3% female, and 87% White (Table 1). There were no significant demographic differences between participants on the basis of protocol type, with the exception of age, which was somewhat younger in the short delay group (mean age 67.5) relative to the long delay (mean age 70.5). There were no significant demographic differences on the basis of feedback condition. Performance on all three M2C2 tasks was associated with age (Table 2). Performance on the Prices and Color Shapes tasks was associated with sex. There were no association between M2C2 performance and education.

### Effects of day and time of day on adherence and performance

We examined the overall completion rate across all 24 assigned M2C2 test sessions and by time of day. Compliance was defined as a completion rate of 80% or higher. Overall, 89.3% of participants met compliance across the 24 assigned sessions within 8 days. Average compliance by time of day ranged from 90.2% for morning sessions, to 77.9% for afternoon sessions, and 84.4% for evening sessions. A Cochran's Q-test revealed differences in completion rates between time of day,  $Q(2) = 10.56$ ,  $p = .01$ , with significant differences between afternoon and morning ( $z = -3.27$ ,

$p = .003$ ), but not between afternoon and evening ( $p = .11$ ) or evening and morning ( $p = .11$ ). There was only one participant who completed less than two sessions on more than one day (days 1 and 8). Compliance declined the most on day 8 (70.5%), with only 97 participants completing assigned sessions. See supplemental Figure 1 for a summary of overall session adherence rates and average performances on M2C2 tasks by study day. Thirty-five (28.7%) participants completed at least one optional or make-up session on day 9. We detected no differences in adherence based on age, education, or sex.

### Effects of feedback condition on adherence and performance

There was no evidence of differences in overall adherence by feedback condition ( $\beta = 0.05$ ,  $p = .56$ ). There was a difference in adherence by feedback condition on day one, but this difference did not reach the threshold for significance ( $\beta = -0.17$ ,  $p = .06$ ), and was absent on subsequent days. There was an effect of feedback condition on Symbol Match (processing speed) performance on day one ( $\beta = 0.17$ ,  $p = .05$ ) through day eight ( $\beta = 0.26$ ,  $p = .01$ ). There was no effect of feedback condition on Prices (episodic memory) or Color Shapes (working memory) performance. All results reported for day 1 or day 8 were aggregated within the day for analysis.

### Convergent and discriminant validity

To assess convergent validity with the standard tests, we looked at associations between performance on each M2C2 cognitive task averaged across sessions with each same-domain standard neuropsychological test, adjusting for feedback condition and protocol type (Table 2). Proportion of correct responses on the Prices task was positively associated with Logical Memory immediate ( $\beta = 0.22$ ) and delayed recall ( $\beta = 0.31$ ). Prices task was also associated with FCSRT Immediate Recall ( $\beta = 0.24$ ). Median response time on the Symbol Match task was positively associated with Trails A completion time ( $\beta = 0.35$ ) and negatively associated with DSST total score ( $\beta = -0.44$ ) and Verbal Fluency (animals) total score ( $\beta = -0.32$ ). Color Shapes task d-prime (discriminability) was positively associated with DSST total score ( $\beta = 0.33$ ) and negatively associated with Trails B completion time ( $\beta = -0.24$ ), but not significantly associated with longest Digit Span backward ( $\beta = 0.16$ ) (Table 3). There were no main effects of protocol type in any of the models.

To evaluate divergent validity, we checked for associations with different domain standard neuropsychological measures, again adjusting for feedback condition and protocol type (Table 4). Proportion of correct responses on the Prices task recall was not associated with measures of processing speed, working memory, or premorbid estimated verbal intelligence (IQ). The Symbol Matching task was associated with verbal fluency ( $\beta = -0.32$ ), but not with untimed language-dependent tasks, such as verbal memory or premorbid verbal IQ estimate. Colors shapes task was not associated with verbal fluency, psychomotor processing speed or premorbid verbal IQ estimate (Table 4).

### Reliability

To evaluate test stability, we examined between- and within-person variance in performance. These results are stratified by feedback condition and provide an update to our preliminary ( $n = 52$ ) reliability data published in 2022 prior to the study completion. On average, in the full sample ( $N = 120$ ), performance

**Table 2.** Correlations between age, sex, education, and M2C2 variables (95% CI)

Test	Age	Female	Education
M2C2 Prices (Proportion Correct)	-.36 (-.51, -.20)	.22 (.05, .38)	-.03 (-.20, .15)
M2C2 Symbol Matching (Median Response Time)	.42 (.27, .56)	-.08 (-.26, .10)	-.15 (-.32, .03)
M2C2 Color Shapes (d-prime discriminability)	-.31 (-.46, -.14)	.27 (.10, .43)	.02 (-.16, .20)

Note: CI = confidence interval, M2C2 = Mobile Monitoring of Cognitive Change. If the confidence interval does not contain 0, the result is significant at  $p < .05$ .

**Table 3.** Convergent validity

Test or Model	Unstd. Est.	$p$ -value	Std. Est.	95% CI
<b>M2C2 Prices (Proportion Correct)</b>				
Logical Memory (Immediate recall)	0.01	.04*	0.22	0.01, 0.42
Logical memory (Delayed recall)	0.01	.003*	0.31	0.12, 0.50
FCSRT Immediate free recall	0.00	.01*	0.24	0.06, 0.43
<b>M2C2 Symbol Matching (Median Response Time)</b>				
Trails A (Completion time)	17.77	<.001*	0.35	0.18, 0.52
Digit symbol substitution score	-22.35	<.001*	-0.44	-0.61, -0.26
Verbal fluency (animals) score	-30.93	<.001*	-0.32	-0.50, -0.14
<b>M2C2 Color Shapes (d-prime discriminability)</b>				
Digits (Longest backwards span)	0.08	.13	0.16	-0.03, 0.35
Trails B (Completion time)	-0.01	.02	-0.24	-0.42, -0.06
Digit symbol substitution score	0.03	<.001*	0.33	0.15, 0.51

Note: All models adjusted for protocol type and feedback group. Unstd. Est. = Unstandardized estimate, CI = Confidence interval, Std. Est. = Standardized estimate, M2C2 = Mobile Monitoring of Cognitive Change. The standardized estimate can be interpreted as an effect size, where 0.10 = small, 0.30 = medium, and 0.50 = large. If the CI does not contain 0, the result is significant at  $p < .05$ . \* = result was still significant at  $p < .05$  after Benjamini-Yekutieli adjustment for multiple comparisons.

improved over time on the Color Shapes and Symbol Match, while performance remained stable on the Prices task (Sup. Figure 1). The within-person reliabilities of average scores aggregated across 9, 15, 18, 21, and 24 sessions were excellent (Table 5). Additionally, we provide the ICCs aggregated from two randomly selected measurement observations from each participant for each task.

## Discussion

The present study adds to a growing body of support for self-administered, HFA via smartphone as a feasible, acceptable, and reliable approach to measuring cognition in older adults, including those at risk for AD/ADRD. In addition to replicating strong test reliability validity, and adherence rates, our study provides new insights into potential factors that may impact adherence and performance on testing in this context (Cerino et al., 2021; Harrington et al., 2021; Nicosia et al., 2023).

In our examination of the potential effects of anticipated feedback on cognitive performance, we found that participants who expected to eventually receive feedback on their performance had faster reaction times on the M2C2 processing speed task. This

**Table 4.** Divergent validity

Test or Model	Unstd. Est.	$p$ -value	Std. Est.	95% CI
<b>M2C2 Prices (Proportion Correct)</b>				
Trails A (Completion time)	0.00	.11	-0.15	-0.36, 0.06
Digit Span Forward	0.01	.10	0.18	-0.02, 0.38
WTAR Scaled score	0.00	.35	-0.10	-0.30, 0.10
<b>M2C2 Symbol Matching (Median Response Time)</b>				
Verbal fluency (animals) score	-30.93	<.001*	-0.32	-0.50, -0.14
Logical memory delayed recall	-4.54	.66	-0.02	-0.21, 0.16
WTAR Scaled score	-4.90	.43	-0.07	-0.27, 0.12
<b>M2C2 Color Shapes (d-prime discriminability)</b>				
Verbal fluency (animals) score	0.01	.38	0.08	-0.11, 0.28
Trails A (Completion time)	-0.01	.12	-0.15	-0.34, 0.03
WTAR Scaled score	0.00	.64	0.03	-0.15, 0.23

Note: All models adjusted for protocol type and feedback group. Unstd. Est. = Unstandardized estimate, CI = Confidence interval, Std. Est. = Standardized estimate, M2C2 = Mobile Monitoring of Cognitive Change. The standardized estimate can be interpreted as an effect size, where 0.10 = small, 0.30 = medium, and 0.50 = large. \* = result was still significant at  $p < .05$  after Benjamini-Yekutieli adjustment for multiple comparisons.

result may suggest that simply knowing that feedback will be received could improve effortful engagement on digital tasks, as processing speed (i.e., how quickly one performs on a task) may be more influenced by subjective control than other cognitive abilities. Overall, these findings on the effects of feedback anticipation have major implications for understanding how to incentivize adherence and performance on remote testing without compromising test validity. Our results suggest that feedback may not need to be given in real-time in order to potentially increase task engagement. Monetary incentives have also been recently associated with strong adherence in a mixed sample of older adults with normal cognition and mild cognitive impairment (Nicosia et al., 2023), but are not appropriate or scalable for clinical settings. In contrast, providing feedback at the end of an assessment period is clinically appropriate and may also be beneficial for patient awareness and engagement in understanding brain health.

Contrary to our expectations, we did not observe an effect of feedback condition on overall protocol adherence. It should be noted that ceiling effects due to high overall study adherence in this sample may have limited our ability to detect significant differences. Additionally, our experimental manipulation was subtle, focusing on adjusting only the anticipation of feedback so as

**Table 5.** M2C2 task reliabilities stratified by feedback condition

Test	Feedback anticipated						No feedback anticipated					
	ICC	9 sessions	15 sessions	18 sessions	21 sessions	24 sessions	ICC	9 sessions	15 sessions	18 sessions	21 sessions	24 sessions
Prices (mean accuracy)	.25	.75	.83	.86	.88	.89	.28	.78	.85	.88	.89	.90
Color Shapes (hit rate)	.42	.86	.91	.93	.94	.94	.40	.86	.91	.92	.93	.94
Color Shapes (d-prime)	.43	.87	.92	.93	.94	.95	.36	.84	.89	.91	.92	.93
Symbol Match (median RT)	.56	.92	.95	.96	.96	.97	.51	.90	.94	.95	.96	.96

Note: M2C2 = Mobile Monitoring of Cognitive Change, ICC = intraclass correlation. The ICC can be interpreted as the correlation between two randomly selected measurement observations of the same participant. The values in the remaining columns estimate the reliability of the test for a participant who completed 9 sessions versus 15 sessions, 18 sessions, etc. The range of values is provided because not all participants completed all 24 sessions. The median number of completed sessions was 24 and the average number of completed sessions was 23.

to avoid introducing any changes to the M2C2 protocol. Future research could also examine the effects of more direct in-app feedback and/or feedback following each M2C2 task session on adherence; however, doing so would likely introduce unwanted variability in cognitive performance and compromise the reliability of the assessments.

We also examined the timing of assessments and found good adherence at all three time points (morning, afternoon, evening). Adherence was strongest for the morning sessions, and significantly less so for afternoon sessions. These findings add to the recent literature on cognitive sundowning captured using remote HFA (Wilks et al., 2021). Despite the observed compliance rate of almost 90% overall, adherence declined notably on Day 8, suggesting that a shorter assessment period might be better tolerated in future studies. Overall, these findings are consistent with prior reports, which support the feasibility of remote, self-administered smartphone-based approaches to cognitive assessment in older adults with and without cognitive impairment (Harrington et al., 2021; Nicosia et al., 2023; Papp et al., 2021).

We further examined the convergent and divergent validity of M2C2 tasks and found consistent associations with same-domain standard neuropsychological assessments sensitive to cognitive deficits in preclinical and prodromal AD. M2C2 associations with standard assessments held true as small or medium effects even after controlling for variability in time delays between our remote and in-person assessments. Most notable for the early detection of AD, we found small but consistent associations between performance on the Prices episodic memory task and measures that are sensitive to early memory decline, including contextual verbal memory (Logical Memory) and verbal paired associates learning (Free and Cued Selective Reminding Test). We further also examined divergent validity and demonstrated an absence of associations between M2C2 tasks and most standard measures assessing different cognitive domains. One exception was a significant association between Symbol Match and verbal fluency, which may be attributable to their shared sensitivity to processing speed. This work, together with our earlier report demonstrating task sensitivity to AD biomarkers, and that of Nicosia et al. (2023), builds a strong foundation of multimodal validation evidence in support of using HFA approaches in the older adult population (Nicosia et al., 2023; Thompson et al., 2023).

Finally, we also examined the reliability of M2C2 smartphone-based assessments. Consistent with prior research, we found high

within-subject reliabilities among M2C2 measures over time (Nicosia et al., 2023; Sliwinski et al., 2018). One advantage of HFA is the ability to obtain high degrees of within-person reliability, which is difficult to achieve with one or two assessment visits. The high M2C2 task reliabilities observed in this study are consistent with good test-retest reliabilities recently reported for other multi-day digital assessment protocols (Stricker et al., 2023; Weizenbaum et al., 2023).

The results of this study provide insights into ways that HFA protocols could be refined and optimized for future research in older adult samples. First, our results support that testing sessions scheduled for the morning and evening may yield better adherence than testing sessions scheduled in the middle of the day. This may in part reflect variability in people's mid-day schedules due to activities such as work or running errands, which may be distracting or more difficult to interrupt. Assessments may fit better into morning and evenings routines and may be easier for people to complete when they are at home. Additionally, our findings indicate that very good within-subject reliabilities can be achieved on measures of processing speed, working memory, and episodic memory within 15 sessions (i.e., three daily sessions for five days). Taken together with the decline in adherence that we observed in our sample after seven days of testing, a five- or six-day assessment period seems optimal. These refinements and other insights, such as the advantages of providing performance feedback for examinees, can inform future clinical implementation research and applications of HFA in decentralized AD/ADRD clinical trials (Leroy et al., 2023). In the future, remote digital assessments could also be particularly useful in combination with blood-based biomarker detection of AD proteinopathy in clinical settings to help guide decision making when obtaining a full cognitive evaluation might not otherwise be feasible (Ashton et al., 2024).

The results of this study should be interpreted with the nature of the cohort in mind. Participants were cognitively normal individuals exclusively recruited from a registry of those expressing interest in research on AD/ADRD. In this context, it is plausible to expect relatively high performance and motivation to be engaged in a self-administered remote assessment protocol, compared to other contexts, such as enrolling symptomatic patients in community clinics. Whereas the M2C2 app has been primarily tested in asymptomatic at-risk populations, others have recently reported positive feasibility and validity data with its use in larger longitudinal cohorts of older adults, including those with mild cognitive impairment (Cerino et al., 2021; Harrington et al., 2021).

Another important contextual consideration in our sample is the variable time delays between the remote and in-person cognitive testing ranging from one to 18 months due to COVID-19-related research restrictions. Given the potential for the remote and in-person task associations to be moderated by a delay, we controlled for protocol type (short or long delay) in our models and additionally checked and found no significant interactions with protocol type.

An important limitation of this study is the narrow demographic features of the sample (i.e., largely White, female, highly educated, maximum age of 80), which limits generalizability to the broader older adult population. Replication of our protocol's feasibility and the validity of the M2C2 measures in a more diverse older adult sample is necessary prior to implementation in larger research trials and clinical settings. Finally, we did not have data on participant employee status or participant's schedules, which may have yielded further insights into time-of-day variations in M2C2 performance and adherence.

As with most mobile cognitive assessments, M2C2 tasks have not been developed for diagnostic purposes and are probably best suited for initial screening and monitoring of cognitive function. Our M2C2 protocol does not include any performance validity checks, but it does have the advantage of being primarily visual-based, which may make it more difficult for participants to take notes or otherwise 'cheat' on the tasks. Nonetheless, we recognize that there are many other potential threats to performance validity including distractions/interruptions or assistance from other individuals, and these are important limitations of self-administered remote assessments. In order to manage potential variability in device processing capabilities and avoid interruptions from calls, texts, or pop-ups we sent all participants a study-managed Android phone to use for the M2C2 protocol; however, this approach is not feasible for scalable clinical implementation. M2C2 and other mobile assessments are now more readily and easily deployed for use on participant's own devices (including iOS and Android), and this is the current approach in use on our own ongoing research and other studies (Harris et al., 2024).

Traditional neuropsychological testing continues to play an important role in capturing the nature and extent of cognitive deficits and aiding clinicians in the diagnostic process. Such measures are administered in carefully controlled conditions that utilize normative data and allow examiners to observe potentially diagnostically relevant behavioral factors such as process-oriented test performance features, emotional affect and social reciprocity, and other non-test behaviors including neurobehavioral signs that may have diagnostic relevance. However, it is likely that in the future, proxies for some of these non-test observational factors can be met via passive monitoring and artificial intelligence-driven capture of behavior and symptoms to further optimize remote assessment. With the current abundance of research in this space, digital cognitive assessment tools are poised to undergo further refinements and integration with other technologies, and will play an important role in AD/ABCD clinical research.

## Conclusions

In summary, this study significantly supports the feasibility, acceptability, and reliability of self-administered, high-frequency cognitive assessments via smartphones in older adults, particularly those at risk for AD. The findings not only replicate strong test reliability, validity, and adherence but also provide valuable

insights into factors affecting adherence and performance. The anticipation of learning one's test results emerges as a potential motivator, having some impact on both reaction times and initial adherence. The study suggests that delayed feedback may still enhance task engagement, offering a practical alternative to delivering real-time results during remote testing. Moreover, our examination of assessment timing reveals a preference for morning sessions, which should be considered for optimizing future protocols. The reliability and validity of M2C2 smartphone-based assessments, including their consistent associations with same-domain standard neuropsychological measures, suggests that with further refinement, such remote cognitive assessment tools could potentially be scalable for screening and monitoring use in clinical settings.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/S1355617724000328>.

**Acknowledgments.** None.

**Financial support.** This work was supported by Alzheimer's Association grant AACSF-20-685786 (Thompson, PI) and by NIA grant T32 AG049676 to Penn State University.

**Competing interests.** Dr Thompson has been a paid consultant for the Davos Alzheimer's Collaborative. Dr Stephen Salloway has been a paid consultant for Lilly, Biogen, Roche, Genentech, Eisai, Bolden, Amylyx, NovoNordisk, Prothena, Ono, and Alnylam. All other authors have nothing to disclose.

## References

- Amjad, H., Roth, D. L., Sheehan, O. C., Lyketsos, C. G., Wolff, J. L., & Samus, Q. M. (2018). Underdiagnosis of dementia: An observational study of patterns in diagnosis and awareness in US older adults. *Journal of General Internal Medicine*, 33(7), 1131–1138. <https://doi.org/10.1007/s11606-018-4377-y>
- Anderson, J. A., Campbell, K. L., Amer, T., Grady, C. L., & Hasher, L. (2014). Timing is everything: Age differences in the cognitive control network are modulated by time of day. *Psychology and Aging*, 29(3), 648–657.
- Ashton, N. J., Brum, W. S., Di Molfetta, G., Benedet, A. L., Arslan, B., Jonaitis, E., Langhough, R. E., Cody, K., Wilson, R., Carlsson, C. M., Vanmechelen, E., Montoliu-Gaya, L., Lantero-Rodriguez, J., Rahmouni, N., Tissot, C., Stevenson, J., Servaes, S., Therriault, J., Pascoal, T., . . . Lleó, A. (2024). Diagnostic accuracy of a plasma phosphorylated tau 217 immunoassay for Alzheimer disease pathology. *JAMA Neurology*, 81(3), 255. <https://doi.org/10.1001/jamaneurol.2023.5319>
- Benjamini, Y., & Yekutieli, D. (2005). False discovery rate-adjusted multiple confidence intervals for selected parameters. *Journal of the American Statistical Association*, 100(469), 71–81.
- Bernstein, I. H., Lacritz, L., Barlow, C. E., Weiner, M. F., & DeFina, L. F. (2011). Psychometric evaluation of the Montreal Cognitive Assessment (MoCA) in three diverse samples. *The Clinical Neuropsychologist*, 25(1), 119–126.
- Brandt, J., Spencer, M., & Folstein, M. (1988). The telephone interview for cognitive status. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, 1, 111–117.
- Burgers, C., Eden, A., Van Engelenburg, M. D., & Buningh, S. (2015). How feedback boosts motivation and play in a brain-training game. *Computers in Human Behavior*, 48, 94–103.
- Cerino, E. S., Katz, M. J., Wang, C., Qin, J., Gao, Q., Hyun, J., Hakun, J. G., Roque, N. A., Derby, C. A., Lipton, R. B., Sliwinski, M. J. (2021). Variability in cognitive performance on mobile devices is sensitive to mild cognitive impairment: Results from the Einstein aging study. *Frontiers in Digital Health*, 3. <https://doi.org/10.3389/fdgh.2021.758031>
- Clark, P. M., Brunell, A. B., & Buelow, M. T. (2024). The effect of false cognitive feedback on subsequent cognitive task performance. *Journal of Clinical and Experimental Neuropsychology*, 46(5), 422–435. <https://doi.org/10.1080/13803395.2024.2360229>

- Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika*, 3(4), 256–266.
- Cook, S. E., Marsiske, M., & McCoy, K. J. M. (2009). The use of the modified telephone interview for cognitive status (TICS-M) in the detection of amnesic mild cognitive impairment. *Journal of Geriatric Psychiatry and Neurology*, 22(2), 103–109. <https://doi.org/10.1177/0891988708328214>
- David, W. (1981). *WAIS-R manual: Wechsler adult intelligence scale-revised*. The Psychological Corporation.
- Di Rosa, E., Schiff, S., Cagnolati, F., & Mapelli, D. (2015). Motivation-cognition interaction: How feedback processing changes in healthy ageing and in Parkinson's disease. *Ageing Clinical and Experimental Research*, 27(6), 911–920.
- Dodge, H. H., Zhu, J., Mattek, N. C., Austin, D., Kornfeld, J., & Kaye, J. A. (2015). Use of high-frequency in-home monitoring data may reduce sample sizes needed in clinical trials. *PLoS One*, 10(9), e0138095. <https://doi.org/10.1371/journal.pone.0138095>
- Donohue, M. C., Sperling, R. A., Salmon, D. P., Rentz, D. M., Raman, R., Thomas, R. G., & Aisen, P. S. (2014). The preclinical Alzheimer cognitive composite. *JAMA Neurology*, 71(8), 961. <https://doi.org/10.1001/jamaneurol.2014.803>
- Emert, S. E., Taylor, D. J., Gartenberg, D., Schade, M. M., Roberts, D. M., Nagy, S. M., Russell, M., Huskey, A., Mueller, M., Gamaldo, A., Buxton, O. M. (2023). A non-pharmacological multi-modal therapy to improve sleep and cognition and reduce mild cognitive impairment risk: Design and methodology of a randomized clinical trial. *Contemporary Clinical Trials*, 132, 107275. <https://doi.org/10.1016/j.cct.2023.107275>
- Gallo, D. A., Shahid, K. R., Olson, M. A., Solomon, T. M., Schacter, D. L., & Budson, A. E. (2006). Overdependence on degraded gist memory in Alzheimer's disease. *Neuropsychology*, 20(6), 625–632. <https://doi.org/10.1037/0894-4105.20.6.625>
- Grober, E., & Buschke, H. (1987). Genuine memory deficits in dementia. *Developmental Neuropsychology*, 3(1), 13–36.
- Grober, E., Veroff, A. E., & Lipton, R. B. (2018). Temporal unfolding of declining episodic memory on the Free and Cued Selective Reminding Test in the predementia phase of Alzheimer's disease: Implications for clinical trials. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*, 10(1), 161–171. <https://doi.org/10.1016/j.dadm.2017.12.004>
- Hampel, H., Au, R., Mattek, S., van der Flier, W. M., Aisen, P., Apostolova, L., Chen, C., Cho, M., De Santi, S., Gao, P., Iwata, A., Kurzman, R., Saykin, A. J., Teipel, S., Vellas, B., Vergallo, A., Wang, H., Cummings, J. (2022). Designing the next-generation clinical care pathway for Alzheimer's disease. *Nature Aging*, 2(8), 692–703. <https://doi.org/10.1038/s43587-022-00269-x>
- Harrington, K., Hakun, J., Zhaoyang, R., Cerino, E., Hyun, J., Katz, M., Lipton, R., & Sliwinski, M. Feasibility, reliability, and validity of novel mobile app-based cognitive assessments for use with older adults at risk of Alzheimer's disease: Results from the Einstein Aging Study. In: *Paper presented at the The International Neuropsychological Society 49th Annual Virtual Meeting, San Diego, CA*, 2021.
- Harris, C., Tang, Y., Birnbaum, E., Cherian, C., Mendhe, D., & Chen, M. H. (2024). Digital neuropsychology beyond computerized cognitive assessment: Applications of novel digital technologies. *Archives of Clinical Neuropsychology*, 39(3), 290–304. <https://doi.org/10.1093/arclin/aca016>
- Hassenstab, J., Aschenbrenner, A. J., Balota, D. A., McDade, E., Lim, Y. Y., Fagan, A. M., Benzinger, T. L. S., Cruchaga, C., Goate, A. M., Morris, J. C., Bateman, R. J., The Dominantly Inherited Alzheimer Network (2020). Remote cognitive assessment approaches in the Dominantly Inherited Alzheimer Network (DIAN). *Alzheimer's & Dementia*, 16(S6). <https://doi.org/10.1002/alz.038144>
- Koo, B. M., & Vizer, L. M. (2019). Mobile technology for cognitive assessment of older adults: A scoping review. *Innovation in Aging*, 3(1), igy038. <https://doi.org/10.1093/geron/igy038>
- Leroy, V., Gana, W., Aidoud, A., N'kodo, J.-A., Balageas, A.-C.é, Blanc, P., Bomia, D., Debacq, C., Fougère, B. (2023). Digital health technologies and Alzheimer's disease clinical trials: Might decentralized clinical trials increase participation by people with cognitive impairment? *Alzheimer's Research & Therapy*, 15, 87. <https://doi.org/10.1186/s13195-023-01227-4>
- Martin, A., & Fedio, P. (1983). Word production and comprehension in Alzheimer's disease: The breakdown of semantic knowledge. *Brain and Language*, 19(1), 124–141.
- Musiek, E. S., Bhimasani, M., Zangrilli, M. A., Morris, J. C., Holtzman, D. M., & Ju, Y.-E. S. (2018). Circadian rest-activity pattern changes in aging and preclinical Alzheimer disease. *JAMA Neurology*, 75(5), 582–590.
- Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(5), 1170–1187. <https://doi.org/10.1037//0278-7393.26.5.1170>
- Nicosia, J., Aschenbrenner, A. J., Balota, D. A., Sliwinski, M. J., Tahan, M., Adams, S., Stout, S. S., Wilks, H., Gordon, B. A., Benzinger, T. L. S., Fagan, A. M., Xiong, C., Bateman, R. J., Morris, J. C., Hassenstab, J. (2023). Unsupervised high-frequency smartphone-based cognitive assessments are reliable, valid, and feasible in older adults at risk for Alzheimer's disease. *Journal of the International Neuropsychological Society : JINS*, 29(5), 459–471. <https://doi.org/10.1017/S135561772200042X>
- Öhman, F., Hassenstab, J., Berron, D., Schöll, M., & Papp, K. V. (2021). Current advances in digital cognitive assessment for preclinical Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 13(1). <https://doi.org/10.1002/dad2.12217>
- 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., Samaroo, A., Chou, H.-Chin, 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). <https://doi.org/10.1002/dad2.12243>
- Parra, M. A., Abrahams, S., Logie, R. H., Méndez, L. G., Lopera, F., & Della Sala, S. (2010). Visual short-term memory binding deficits in familial Alzheimer's disease. *Brain*, 133(9), 2702–2713. <https://doi.org/10.1093/brain/awq148>
- Parra, M. A., Sala, S. D., Abrahams, S., Logie, R. H., Guillermo Méndez, L., & Lopera, F. (2011). Specific deficit of colour-colour short-term memory binding in sporadic and familial Alzheimer's disease. *Neuropsychologia*, 49(7), 1943–1952. <https://doi.org/10.1016/j.neuropsychologia.2011.03.022>
- Pratap, A., Neto, E. C., Snyder, P., Stepnowsky, C., Elhadad, Némie, Grant, D., Mohebbi, M. H., Mooney, S., Suver, C., Wilbanks, J., Mangravite, L., Heagerty, P. J., Areán, P., Omberg, L. (2020). Indicators of retention in remote digital health studies: A cross-study evaluation of 100,000 participants. *NPI Digital Medicine*, 3(1). <https://doi.org/10.1038/s41746-020-0224-8>
- Reitan, R. M., & Wolfson, D. (1985). *The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation*, vol. 4. Neuropsychology Press.
- Schindler, S. E., Jasielc, M. S., Weng, H., Hassenstab, J. J., Grober, E., McCue, L. M., Morris, J. C., Holtzman, D. M., Xiong, C., Fagan, A. M. (2017). Neuropsychological measures that detect early impairment and decline in preclinical Alzheimer disease. *Neurobiology of Aging*, 56, 25–32. <https://doi.org/10.1016/j.neurobiolaging.2017.04.004>
- Scott, S. B., Graham-Engeland, J. E., Engeland, C. G., Smyth, J. M., Almeida, D. M., Katz, M. J., Lipton, R. B., Mogle, J. A., Munoz, E., Ram, N., Sliwinski, M. J. (2015). The effects of stress on cognitive aging, physiology and emotion (ESCAPE) project. *BMC Psychiatry*, 15, 146. <https://doi.org/10.1186/s12888-015-0497-7>
- Sliwinski, M. J., Mogle, J. A., Hyun, J., Munoz, E., Smyth, J. M., & Lipton, R. B. (2018). Reliability and validity of ambulatory cognitive assessments. *Assessment*, 25(1), 14–30. <https://doi.org/10.1177/1073191116643164>
- Spencer, R., Wendell, C., Giggey, P., Katzel, L., Lefkowitz, D., Siegel, E., & Waldstein, S. (2013). Psychometric limitations of the mini-mental state examination among nondemented older adults: An evaluation of neuro-cognitive and magnetic resonance imaging correlates. *Experimental Aging Research*, 39(4), 382–397. <https://doi.org/10.1080/0361073X.2013.808109>



- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31(1), 137–149. <https://doi.org/10.3758/bf03207704>
- Stricker, J. L., Corriveau-Lecavalier, N., Wiepert, D. A., Botha, H., Jones, D. T., & Stricker, N. H. (2023). Neural network process simulations support a distributed memory system and aid design of a novel computer adaptive digital memory test for preclinical and prodromal Alzheimer's disease. *Neuropsychology*, 37(6), 698–715. <https://doi.org/10.1037/neu0000847>
- Thompson, L. I., Harrington, K. D., Roque, N., Strenger, J., Correia, S., Jones, R. N., Salloway, S., Sliwinski, M. J. (2022). A highly feasible, reliable, and fully remote protocol for mobile app-based cognitive assessment in cognitively healthy older adults. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 14(1). <https://doi.org/10.1002/dad2.12283>
- Thompson, L. I., Kunicki, Z. J., Emrani, S., Strenger, J., De Vito, A. N., Britton, K. J., Dion, C., Harrington, K. D., Roque, N., Salloway, S., Sliwinski, M. J., Correia, S., Jones, R. N. (2023). Remote and in-clinic digital cognitive screening tools outperform the MoCA to distinguish cerebral amyloid status among cognitively healthy older adults. *Alzheimer's & Dementia (Amsterdam, Netherlands)*, 15(4), e12500–e12500. <https://doi.org/10.1002/dad2.12500>
- van Dokkum, L. E. H., Ward, T., & Laffont, I. (2015). Brain computer interfaces for neurorehabilitation – its current status as a rehabilitation strategy post-stroke. *Annals of Physical and Rehabilitation Medicine*, 58(1), 3–8. <https://doi.org/10.1016/j.rehab.2014.09.016>
- Wechsler, D. (1987). *Wechsler memory scale-revised*. Psychological Corporation.
- Wechsler, D. (2001). *Wechsler test of adult reading: WTAR*. Psychological Corporation.
- Weizenbaum, E. L., Soberanes, D., Hsieh, S., Molinare, C. P., Buckley, R. F., Betensky, R. A., Properzi, M. J., Marshall, G. A., Rentz, D. M., Johnson, K. A., Sperling, R. A., Amariglio, R. E., & Papp, K. V. (2023). Capturing learning curves with the multiday Boston Remote Assessment of Neurocognitive Health (BRANCH): Feasibility, reliability, and validity. *Neuropsychology*, 38, 198–210. <https://doi.org/10.1037/neu0000933>
- 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>