



Published in final edited form as:

*Clin Neuropsychol.* 2020 May ; 34(4): 678–699. doi:10.1080/13854046.2020.1733097.

## Bridging the Gap between Performance-based Assessment and Self-reported Everyday Functioning: An Ecological Momentary Assessment Approach

Maureen Schmitter-Edgecombe<sup>1</sup>, Catherine Sumida<sup>1</sup>, Diane J. Cook<sup>2</sup>

<sup>1</sup>Department of Psychology, Washington State University, Pullman, WA

<sup>2</sup>School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA

### Abstract

**Objective:** In the real-world environment, multiple and interacting state-dependent factors (e.g., fatigue, distractions) can cause cognitive failures and negatively impact everyday activities. This study used ecological momentary assessment (EMA) and a n-back task to examine the relationship between fluctuating levels of cognition measured in the real-world environment and self-report and performance-based measures of functional status.

**Method:** Thirty-five community-dwelling older adults ( $M$  age = 71.80) completed a brief battery of objective and self-report measures of cognitive and functional status. After completing 100, 45-second trials to reach stable performance on a n-back task, EMA data collection began. Four times daily for one week, participants received prompts on a tablet to complete a n-back task and a brief survey. From the EMA n-back trials, measures of EMA average performance and intra-individual variability (IIV) across performances were created.

**Results:** For the EMA n-back, the correlation between IIV and EMA average was weak and non-significant. IIV associated with self-report measures, and EMA average with the objective, performance-based functional status composite. Hierarchical regressions further revealed that IIV was a significant predictor of self-reported functional status and cognitive failures over and above EMA average performance and global cognitive status. In contrast, for the objective, functional status composite, IIV did not explain additional variance.

**Conclusions:** The findings suggest that IIV and self-report measures of functional status and cognitive failures may capture a real-world cognitive capacity that fluctuates over time and with context; one that may not easily be captured by objective, performance-based measures designed to assess optimal function.

---

Corresponding author at: Department of Psychology, Washington State University, Pullman, WA 99164-4820, USA, Tel.: +1-509-332-1218; Fax: +1-509-335-5043; schmitter-e@wsu.edu (M. Schmitter-Edgecombe).  
ResearchGate: [https://www.researchgate.net/profile/Maureen\\_Schmitter-Edgecombe](https://www.researchgate.net/profile/Maureen_Schmitter-Edgecombe)

Disclosure statement: The authors have no disclosures.

## Keywords

Ecological Momentary Assessment; Executive Function; Cognitive Control; Intra-individual Variability; Activities of Daily Living

---

Significant negative health consequences, including social isolation, conversion to dementia, and placement in long-term care facilities, are associated with functional impairment in older adults (Fauth et al., 2013; Nourhashemi et al., 2001). Neuropsychologists are frequently asked to use assessments conducted in a clinical setting, typically during one testing session, to predict older individuals' ability to complete activities of daily living in their everyday environments. However, these predictions can be negatively impacted by multiple influences, including the standardized test environment, which may not reflect the complex nature of everyday environments (Small et al., 2019), and unmeasured sources of within-person variability (Sliwinski et al., 2018). For example, an individual's cognitive performance in daily life can fluctuate as a function of stress, fatigue, lack of sleep, a distracting environment, and circadian rhythms (Metternich, Schmidtke, & Hüll, 2009; Schmidt, Collette, Cajochen, & Peigneux, 2007). Such fluctuating or state-dependent factors can impact compensatory cognitive resources, which may be more available under ideal conditions (e.g., a quiet and controlled clinical space) but to a lesser degree in varying real-world contexts (Small et al., 2019). Recently, Carigan and Barkus (2016) proposed that self-reported cognitive failures may measure a real-world cognitive capacity (i.e., one's level of cognitive performance in a particular situation that may shift over time and with context), which is distinct from cognitive ability (i.e., one's relatively stable level of optimal function). In this study, we use ecological momentary assessment (EMA) and a n-back task to examine the relationship between fluctuating levels of cognition captured in the real-world environment and clinical proxy measures of functional status.

Clinicians typically use performance-based assessments and self-report measures as proxies to assess functional status (Schmitter-Edgecombe & Farias, 2018). Similar to neurocognitive tests, performance-based measures have reliable coding systems and are objective, but represent a single evaluation point devoid of personal everyday environmental cues and can be time-consuming to administer (Marson & Hebert, 2006; Myers, Holliday, Harvey, & Hutchinson, 1993; Zimmerman & Magaziner, 1994). Conversely, self-report questionnaires are inexpensive, easy to administer, and may reflect a summarization of the reporter's experiences across multiple real-world situations but are subject to reporter bias (e.g., Dassel & Schmitt, 2008; Richardson, Nadler, & Malloy, 1995). To date, weak correlations between these two methods of functional status assessment have been reported in the aging population (e.g., Burton, Strauss, Bunce, Hunter, & Hultsch, 2009; Finlayson, Havens, Holm, & Van Denend et al., 2003; Jefferson et al., 2008; Tabert et al., 2002). A potential explanation for this weak relationship is that these measures may be capturing different aspects of functional status, a multi-dimensional construct impacted by many factors. In support of this explanation, a study with community-dwelling older adults found that objective, performance-based and self-report measures were unrelated to each other, but both correlated with a directly-observed measure of everyday functioning (Schmitter-Edgecombe, Parsey & Cook, 2011).

A recent systematic review of self-reported cognitive failures in daily life found a similar pattern of small correlations between performance on domain-specific neuropsychological tests and self-report measures of cognitive failure (Carigan & Barkus, 2016). Rather than take the stance that self-report measures are unreliable or reflect poor self-monitoring (Rodriguez et al., 2013; Wilhelm, Witthöft & Schipolowski, 2010), the authors proposed that objective and subjective measurements of cognition could represent equally valuable but different concepts. More specifically, most neuropsychological tests are thought to measure a relatively stable level of cognitive abilities that reflect optimal performance given ideal circumstances. However, optimal performance levels can be altered by state-like factors (Carigan and Barkus, 2016). Given that self-reported cognitive failures have been related to significant real-life outcomes (e.g., at-fault driver in a car accident; university entrance scores; Larson & Merritt, 1991; Unsworth, Brewer, & Spillers, 2012), capturing day-to-day cognitive fluctuations within the real-world environment may improve predictions of functional status and enrich the field's understanding of subjective, self-report data.

Ecological momentary assessment is increasingly being used as a method for capturing fluctuating variables (e.g., fatigue, stress, mood) within an individual's everyday environment (Powell, McMinn & Allan, 2017). Specifically, EMA is a data capture technique where participants are prompted multiple times per day (i.e., times may be random or predetermined) within the naturalistic environment to assess their current behaviors (Shiffman, Stone & Hufford, 2008). Because EMA allows for repeated assessment in a normal daily environment, fluctuations in performance as a result of multiple state-dependent factors can be captured. Recent work has demonstrated the feasibility, reliability and validity of using app-based EMA procedures with older adults to collect cognitive data several times a day (Brouillette et al., 2013; Moore, Depp, Wetherell, & Lenze, 2016; Schweitzer et al., 2017). Cognitive performance data can also be measured together with other important state-like factors, such as fatigue or mood. In fact, a growing literature using EMA data capture techniques have demonstrated real-time associations between fluctuations in cognition and symptom expression, including fatigue (Small et al., 2019), physiological activation (Riediger et al., 2014), and side effects of medication (Frings et al., 2008). These types of EMA studies are improving the field's understanding of the dynamic associations that can occur between daily life state-like factors and cognitive functioning.

In the current study, we take a different approach. Rather than directly investigating dynamic associations, we examine the relationship between fluctuating levels of cognitive performance captured in the real-world environment and objective and subjective proxy measures of functional status collected during a clinical assessment. Specifically, we use EMA to capture within-person fluctuations or intra-individual variability (IIV) in everyday cognitive performance on a n-back task. Typically, IIV is measured by examining variability in trial-to-trial reaction times for a single cognitive task (Jensen, 1992; MacDonald, Li, & Bäckman, 2009). IIV can also be measured by calculating dispersion across a battery of neuropsychological tasks (e.g., Fellows & Schmitter-Edgecombe, 2015) or variability in performance on the same cognitive task administered at different time points (e.g., Hertzog, Dixon, & Hulstsch, 1992). Theoretically, IIV is thought to represent lapses of attention or fluctuations in executive brain function with higher IIV associated with poorer overall cognitive control (Jensen, 1992; Johnson et al., 2015; MacDonald et al., 2009; McVay &

Kane, 2012; Unsworth, 2015). Therefore, when compensatory cognitive resources are compromised by varying state-dependent influences, older individuals who experience difficulties maintaining cognitive control may encounter problems completing daily tasks. Consistent with this supposition, IIV in reaction times has been shown to be sensitive to aging (MacDonald et al., 2009), predictive of cognitive decline (Bielak, Hultsch, Strauss, MacDonald, & Hunter, 2010; Holtzer, Verghese, Wang, Hall, & Lipton, 2008), and associated with prefrontal gray matter atrophy (Hines et al., 2016). Furthermore, IIV as measured through cognitive dispersion has been correlated with everyday prospective memory performance (Sullivan, Woods, Bucks, Loft, & Weinborn, 2018), naturalistic multitasking (Fellows & Schmitter-Edgecombe, 2015), functional disability (Rapp, Schnaider-Berri, Sano, Silverman, & Haroutunian, 2005), and ability to perform activities of daily living (Christensen et al., 1999).

In the current study, neuropsychological tests and proxy measures of functional status, including objective and self-report measures, were used to assess cognition and functional status. EMA data collection was also used to capture fluctuating levels of cognition on a n-back task within the real-world environment (i.e., IIV) as well as an average performance level. Specifically, participants were prompted to complete a brief survey and a n-back task on a tablet device multiple times per day. Because performing the n-back task requires a demanding cascade of cognitive processes, it provides a strong tool for capturing cognitive capacity that could be impacted by multiple state-dependent everyday influences (e.g., fatigue, distractions). That is, n-back tasks require that stimuli be encoded, temporarily stored and continuously updated. At the same time, irrelevant information must be inhibited and abandoned from working memory. Furthermore, the n-back task matches the criteria of a domain-general executive attention task (Kane et al., 2004; Wilhelm, Hildebrandt, & Oberauer, 2013), and executive functioning deficits have most consistently been linked with poorer performance in activities of daily living in the aging population (e.g., Koehler et al., 2011; Rapp & Reischies, 2005; Cahn-Weiner, Malloy, Boyle, Marran, & Salloway, 2000).

We hypothesized that if higher IIV (i.e., greater variability) represents greater difficulty maintaining optimal cognitive control in everyday life, then significant correlations should emerge between IIV and self-report measures of functional status and cognitive failures, which may capture a more real-world cognitive capacity. In contrast, the objective, performance-based functional status measure (i.e., a reflection of performance under ideal conditions) was not expected to be associated with IIV but rather to demonstrate a significant relationship with average performance on the n-back task. We further hypothesized that IIV would explain additional variance in self-report measures of functional status and cognitive failures over and above variance accounted for by global cognitive status and average n-back performance. We also examined for relationships between average n-back performance and IIV, and standardized tests assessing executive functioning and prospective memory. Given that laboratory tests of prospective memory may be sensitive to fluctuating levels of attention (Sullivan et al., 2018), we were especially interested in determining whether tests assessing prospective memory might correlate with the computed naturalistic measure of IIV.

## Methods

### Participants

The sample is composed of 35 community-dwelling older adults. All participants were recruited through community advertisements or referrals from physicians and local community agencies in the Whitman and Spokane counties of Washington State. Prior to being selected to participate in this study, participants completed a health screening interview and the Telephone Interview of Cognitive Status (TICS, Brandt & Folstein, 2003) over the phone. The interview was used to rule out exclusion criteria, including age less than 45 and the possibility of dementia (TICS < 26) or lack of insight that could lead to unreliable EMA self-report data. The data was collected between 2016 and 2018 as part of a larger study focused on developing and validating health algorithms using smart environment technologies. All participant homes were converted to smart homes for 3–4 months, which involved installing ambient sensors that collected continuous data while residents performed their normal routines. To be eligible for this study, participants must have completed a minimum of 16 EMA responses within the first seven days of EMA data collection. Of the 44 participants enrolled in the larger study, EMA data from four participants could not be collected due to technical issues and data from five participants did not reach a 16 EMA-response minimum within the initial 7-day period. The study participants consisted of 24 female and 11 male participants, with a mean age of 71.80 years (SD = 9.64; range 48–89 years) and a mean education of 15.83 years (SD = 2.68; range = 12–20 years). This protocol was reviewed and approved by the Washington State University Institutional Review Board, and all participants gave verbal consent and signed the institutional review approved consent document.

### Measures

The repeatable battery for the assessment of neuropsychology status (RBANS; Randolph, 2012) is a brief (approximately 30 minutes) cognitive battery of 12 tests used to establish neuropsychological status and detect cognitive decline. The total scale index score was used as a measure of global cognitive status and entered into the regression analysis to control for cognitive status. To reduce the number of correlation and regression analyses conducted between the objective and subjective (self-report) measures of cognition and functional status, composite scores were created when possible to represent constructs for each domain of functioning.

### Objective Measures

**Executive Functioning:** The letter fluency and design fluency tests were administered from the Delis-Kaplan Executive Function System (D-KEFS; Delis, Kaplan & Kramer, 2001) to capture executive functioning. The letter fluency subtest requires individuals to quickly retrieve and name words beginning with specified letters (i.e., three trials, one for the letters F, A, and S) while not providing names of people, places and numbers. For the design fluency test, participants were presented with an 8.5 by 11-inch sheet of paper with 35 boxes containing uniformly-positioned dots inside each box. Participants are required to quickly create different designs by connecting dots, using four straight lines, and connecting all lines to at least one dot. The design fluency test consists of three subtests. The raw total

number correct from the three-letter fluency trials was transformed to a z-score as was the raw total designs produced across the three design fluency subtests. The z-scores were then summed to create an executive functioning composite.

**Prospective Memory:** Given that prospective memory tests may be sensitive to fluctuating levels of cognition across time, we captured prospective memory performance during the assessment using an activity memory paradigm (Schmitter-Edgecombe, Woo & Greeley, 2009) and a hidden object trial. The activity memory paradigm required participants to remind the examiner to give their friend pain medication each time they were presented with a scale card (1–5 Likert scale; the “event-cue”) and asked to rate how challenging they found the task they just completed. Participants received eight event-based cues over the course of approximately one hour. Each cue was scored as 0 (no response), .5 (early or late response), or 1 (correct response). For the hidden object trial, a personal object (e.g., watch) from the participant was taken at the beginning of testing and hidden. The participant was required to ask for the hidden object at the end of testing when the examiner said, “We have now completed all tasks and testing is over”. Scoring ranged from 0 to 3 and was dependent on whether the participant completed none, part or all of the task correctly, at the right time, and without cueing (Raskin, Buckheit, & Sherrod, 2010). The raw score from the hidden object trial was added to the raw score from the activity memory paradigm to create the prospective memory total score.

**Performance-based Functional Status:** The Independent Living Scale (ILS; Loeb, 1996) assesses functional abilities by providing examples of various everyday situations and requiring participants to verbally state or demonstrate how they would manage those situations to remain living independently. The ILS was used as the performance-based measure of functional status given that it requires multiple executive skills, including planning, problem-solving, and flexibility in thinking. Two of the five ILS scales were administered to participants. The Managing Home and Transportation scale (15 items), which assesses a participant’s ability to manage a home, use a telephone, and travel using public transportation, and the Health and Safety scale (20 items), which assesses a participant’s ability to identify and manage various health problems and report relevant safety precautions. Each item is scored on a 0 (no points) to 2 (full points) scale. The raw total for each scale was converted to a z-score. The z-scores were then summed to create the performance-based everyday functioning score. This measure has a lower sample size ( $n=17$ ) than the other measures because it was associated with a different study that finished prior to the end of the current study.

## Subjective Measures

**Self-report Cognitive Failures:** Participants were administered the dysexecutive questionnaire (DEX, Wilson et al., 1996) from the Behavioural Assessment of the Dysexecutive Syndrome test battery as well as the prospective and retrospective memory questionnaire (PRMQ; Smith, Dela Sala, Logie & Maylor, 2000). The DEX requires participants to respond to 20 questions assessing everyday problems commonly associated with executive dysfunction using a 5-point Likert scale ranging from 0 (never) to 4 (very often). A sum score is computed from the questions. The PRMQ is a self-report



questionnaire made up of 16 items assessing everyday prospective and retrospective memory difficulties in daily life. A sum is calculated using participant ratings for each memory problem on a 5-point Likert scale, ranging from 1 (never) to 5 (always). The raw total score for each questionnaire was converted to a z-score. The z-scores were then summed to create a measure of everyday cognitive failures, with higher scores indicating greater self-reported cognitive failures.

**Self-Report Functional Status:** Participants were administered the self-report form of the Instrumental Activities of Daily Living: Compensation (IADL-C; Schmitter-Edgecombe, Parsey & Lamb, 2014) scale. This 27-item questionnaire assesses participant's ability to complete varying everyday activities of daily living (e.g., money management, travel). Participants respond using an 8-point scale, ranging from 1 ("independent, no aid") to 8 ("cannot complete this activity anymore"); higher scores represent greater difficulty with everyday tasks. The scale is sensitive to early cognitive decline (Schmitter-Edgecombe et al., 2014).

### EMA Data

**N-back Task:** Participants completed the n-back task on an iPad (9.7-inch diagonal). Prior work suggests that tablets may be better data collection tools than smart phones with the older adult population given the larger screen (Brouillette et al., 2013). The n-back task used in this study was a 1-back. Simple shapes (i.e., circle, triangle, diamond) were presented one at a time to participants (see Figure 1). The shapes were large, with good contrast (blue on a white background) and the "yes" (green) and "no" (red) buttons were large, appropriately colored and separated in distance to prevent inadvertent selection of the wrong button. In addition, the following instructions were always visible above the "yes" and "no" buttons: "Is this shape the same as the previous one?". Furthermore, a counter indicating the amount of time left in the trial was also visible above the shape. Each n-back trial presented shape comparison for 45 seconds. Shapes were randomly presented for all n-back trials. Prior to beginning each 45-second n-back trial, a shape with a visible 3-second countdown timer (i.e., 3, 2 and 1) appeared to give the participant time to get ready for the trial. The first shape (i.e., one with countdown timer) was followed by a second shape, and participants indicated by selecting the "yes" or "no" button whether the first shape was the same or different than the second shape. The 45-second n-back trial began when the second shape appeared. Following the participant's response, a third shape appeared, which the participant then compared to the second shape and so on. The application continued to present shapes to the participant until the 45 seconds ended. Participants were told to respond as quickly and accurately as possible. The next shape appeared on the screen approximately 500 milliseconds after the participant selected a response. At the completion of each n-back trial, participants could navigate to a "Shape Tests Results" page, which displayed the total number of accurate shape comparisons they completed within the most recent 45-second trial and for all prior trials, allowing participants to monitor their progress.

### Procedures

Participants completed a brief battery of cognitive tests (approximately 1.5–2 hours) with an examiner in their own homes. At the end of the assessment, the examiner set up an iPad in a

central location (e.g., kitchen counter). Participants were informed that they would need to first complete 100 trials of the n-back task prior to the start of EMA data collection. The examiner then used an EMA simulator to demonstrate an EMA data collection, which allowed the participant to pretend to respond to the 7-item questionnaire and provided a teaching tool for the n-back task. Performance on the EMA simulator's n-back task did not contribute to the participant's required 100 trials but rather served to familiarize the participant with how to complete the n-back task. Once the participant understood how to complete the n-back task, they completed 10, 45-second trials, with the examiner and demonstrated their ability to navigate to the "Shape Tests Results" page to view their results and progress. Participants were then asked to complete an additional 90 trials of the n-back task as practice during the course of the following week. Progress was monitored by laboratory personnel remotely and reminder calls were made when needed.

After completion of the 100 practice trials, the 7-day EMA data collection period began. During EMA collection, participants received prompts to complete an EMA data collection at random times within four standard three-hour blocks: 8:00 AM – 11:00 AM; 11:30 AM – 2:30 PM; 3:00 PM – 6:00 PM; and 6:30 PM – 9:30 PM. For a few participants, the blocks were shifted to accommodate sleep/wake schedules and/or other scheduled activities where the individual was consistently out of the home (e.g., volunteering, work) because the tablets remained in the participant's home. If the participant did not initially respond to the prompt, a second prompt was initiated after 10 minutes. If there was no response following the second prompt, the application would not prompt again until the next scheduled time. The average time between prompts across participants was approximately 3 hours.

EMA collection consisted of a 7-item questionnaire followed by the 45-second n-back task. Each EMA data collection took 2–3 minutes to complete. Only the n-back task data was analyzed for this study. Specifically, four n-back variables were used in data analyses. The 100 trials of n-back practice resulted in the following two n-back variables: clinical average and asymptote. *Clinical average*, which represented n-back performance collected during the clinical assessment, was computed as the average of the total correct responses for each of the 45-second n-back practice trials 1–10. *Asymptote*, which represented a stable level of performance on the n-back task, was computed as the average of the total correct responses for each of the 45-second n-back trials 91–100. The EMA data collection resulted in the following two variables: EMA average and IIV. The *EMA average*, which was computed as the average of the total correct responses for each 45-second n-back task completed during the EMA data collection period, was used to represent the participant's mean EMA n-back performance. Consistent with prior work using standard deviations of RTs to index IIV (e.g., MacDonald et al., 2009; Lu & Wang, 2018), we computed *IIV* as the standard deviation of the total correct responses for each of the 45-second n-back tasks completed by each individual during the EMA data collection period. The IIV measure was considered to capture fluctuations in cognitive performance and represent the impact that real-life momentary influences can have on everyday functioning.



## Analyses

Before computing the EMA average and IIV measures, we examined n-back performance for outlier trials that may have resulted from spurious situations (e.g., interruption occurring in the real-world environment). One data point was removed from EMA analysis due to only 13 n-back task stimuli being displayed for response during the 45-second EMA n-back trial. This was considerably less than the next minimum number, which occurred numerous times, of 22 n-back task stimuli displayed for response during an EMA n-back trial. With the exception of one participant, accuracy of responding on the n-back task was 94% for all study participants and the overall median accuracy rate was 98.90%. The mean number of EMA n-back task trials completed per participant across the 7 days of EMA data collection and available for IIV calculation was 20 EMA n-back task trials ( $SD = 2.65$ , range = 16–25).

The mean, standard deviation and ranges for the clinical variables and four n-back variables are displayed in Table 1. With the exception of the composite measure of prospective memory, all measures were normally distributed, with skewness and kurtosis values  $< 1.00$ . Because skewness ( $-1.48$ ) and kurtosis were not heightened too greatly (1.52), the prospective memory composite was not transformed and parametric tests were used for all analyses.

To assess practice effects and to determine whether n-back asymptotic performance (i.e., a stable level of performance) had been reached, we used a repeated-measures analysis of variance (ANOVA). Correlation analyses were used to examine for associations among the n-back variables (i.e., clinical average, asymptote, EMA average, IIV) and for relationships between the n-back variables, demographic variables (i.e., age, gender, education) and global cognitive status. Correlation analyses were also used to explore relationships between the n-back variables and the objective and subjective clinical measures of executive functioning, prospective memory, cognitive failures and functional status. Given the sample size and early nature of this work, we chose to keep the p-value for the correlation analyses at .05. All significant correlations between the n-back variables and clinical measures fell above  $r = .40$ , and can be considered to suggest a moderate relationship. Finally, we conducted multiple hierarchical regression analyses to determine whether IIV (influenced by state-dependent factors) (entered in block 2) would account for a significant amount of variance on the self-report measures of cognitive failures and functional status over and above variance accounted for by global cognitive status (i.e., RBANS total) and EMA average performance (entered in block 1).

## Results

### Practice Effects and Asymptote Performance:

The repeated measures ANOVA comparing the n-back clinical average (first 10 n-back trials: 1–10), asymptote (last 10 n-back trials: 91–100) and EMA average scores revealed a significant main effect of time,  $F(2, 33) = 106.18$ ,  $MSE = 1148.30$   $p < .001$ ,  $\eta_p^2 = .76$ . Post hoc tests revealed the presence of a significant practice effect, with ten additional accurate responses on the n-back task being completed at asymptote ( $M = 33.53$ ,  $SD = 4.14$ ) and during EMA data collection ( $M = 33.19$ ,  $SD = 4.30$ ) compared to the clinical average ( $M =$

23.44,  $SD = 5.55$ ),  $F_s(1, 34) > 110.45$ ,  $p_s < .001$ . The lack of a difference between the asymptote and EMA average scores,  $F = .59$ , along with visual inspection of each participant's 100 n-back practice trials for asymptote, suggested that participants were accurately completing a stable number of n-back responses during each 45-second task trial prior to EMA data collection. Therefore, variations in performance across EMA data collection can be interpreted as representing fluctuating or state-dependent influences on performance (e.g., fatigue, mood, distractions etc.) rather than continued learning (i.e., practice effects).

### **Correlations Among n-back Task Measures and with Demographic Variables and Global Cognitive Status.**

Correlational analyses were conducted to examine associations among the following four n-back scores: clinical average, asymptote, EMA average, and IIV. As can be seen in Table 2, no significant correlations emerged between the IIV score and the three average performance scores ( $r$ 's =  $-.16$  to  $-.05$ ), which captured the number of n-back responses accurately completed both prior to (clinical average, asymptote) and during (EMA average) EMA data collection. This suggests that the IIV variable is capturing a different neurocognitive capacity than the average performance measures. Furthermore, there was a moderate correlation between the clinical average and both asymptote ( $r = .44$ ,  $p = .009$ ) and EMA average ( $r = .40$ ,  $p = .02$ ), and the expected strong correlation between asymptote and EMA average ( $r = .81$ ,  $p < .001$ ). These findings again suggest that a successful asymptote in mean n-back performance was reached prior to EMA data collection. Correlation analyses further revealed that neither age nor sex correlated with the n-back scores (see Table 2). Education level correlated significantly with the average performance scores ( $r$ 's  $> .34$ ) but not with IIV ( $r = -.21$ ,  $p = .23$ ). Furthermore, global cognitive status (i.e., RBANS total) correlated significantly with the clinical average ( $r = .44$ ,  $p = .008$ ) score. Of note, after a stable level of performance was reached on the n-back task, there was almost no correlation between global cognitive status and the asymptote or EMA average scores (see Table 2).

### **Correlations with Cognitive and Functional Status Measures.**

Correlations were then conducted between the EMA n-back variables and the clinical measures of executive functioning, prospective memory, cognitive failures and functional status. As can be seen in Table 3, there was a significant positive correlation between EMA average and the performance-based measure of functional status ( $r = .60$ ,  $p = .01$ ). The remaining correlations between the clinical measures and EMA average were weak and non-significant ( $r$ 's between  $-.16$  and  $.17$ ). In contrast, for the IIV measure, correlation with the performance-based measure of functional status was weak ( $r = -.15$ ,  $p = .56$ ), while correlations with the self-report measures of cognitive failures ( $r = .53$ ,  $p = .003$ ) and functional status ( $r = .54$ ,  $p = .002$ ), and with the prospective memory measure ( $r = -.42$ ,  $p = .02$ ), were all moderate. Figure 2 graphically displays with scatterplots the contrasting relationships between the n-back EMA average and IIV scores and the proxy measures of functional status. Because fewer individuals completed the performance-based functional status test, to check that the IIV correlations showed a similar pattern with this smaller sample, we re-ran the IIV correlation analyses using only the subsample of individuals that completed the performance-based functional status test. The IIV correlations with the self-

report cognitive failures ( $r = .59$ ) and functional status ( $r = .50$ ) measures, but not the prospective memory measure ( $r = -.18$ ), remained similar in magnitude.

### Regression Analyses.

Table 4 displays the hierarchical regression analyses. For the executive composite, RBANS total and EMA average explained a significant amount of variance,  $R^2 = .18$ ,  $F(2,32) = 3.58$ ,  $p = .04$ . Addition of the IIV variable in block 2 resulted in little additional variance being explained,  $R^2 = .02$ ,  $F(1,31) = .65$ ,  $p = .43$ . The RBANS measure of global cognitive status,  $t = 2.52$ ,  $p = .02$ , emerged as the only significant predictor of the executive function composite after controlling for all other variables.

The regression analysis for the prospective memory total revealed that IIV accounted for a significant amount of the variance above and beyond RBANS total and EMA average,  $R^2 = .18$ ,  $F(1,28) = 6.74$ ,  $p = .02$ , which were not significant predictors in block 1,  $R^2 = .08$ ,  $F(2,29) = 1.23$ ,  $p = .31$ . Furthermore, the IIV measure,  $t = -2.59$ ,  $p = .02$ , emerged as the only significant predictor of the prospective memory total.

Analysis of the regression models for the self-report measures revealed a similar pattern, in that IIV accounted for a significant amount of the variance above and beyond RBANS total and EMA average for both the self-report cognitive failures,  $R^2 = .25$ ,  $F(1,26) = 8.94$ ,  $p = .006$ , and functional status,  $R^2 = .27$ ,  $F(1,26) = 10.57$ ,  $p = .003$ , measures. Furthermore, RBANS total and EMA average did not account for a significant amount of the variance in block 1 for either analysis. After controlling for all other variables in block 2, IIV emerged as the only significant predictor for both self-reported cognitive failures,  $t = 2.99$ ,  $p = .006$ , and functional status,  $t = 3.25$ ,  $p = .003$ . Overall, the set of predictors accounted for 28% and 34% of the total variance in self-reported cognitive failures and functional status, respectively, with the majority of variance explained by the IIV measure.

In contrast, for the objective, performance-based functional status measure, a significant amount of the variance was explained by global cognitive status (RBANS total) and EMA average,  $R^2 = .37$ ,  $F(2,14) = 4.04$ ,  $p = .04$ . No additional variance was accounted for by adding IIV in block 2,  $R^2 = .008$ ,  $F(1,13) = .17$ ,  $p = .69$ . The EMA average n-back performance,  $t = 2.70$ ,  $p = .02$ , emerged as the only significant predictor in the final model.

### Discussion

Real-world environments have multiple and interacting state-dependent factors (e.g., mood, anxiety, chaotic surroundings), which can cause cognitive failures and impact the quality of daily living tasks (e.g., remembering to take clothes out of washer). Many people experience occasions when their ability to perform everyday activities is compromised by environmental or other state-dependent influences. In the current study, EMA data collection was used to capture both mean cognitive performance and fluctuating levels of cognition in the real-world environment as measured by a n-back task. We are unaware of prior studies that have examined the impact of fluctuating everyday factors on measures of functional status by computing IIV from cognitive EMA data.

In line with prior studies (Brouillette et al., 2013; Moore et al. 2016; Schweitzer et al., 2017), the data illustrate that EMA can be used to capture information about cognition in an older adult population. High participant accuracy rates suggest that the large interface (i.e., tablet) and stimulus display designed for older adults (e.g., high contrast, large stimuli, intuitive design) successfully reduced inadvertent n-back task errors (e.g., tapping wrong response button). Prior studies have also observed practice effects when using repeated cognitive assessments in the real-world environment (e.g., Waters & Li, 2008; Bouvard et al., 2018). The provision of 100, 45-sec n-back practice trials, prior to EMA data collection successfully reduced the impact of continued learning. In comparison to the clinical average (trials 1–10), ten additional accurate n-back responses were completed at asymptote and during EMA data collection. Furthermore, the asymptote and EMA data average scores were strongly correlated. These findings increase confidence in our interpretation that variations in an individual's performance across the EMA trials more purely reflects the impact of momentary factors (e.g., distraction, fatigue) on performance rather than learning.

The data also showed a weak and nonsignificant relationship between the IIV and EMA average scores, suggesting that these measures may reflect different components of cognitive performance as captured through repeated n-back assessment in a real-world environment. This supposition was further supported by the contrasting pattern of correlations found between the EMA variables and the subjective and objective measures. Consistent with our hypothesis that higher IIV may represent greater difficulty maintaining optimal cognitive control in everyday life, higher IIV was associated with higher levels of self-reported cognitive failures (i.e., executive and memory) and with poorer self-reported functional status. IIV did not associate with the objective, performance-based composite of functional status (i.e., ILS subtests). In contrast, higher EMA average scores were associated with higher scores on the objective measure of functional status, but did not associate with self-reported functional status or cognitive failures.

While objective measures are designed to capture cognitive performance under optimal conditions, greater IIV has been theorized to represent poorer executive brain function and cognitive control, in addition to lapses of attention (Williams, Thayer, & Koenig, 2016). Therefore, the contribution of IIV to predicting everyday failures may reflect its role as a measure of overall cognitive control. From a psychometric and clinician perspective, these findings are consistent with the notion that objective (i.e., performance-based) and subjective (i.e., self-report) measurements of functional status could possibly represent different but equally valuable concepts. Specifically, an individuals' relatively stable level of functional capacity may be best assessed by performance-based measures of functional status. In contrast, difficulties with maintaining cognitive control in the real-world environment in the face of fluctuating demands, which a clinician cannot directly observe or measure, might be better captured by self-report or informant-report measures of functional status. When neuropsychologists are tasked with making predictions about functional status, the current findings suggest that it may be important to assess and weigh findings from both self-report and objective measures of functional status.

Prior work suggests that cognitive abilities typically predict between 20–25% of the total variance in functional status (McAlister, Schmitter-Edgecombe & Lamb, 2016; Royall et al.,

2007; Tucker-Drob, 2011). An objective measure of the ability to maintain control when faced with changing real-world challenges may have the potential to improve ability of neuropsychologists to predict functional status from clinical data. It is possible that lower IIV (i.e., better cognitive control) may serve as a protective factor against fluctuations in momentary influences that could lead to cognitive failures and poorer completion of daily activities. For example, an older individual with better cognitive control may be able to compensate for low arousal or fatigue by adapting their processing or behavior (e.g., use a compensatory strategy or complete the task at an alternate time). Prior studies have also shown that measures of IIV calculated from trial-to-trial RTs are sensitive to cognitive decline (Bielak et al., 2010; Holtzer et al., 2008) and may represent a measure of neuronal change (Hines et al., 2016). Given the early nature of this work linking IIV captured using EMA and a brief n-back task to real-world cognitive control, it should be considered exploratory. Future work is needed to examine the potential for measures of IIV, including EMA derived IIV from cognitive tasks administered repeatedly in the real-world environment, to predict functional status as assessed through real-world observation of older individuals completing tasks across changing conditions in their everyday environments.

Evidence linking measures of IIV to regions of executive brain function come from studies of individuals with damage to the prefrontal cortex (e.g., Murtha, Cismaru, Waechter, & Chertkow, 2002), behavioral studies (e.g., Bellgrove, Hester, & Garavan, 2004) and from functional magnetic imaging work (see MacDonald et al., 2009 for a review). IIV has been conceptualized as representing attention lapses through wavering cognitive control (Vasquez, Binns, & Anderson, 2018). Although the largest associations in this study were found between IIV and self-report measures, IIV also associated moderately with prospective memory, which can be impacted by lapses in attention. In addition, IIV predicted prospective memory performance over and above EMA average and global cognitive status. Recent work has shown that cognitive control, as indexed by a dispersion measure of IIV, may play a role in naturalistic and clinic-based strategic prospective memory performance in older adults (Ihle, Ghisletta, & Kliegel, 2017; Sullivan et al., 2018). Prior studies have also shown that prospective memory plays an important role in supporting everyday functional status (Schmitter-Edgecombe et al., 2009; Schmitter-Edgecombe, McAlister, & Weakley, 2012; Woods, Weinborn, Velnoweth, Rooney, & Bucks, 2012).

Also of interest, the RBANS total score was significantly related to the n-back clinical average (trials 1–10) but showed virtually no relationship with the asymptote and EMA average scores. It has been suggested that repeated EMA cognitive assessment may lead to a more stable characterization of a person's average level of cognitive performance (Allard et al., 2014; Sliwinski et al., 2018). Furthermore, global cognitive status, but not n-back EMA average or IIV, was a significant predictor of the executive composite. It is possible that global cognitive status may play a more significant role when predicting initial, as opposed to practiced performance on a novel task. This might be because higher cognitive functioning individuals can more quickly process and remember instructions and learn and develop strategies, but over repeated exposures these advantages to performance have less of an impact.

Strengths of this study include having participants reach asymptote on the n-back task prior to beginning EMA data collection. This reduced the influence of practice effects, which have been observed in other studies that have used repeated real-world cognitive assessments (e.g., Waters & Li, 2008; Bouvard et al., 2018), and increased confidence that variations in n-back performance reflected momentary factors. Furthermore, our dynamic association work with the survey question data and n-back task provides evidence for an influencing role of state-dependent effects (Caffery et al., submitted). More specifically, we found that when the older adult participants reported experiencing greater fatigue than typical or reported being more physically active than typical, subsequent (3-hour time lag) cognitive performance was worse (Caffery et al., submitted).

Study limitations include our relatively small sample of highly educated older adults, which may have artificially inflated some of the findings and limits generalization to other older adult populations or other age ranges. Analyses conducted with the performance-based functional status measure were especially impacted by sample size and will need to be replicated. Generalization of our findings is also limited by the cognitive and functional tests administered in this study such that the findings will need to be replicated. In addition, future research will be required to determine whether the findings extend beyond the specific n-back task used in this study and to other cognitive domains (e.g., memory) whose variability could be captured through EMA data collection. Furthermore, we did not ask participants about interruptions during each EMA data collection period and future studies may want to do so. One study that asked participants about distractions and interruptions during each EMA data entry found an expected substantial amount of distraction but few reports of interruptions (Tiplady, Oshinowo, Thomson, & Drummond, 2009).

In summary, the findings suggest that cognitive control abilities may interact with momentary factors and influence the quality of everyday task completion. Additional research is needed to determine whether IIV, as measured using EMA and a brief n-back task, could be used to provide an objective method for capturing the impact of real-world lapses in executive brain function and attention that could negatively impact daily life activities and cause cognitive failures. Such a method may also assist in better understanding real-world functioning when individuals have poor insight. For a healthy older adult population, the data suggest that self-report measures assessing cognitive failures and functional status limitations may capture information about the real-world environments impact on cognitive capacity, which is not being captured by objective, performance-based measures of functional status. This has important implications for how clinicians interpret data from self-report measures assessing functional status when they do not align with performance-based assessment.

## Acknowledgements:

We thank Sherin Shahsavand for her assistance in developing the n-back app. We also thank members of the WSU Neuropsychology and Aging laboratory and the CASAS team for their help with data collection.

Funding details: This work was supported by the NIBIB under Grant #R01 EB009675 and NIA under grant #R25 AG046114.



## References

- Allard M, Husky M, Catheline G, Pelletier A, Dilharreguy B, Amieva H, ... Swendsen J (2014). Mobile Technologies in the Early Detection of Cognitive Decline. *PLoS ONE*, 9(12), e112197. 10.1371/journal.pone.0112197 [PubMed: 25536290]
- Bellgrove MA, Hester R, & Garavan H (2004). The functional neuroanatomical correlates of response variability: evidence from a response inhibition task. *Neuropsychologia*, 42(14), 1910–1916. 10.1016/j.neuropsychologia.2004.05.007 [PubMed: 15381021]
- Bielak AAM, Hultsch DF, Strauss E, MacDonald SWS, & Hunter MA (2010). Intraindividual variability in reaction time predicts cognitive outcomes 5 years later. *Neuropsychology*, 24(6), 731–741. 10.1037/a0019802 [PubMed: 20853957]
- Bouvard A, Dupuy M, Schweitzer P, Revranche M, Fatseas M, Serre F, ... Swendsen J (2018). Feasibility and validity of mobile cognitive testing in patients with substance use disorders and healthy controls: Mobile Cognitive Testing in Patients With SUD. *The American Journal on Addictions*, 27(7), 553–556. 10.1111/ajad.12804 [PubMed: 30260085]
- Brandt J, & Folstein M (2003). Telephone Interview for Cognitive Status. Lutz, FL: Psychological Assessment Resources, Inc.
- Brouillette RM, Foil H, Fontenot S, Correro A, Allen R, Martin CK, ... Keller JN (2013). Feasibility, Reliability, and Validity of a Smartphone Based Application for the Assessment of Cognitive Function in the Elderly. *PLoS ONE*, 8(6), e65925. 10.1371/journal.pone.0065925 [PubMed: 23776570]
- Burton CL, Strauss E, Bunce D, Hunter MA, & Hultsch DF (2009). Functional Abilities in Older Adults with Mild Cognitive Impairment. *Gerontology*, 55(5), 570–581. 10.1159/000228918 [PubMed: 19602873]
- Caffrey KS, Preszler J, Burns GL, Wright BR, Cook DJ, & Schmitter-Edgecombe M (2019). Dynamic SEM for lifestyle factors associated with cognitive performance. Manuscript submitted for publication.
- Cahn-Weiner DA, Malloy PF, Boyle PA, Marran M, & Salloway S (2000). Prediction of functional status from neuropsychological tests in community-dwelling elderly individuals. *The Clinical Neuropsychologist*, 14(2), 187–195. 10.1076/1385-4046(200005)14:2:1-Z;FT187 [PubMed: 10916193]
- Carrigan N, & Barkus E (2016). A systematic review of cognitive failures in daily life: Healthy populations. *Neuroscience & Biobehavioral Reviews*, 63, 29–42. 10.1016/j.neubiorev.2016.01.010 [PubMed: 26835660]
- Christensen H, Mackinnon AJ, Korten AE, Jorm AF, Henderson AS, & Jacomb P (1999). Dispersion in Cognitive Ability as a Function of Age: A Longitudinal Study of an Elderly Community Sample. *Aging, Neuropsychology, and Cognition*, 6(3), 214–228. 10.1076/anec.6.3.214.779
- Dassel KB, & Schmitt FA (2008). The Impact of Caregiver Executive Skills on Reports of Patient Functioning. *The Gerontologist*, 48(6), 781–792. 10.1093/geront/48.6.781 [PubMed: 19139251]
- Delis DC, Kaplan E, Kramer JH, & Psychological Corporation. (2001). *Delis-Kaplan Executive Function System: Technical Manual*. San Antonio, TX: Psychological Corp.
- Fauth EB, Schwartz S, Tschanz JT, Østbye T, Corcoran C, & Norton MC (2013). Baseline disability in activities of daily living predicts dementia risk even after controlling for baseline global cognitive ability and depressive symptoms: ADL disability predicts dementia risk. *International Journal of Geriatric Psychiatry*, 28(6), 597–606. 10.1002/gps.3865 [PubMed: 22968965]
- Fellows RP, & Schmitter-Edgecombe M (2015). Between-domain cognitive dispersion and functional abilities in older adults. *Journal of Clinical and Experimental Neuropsychology*, 37(10), 1013–1023. 10.1080/13803395.2015.1050360 [PubMed: 26300441]
- Finlayson M, Havens B, Holm MB, & Van Denend T (2003). Integrating a Performance-Based Observation Measure of Functional Status into a Population-Based Longitudinal Study of Aging. *Canadian Journal on Aging / La Revue Canadienne Du Vieillissement*, 22(2), 185–195. 10.1017/S0714980800004505

- Frings L, Wagner K, Maiwald T, Carius A, Schinkel A, Lehmann C, & Schulze-Bonhage A (2008). Early detection of behavioral side effects of antiepileptic treatment using handheld computers. *Epilepsy & Behavior*, 13(2), 402–406. 10.1016/j.yebeh.2008.04.022 [PubMed: 18539084]
- Hertzog C, Dixon RA, & Hultsch DF (1992). Intraindividual change in text recall of the elderly. *Brain and Language*, 42(3), 248–269. 10.1016/0093-934X(92)90100-S [PubMed: 1606486]
- Hines LJ, Miller EN, Hinkin CH, Alger JR, Barker P, Goodkin K, ... Becker JT (2016). Cortical brain atrophy and intra-individual variability in neuropsychological test performance in HIV disease. *Brain Imaging and Behavior*, 10(3), 640–651. 10.1007/s11682-015-9441-1 [PubMed: 26303224]
- Holtzer R, Verghese J, Wang C, Hall C, & Lipton R (2008). Within-Person Across-Neuropsychological Test Variability and Incident Dementia. *JAMA*, 300(7), 823. 10.1001/jama.300.7.823 [PubMed: 18714062]
- Ihle A, Ghisletta P, & Kliegel M (2017). Prospective memory and intraindividual variability in ongoing task response times in an adult lifespan sample: the role of cue focality. *Memory*, 25(3), 370–376. 10.1080/09658211.2016.1173705 [PubMed: 27088610]
- Jefferson AL, Byerly LK, Vanderhill S, Lambe S, Wong S, Ozonoff A, & Karlawish JH (2008). Characterization of Activities of Daily Living in Individuals With Mild Cognitive Impairment. *The American Journal of Geriatric Psychiatry*, 16(5), 375–383. 10.1097/JGP.0b013e318162f197 [PubMed: 18332397]
- Jensen AR (1992). The importance of intraindividual variation in reaction time. *Personality and Individual Differences*, 13(8), 869–881. 10.1016/0191-8869(92)90004-9
- Johnson BP, Pinar A, Fornito A, Nandam LS, Hester R, & Bellgrove MA (2015). Left anterior cingulate activity predicts intra-individual reaction time variability in healthy adults. *Neuropsychologia*, 72, 22–26. 10.1016/j.neuropsychologia.2015.03.015 [PubMed: 25791710]
- Kane MJ, Hambrick DZ, Tuholski SW, Wilhelm O, Payne TW, & Engle RW (2004). The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. *Journal of Experimental Psychology: General*, 133(2), 189–217. 10.1037/0096-3445.133.2.189 [PubMed: 15149250]
- Koehler M, Kliegel M, Wiese B, Bickel H, Kaduszkiewicz H, van den Bussche H, ... Pentzek M (2011). Malperformance in Verbal Fluency and Delayed Recall as Cognitive Risk Factors for Impairment in Instrumental Activities of Daily Living. *Dementia and Geriatric Cognitive Disorders*, 31(1), 81–88. 10.1159/000323315 [PubMed: 21242689]
- Larson GE, & Merritt CR (1991). Can Accidents be Predicted? An Empirical Test of the Cognitive Failures Questionnaire. *Applied Psychology*, 40(1), 37–45. 10.1111/j.1464-0597.1991.tb01356.x
- Loeb PA (1996). *Independent living scales (ILS) manual*. San Antonio: Psychological Corp.
- Lü W, & Wang Z (2018). Associations between resting respiratory sinus arrhythmia, intraindividual reaction time variability, and trait positive affect. *Emotion*, 18(6), 834–841. 10.1037/emo0000392 [PubMed: 29265841]
- MacDonald SWS, Li S-C, & Bäckman L (2009). Neural underpinnings of within-person variability in cognitive functioning. *Psychology and Aging*, 24(4), 792–808. 10.1037/a0017798 [PubMed: 20025396]
- Marson D, & Hebert KR (2006). *Functional Assessment In Attix DK & Welsh-Bohmer KA (Eds.), Geriatric neuropsychology: Assessment and intervention* (pp. 158–197). New York, NY, US: Guilford Publications.
- McAlister C, Schmitter-Edgecombe M, & Lamb R (2016). Examination of Variables That May Affect the Relationship Between Cognition and Functional Status in Individuals with Mild Cognitive Impairment: A Meta-Analysis. *Archives of Clinical Neuropsychology*, acv089. 10.1093/arclin/acv089
- McVay JC, & Kane MJ (2012). Drifting from slow to “D’oh!” Working memory capacity and mind wandering predict extreme reaction times and executive-control errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 525–549. 10.1037/a0025896
- Metternich B, Schmidtke K, & Hüll M (2009). How are memory complaints in functional memory disorder related to measures of affect, metamemory and cognition? *Journal of Psychosomatic Research*, 66(5), 435–444. 10.1016/j.jpsychores.2008.07.005 [PubMed: 19379960]

- Moore RC, Depp CA, Wetherell JL, & Lenze EJ (2016). Ecological momentary assessment versus standard assessment instruments for measuring mindfulness, depressed mood, and anxiety among older adults. *Journal of Psychiatric Research*, 75, 116–123. 10.1016/j.jpsychires.2016.01.011 [PubMed: 26851494]
- Murtha S, Cismaru R, Waechter R, & Chertkow H (2002). Increased variability accompanies frontal lobe damage in dementia. *Journal of the International Neuropsychological Society*, 8(3), 360–372. 10.1017/S1355617702813170 [PubMed: 11939695]
- Myers AM, Holliday PJ, Harvey KA, & Hutchinson KS (1993). Functional Performance Measures: Are They Superior to Self-Assessments? *Journal of Gerontology*, 48(5), M196–M206. 10.1093/geronj/48.5.M196 [PubMed: 8366262]
- Nourhashemi F, Andrieu S, Gillette-Guyonnet S, Vellas B, Albaredo JL, & Grandjean H (2001). Instrumental Activities of Daily Living as a Potential Marker of Frailty: A Study of 7364 Community-Dwelling Elderly Women (the EPIDOS Study). *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 56(7), M448–M453. 10.1093/gerona/56.7.M448
- Powell DJH, McMinn D, & Allan JL (2017). Does real time variability in inhibitory control drive snacking behavior? An intensive longitudinal study. *Health Psychology*, 36(4), 356–364. 10.1037/hea0000471 [PubMed: 28192005]
- Randolph C (2012). *RBANS Update: Repeatable Battery for the Assessment of Neuropsychological Status*. Bloomington, Minn.: Pearson.
- Rapp MA, & Reischies FM (2005). Attention and Executive Control Predict Alzheimer Disease in Late Life: Results From the Berlin Aging Study (BASE). *American Journal of Geriatric Psychiatry*, 13(2), 134–141. 10.1176/appi.ajgp.13.2.134 [PubMed: 15703322]
- Rapp MA, Schnaider-Beri M, Sano M, Silverman JM, & Haroutunian V (2005). Cross-Domain Variability of Cognitive Performance in Very Old Nursing Home Residents and Community Dwellers: Relationship to Functional Status. *Gerontology*, 51(3), 206–212. 10.1159/000083995 [PubMed: 15832049]
- Raskin S, Buckheit C, & Sherrod C (2010). *Memory for intentions test*. Lutz: Psychological Assessment Resources.
- Richardson ED, Nadler JD, & Malloy PF (1995). Neuropsychologic prediction of performance measures of daily living skills in geriatric patients. *Neuropsychology*, 9(4), 565–572. 10.1037/0894-4105.9.4.565
- Riediger M, Wrzus C, Klipker K, Müller V, Schmiedek F, & Wagner GG (2014). Outside of the laboratory: Associations of working-memory performance with psychological and physiological arousal vary with age. *Psychology and Aging*, 29(1), 103–114. 10.1037/a0035766 [PubMed: 24660800]
- Rodriguez C, Ruggero CJ, Callahan JL, Kilmer JN, Boals A, & Banks JB (2013). Does risk for bipolar disorder heighten the disconnect between objective and subjective appraisals of cognition? *Journal of Affective Disorders*, 148(2–3), 400–405. 10.1016/j.jad.2012.06.029 [PubMed: 22884811]
- Royall DR, Lauterbach EC, Kaufer D, Malloy P, Coburn KL, & Black KJ (2007). The cognitive correlates of functional status: A review from the committee on research of the American neuropsychiatric association. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 19, 249–265. 10.1176/jnp.2007.19.3.249. [PubMed: 17827410]
- Schmidt C, Collette F, Cajochen C, & Peigneux P (2007). A time to think: Circadian rhythms in human cognition. *Cognitive Neuropsychology*, 24(7), 755–789. 10.1080/02643290701754158 [PubMed: 18066734]
- Schmitter-Edgecombe M, & Farias ST (2018). Aging and everyday functioning: measurement, correlates and future directions In Smith GE & Farias ST (Ed.). *APA Handbook of Dementia* (pp. 187–218). Washington, DC: American Psychological Association.
- Schmitter-Edgecombe M, McAlister C, & Weakley A (2012). Naturalistic assessment of everyday functioning in individuals with mild cognitive impairment: The day-out task. *Neuropsychology*, 26(5), 631–641. 10.1037/a0029352 [PubMed: 22846035]
- Schmitter-Edgecombe M, Parsey C, & Cook DJ (2011). Cognitive Correlates of Functional Performance in Older Adults: Comparison of Self-Report, Direct Observation, and Performance-

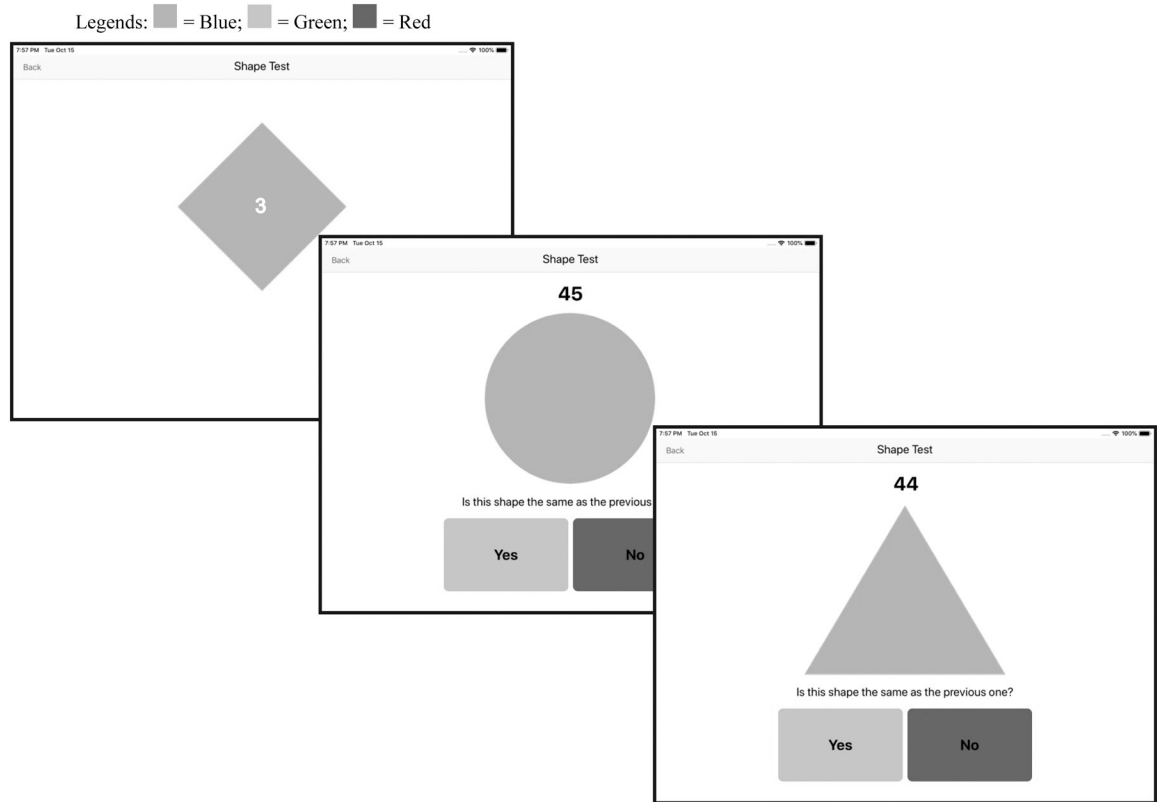
- Based Measures. *Journal of the International Neuropsychological Society*, 17(5), 853–864. 10.1017/S1355617711000865 [PubMed: 21729400]
- Schmitter-Edgecombe M, Parsey C, & Lamb R (2014). Development and psychometric properties of the instrumental activities of daily living: compensation scale. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 29(8), 776–792. 10.1093/arclin/acu053 [PubMed: 25344901]
- Schmitter-Edgecombe M, Woo E, & Greeley DR (2009). Characterizing multiple memory deficits and their relation to everyday functioning in individuals with mild cognitive impairment. *Neuropsychology*, 23(2), 168–177. 10.1037/a0014186 [PubMed: 19254090]
- Schweitzer P, Husky M, Allard M, Amieva H, Pérès K, Foubert-Samier A, ... Swendsen J (2017). Feasibility and validity of mobile cognitive testing in the investigation of age-related cognitive decline: Feasibility and Validity of Mobile Cognitive Testing in the Elderly. *International Journal of Methods in Psychiatric Research*, 26(3), e1521 10.1002/mpr.1521
- Shiffman S, Stone AA, & Hufford MR (2008). Ecological Momentary Assessment Annual Review of Clinical Psychology, 4(1), 1–32. 10.1146/annurev.clinpsy.3.022806.091415
- Sliwinski MJ, Mogle JA, Hyun J, Munoz E, Smyth JM, & Lipton RB (2018). Reliability and Validity of Ambulatory Cognitive Assessments. *Assessment*, 25(1), 14–30. 10.1177/1073191116643164 [PubMed: 27084835]
- Small BJ, Jim HSL, Eisel SL, Jacobsen PB, & Scott SB (2019). Cognitive performance of breast cancer survivors in daily life: Role of fatigue and depressed mood. *Psycho-Oncology*. 10.1002/pon.5203
- Smith G, Del Sala S, Logie RH, & Maylor EA (2000). Prospective and retrospective memory in normal ageing and dementia: A questionnaire study. *Memory*, 8(5), 311–321. 10.1080/09658210050117735 [PubMed: 11045239]
- Sullivan KL, Woods SP, Bucks RS, Loft S, & Weinborn M (2018). Intraindividual variability in neurocognitive performance is associated with time-based prospective memory in older adults. *Journal of Clinical and Experimental Neuropsychology*, 40(7), 733–743. 10.1080/13803395.2018.1432571 [PubMed: 29463204]
- Tabert MH, Albert SM, Borukhova-Milov L, Camacho Y, Pelton G, Liu X, ... Devanand DP (2002). Functional deficits in patients with mild cognitive impairment: Prediction of AD. *Neurology*, 58(5), 758–764. 10.1212/WNL.58.5.758 [PubMed: 11889240]
- Tiplady B, Oshinowo B, Thomson J, & Drummond GB (2009). Alcohol and Cognitive Function: Assessment in Everyday Life and Laboratory Settings Using Mobile Phones. *Alcoholism: Clinical and Experimental Research*, 33(12), 2094–2102. 10.1111/j.1530-0277.2009.01049.x
- Tucker-Drob EM (2011). Neurocognitive functions and everyday functions change together in old age. *Neuropsychology*, 25(3), 368–377. 10.1037/a0022348 [PubMed: 21417532]
- Unsworth N (2015). Consistency of attentional control as an important cognitive trait: A latent variable analysis. *Intelligence*, 49, 110–129. 10.1016/j.intell.2015.01.005
- Unsworth N, Brewer GA, & Spillers GJ (2012). Variation in cognitive failures: An individual differences investigation of everyday attention and memory failures. *Journal of Memory and Language*, 67(1), 1–16. 10.1016/j.jml.2011.12.005
- Vasquez BP, Binns MA, & Anderson ND (2018). Response Time Consistency Is an Indicator of Executive Control Rather than Global Cognitive Ability. *Journal of the International Neuropsychological Society*, 24(5), 456–465. 10.1017/S1355617717001266 [PubMed: 29208077]
- Waters AJ, & Li Y (2008). Evaluating the utility of administering a reaction time task in an ecological momentary assessment study. *Psychopharmacology*, 197(1), 25–35. 10.1007/s00213-007-1006-6 [PubMed: 18064440]
- Wilhelm O, Hildebrandt A, & Oberauer K (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, 4 10.3389/fpsyg.2013.00433
- Wilhelm O, Witthöft M, & Schipolowski S (2010). Self-Reported Cognitive Failures: Competing Measurement Models and Self-Report Correlates. *Journal of Individual Differences*, 31(1), 1–14. 10.1027/1614-0001/a000001
- Williams DP, Thayer JF, & Koenig J (2016). Resting cardiac vagal tone predicts intraindividual reaction time variability during an attention task in a sample of young and healthy adults: Vagal

tone and reaction time variability. *Psychophysiology*, 53(12), 1843–1851. 10.1111/psyp.12739 [PubMed: 27658566]

Wilson BA (Ed.). (1996). *Behavioural assessment of the dysexecutive syndrome: BADS*. London: Pearson.

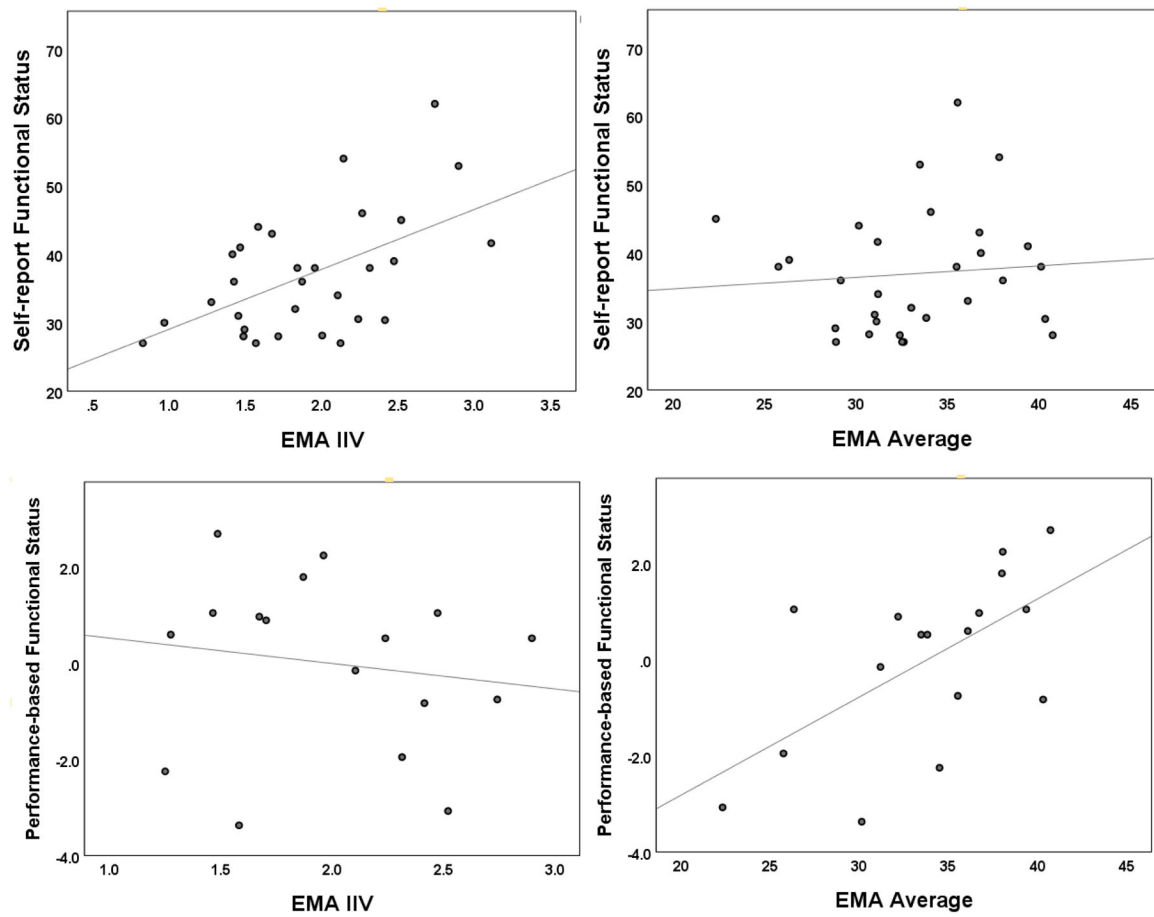
Woods SP, Weinborn M, Velnoweth A, Rooney A, & Bucks RS (2012). Memory for Intentions is Uniquely Associated with Instrumental Activities of Daily Living in Healthy Older Adults. *Journal of the International Neuropsychological Society*, 18(1), 134–138. 10.1017/S1355617711001263 [PubMed: 22032776]

Zimmerman SI, & Magaziner J (1994). Methodological issues in measuring the functional status of cognitively impaired nursing home residents: the use of proxies and performance-based measures. *Alzheimer Disease and Associated Disorders*, 8 Suppl 1, S281–290. [PubMed: 8068270]



**Figure 1.** Example screenshot stimuli sequence (including countdown shape) for the 1-back shape comparison n-back task.





**Figure 2.**

Scatterplots illustrating contrasting findings between IIV (left graphs) and EMA average (right graphs) n-back measure associations with self-report (top graphs) and performance-based (bottom graphs) measures of functional status. Note: IIV = Intra-individual variability; EMA = ecological momentary assessment

**Table 1.**

Means, Standard Deviations and Ranges for the Clinical and N-back Task Data

	Sample Size	Range	Mean	SD
<u>General Cognitive Status</u>				
RBANS <sup>a</sup>				
Immediate Memory	35	69 – 126	102.37	13.65
Visuospatial Construction	35	58 – 112	87.43	14.54
Language	35	85 – 127	102.09	11.97
Attention	35	72 – 138	102.57	15.85
Delayed Memory	35	68 – 130	100.94	12.29
Global Cognitive Status (total)	35	80 – 122	98.09	11.20
<u>Objective Measures</u>				
Executive Functioning: D-KEFS				
Letter Fluency	35	18 – 58	39.66	9.66
Design Fluency	35	14 – 43	25.29	7.23
Executive Composite <sup>b</sup>	35	-2.70 – 3.24	.00	1.57
Prospective Memory				
Hidden Objects	32	0 – 3	2.13	.91
Activity Paradigm	32	0 – 8	6.26	2.59
Prospective Memory Total	32	0 – 11	8.39	3.01
Performance-based Functional Status				
ILS Managing Home	17	20 – 29	26.53	2.67
ILS Health & Safety	17	31 – 38	34.06	2.22
Functional Status Composite <sup>b</sup>	17	-3.37 – 2.70	.00	1.79
<u>Self-Report Measures</u>				
Cognitive Failures				
DEX <sup>c</sup>	30	1 – 28	11.19	7.82
PRMQ <sup>c</sup>	30	20 – 51	33.87	7.20
Cognitive Failures Composite <sup>b</sup>	30	-3.10 – 4.40	.00	1.84
Self-report Functional Status				
IADL-C <sup>c</sup>	30	27 – 62	36.99	8.86
<u>N-back Task Variables</u>				
Clinical Average <sup>d</sup>	35	11 – 33	23.44	5.55
Asymptote <sup>e</sup>	35	24 – 40	33.53	4.14
EMA Average	35	22 – 41	33.19	4.30
IIV <sup>c</sup>	35	.83 – 3.11	1.86	.53

*Note.* RBANS = Repeatable battery for the assessment of neuropsychology status; D-KEFS = Delis-Kaplan Executive Function System; PM = Prospective Memory; DEX = Dysexecutive Questionnaire; PRMQ = Prospective and Retrospective Memory Questionnaire; ILS = Independent Living Scale; IADL-C = Instrumental Activities of Daily Living: Compensation; EMA = Ecological Momentary Assessment; IIV = intra-individual variability

<sup>a</sup>Presented as scaled scores,

<sup>b</sup>Presented as z-scores,

<sup>c</sup>Higher score = poorer performance,

<sup>d</sup>n-back practice trials 1–10,

<sup>e</sup>n-back practice trials 91–100

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2:**

## N-back Variable Correlations with Demographics and Global Cognitive Status

	N-back variables			
	Clinical Average	Asymptote	EMA Average	IIV
N-back Task Variables				
Clinical Average	-			
Asymptote	.44**	-		
EMA Average	.40*	.81**	-	
IIV	-.16	-.05	-.11	-
Demographics				
Age	.15	-.24	-.15	-.17
Education	.49**	.34*	.37*	-.20
Sex <sup>a</sup>	.00	.15	.10	.07
Global Cognitive Status				
RBANS total	.44*	-.07	-.01	-.12

*Note.* EMA = ecological momentary assessment; IIV – intra-individual variability; RBANS = Repeatable Battery for the Assessment of Neuropsychology Status

<sup>a</sup>Spearman's rho correlations conducted for sex

\*  $p < .05$ ;

\*\*  $p < .005$

**Table 3:**

EMA n-back Variable Correlations with Objective and Subjective Clinical Measures

Construct	Objective			Subjective	
	Executive Composite	Prospective Memory <sup>a</sup>	Performance-based Functional Status <sup>b</sup>	Self-report Cognitive Failures <sup>c</sup>	Self-report Functional Status <sup>c</sup>
N-back Variables					
EMA Average	.17	-.16	.60**	-.14	.09
IIV	.06	-.42*	-.15	.53**	.54**

*Note.* EMA = ecological momentary assessment; IIV - intra-individual variability. See methods section and Table 1 for description of tests used to define each construct.

<sup>a</sup>  $n = 32$

<sup>b</sup>  $n = 17$

<sup>c</sup> higher score represents greater self-reported difficulties;  $n = 30$

\*  $p < .05$ .

\*\*  $p < .01$ .

**Table 4.**

Summary of Hierarchical Regression Results with Global Cognitive Status and EMA n-back Variables as predictors of the Objective and Subjective Clinical Measures

Variables	Dependent Measures				
	Executive Composite	Objective Prospective Memory <sup>a</sup>	Performance-based Everyday Functioning <sup>b</sup>	Self-report Cognitive Failures <sup>c</sup>	Self-report Functional Status <sup>c</sup>
<i>Model 1</i>					
RBANS total	.39*	.23	-.03	-.12	-.25
EMA average	.17	-.16	.61*	-.14	.08
Total R <sup>2</sup>	.18*	.08	.37*	.03	.07
<i>Model 2</i>					
RBANS total	.41*	.19	-.02	-.03	-.14
EMA average	.19	-.20	.63*	-.07	.16
IIV	.13	-.43*	.07	.51**	.53**
Change in R <sup>2</sup>	.02	.18*	.01	.25**	.27**
Total R <sup>2</sup>	.20	.26	.37	.28	.34

*Note.* Standardized Beta Coefficients presented for predictors. RBANS = Repeatable Battery of Neuropsychological Status; EMA = ecological momentary assessment; IIV – intra-individual variability. See methods section and Table 1 for description of tests used to define each construct.

<sup>a</sup>  
*n* = 32

<sup>b</sup>  
higher score represents greater self-reported difficulties; *n* = 30

<sup>c</sup>  
*n* = 17

\*  
*p* < .05.

\*\*  
*p* < .01.