2021

# Cognition 'In the Wild': using smartphones to assess cognitive variability in healthy adults and individuals with Parkinson's disease

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

# COGNITION 'IN THE WILD': USING SMARTPHONES TO ASSESS COGNITIVE VARIABILITY IN HEALTHY ADULTS AND INDIVIDUALS WITH PARKINSON'S DISEASE

by

**EMMA L. WEIZENBAUM**

B.A., Wesleyan University, 2013
M.A., Boston University, 2016

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2021

Approved by

First Reader ______________________________________________

Alice Cronin-Golomb, Ph.D.
Professor of Psychological and Brain Sciences

Second Reader ______________________________________________

Daniel Fulford, Ph.D.
Assistant Professor of Occupational Therapy & Rehabilitation Science
Adjunct Assistant Professor of Psychological and Brain Sciences

Third Reader ______________________________________________

Vijaya Kolachalama, Ph.D.
Assistant Professor of Computational Biomedicine

## ACKNOWLEDGMENTS

I wish to thank my advisors Alice and Dan who have each provided invaluable support and guidance in many forms. You each had the ability to see the big picture, even when I found it fuzzy, and I am so grateful to have had you both as advocates for me and my training at each step of this process. To my family of labmates in the AMPLab and Vision & Cognition Lab who have been a tireless sounding board for me these last few years and have helped me to grow as a critical thinker and a teammate. A special thank you to Emma, Anup, Aditya, and Phil who each were central to making this project come to life with their incredible help in data collection and data management. I am grateful to my cohort, Julianne, Esther, Emily, Libby, and Kat, who have shown me a way to celebrate life along this path of clinical science training, all while pushing boundaries within myself and the field. Thank you to my family for showing me a quiet and constant sense of perspective. Throughout my graduate career I have learned the power of flexibility, and I thank those I love for showing me the unexpected surprises that can follow with a touch of flexible thinking. Thank you to my parents who taught me from the beginning to lean in with curiosity and to ask lots of follow up questions. To my sister who has shown love and support by working alongside me as she has lived near and far. A special thank you to Nana whose meticulous tracking of the specifics of my academic career has provided me with enduring companionship and reminded me the value of my work. Lastly, thank you to my participants for their dedication to science, without their efforts this body of research could not exist.

**COGNITION ‘IN THE WILD’: USING SMARTPHONES TO ASSESS COGNITIVE VARIABILITY IN HEALTHY ADULTS AND INDIVIDUALS WITH PARKINSON’S DISEASE**

**EMMA L. WEIZENBAUM**

Boston University Graduate School of Arts and Sciences, 2021

Major Professor: Alice Cronin-Golomb, PhD, Professor of Psychological and Brain Sciences

ABSTRACT

The prevalence of neurodegenerative disorders affecting cognition is rising as the population ages, underscoring the need for accessible methods of assessment. Repeated remote testing in one’s context may reveal cognition’s dynamics and determinants and advance ecologically valid precision medicine.

The hypotheses of Study 1 were that cognition would vary with context, and that individual differences would predict patterns of cognitive variability. 177 young adults (mean age 19.2, 100 women) were prompted via smartphone app 5x/day for 10 days to complete brief cognitive tasks and surveys about their context. Psychometric analyses showed that the smartphone working memory task, more than the executive function task, had greater between-subject reliability and within-subject variability in performance. Multi-level modeling revealed no contextual variables that were directly predictive of smartphone performance. There was indication of significant between person differences in the association between momentary motivation and working memory, but no between-person moderators of this relation were identified.

Study 2 proposed the same hypotheses as Study 1 while assessing the feasibility of smartphone cognitive assessment in Parkinson's disease (PD) and convergent validity between performance on the smartphone tasks and traditional neuropsychological tests. Twenty-seven participants with mild-moderate PD (mean age 63.2, 13 women) completed surveys and games 5x/day for 10 days. The results supported both hypotheses. Response rate to prompts was high, demonstrating feasibility of the approach. Between-subject reliability was high on both games. Within-subject variability was higher for working memory. Multi-level modeling indicated that performance was better on smartphone working memory when participants reported being "home" and having recently exercised, particularly in individuals who endorsed exercise less frequently. Multi-level modeling also revealed that less self-reported daytime sleepiness, and lower PD symptom burden, predicted a stronger association between time of day and smartphone test performance. Strong convergent validity was seen between traditional tests and smartphone working memory but not executive function.

Together these findings support the use of repeat smartphone assessments to understand how context affects cognitive performance. Further development of this assessment method could be useful in increasing sensitivity and specificity regarding cognitive dysfunction, and ultimately lead to personalized treatment recommendations.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| ADHD | Attention Deficit Hyperactivity Disorder |
| BOLD | Blood Oxygen Level-Dependent [response] |
| BP | Between Person [variability] |
| EMA | Ecological momentary assessment |
| ESM | Ecological sampling method |
| GPS | Global Positioning System |
| H&Y | Hoehn & Yahr motor stage |
| HIPAA | Health Insurance Portability and Accountability Act of 1996 |
| mindLAMP | mind Learn Assess Manage Prevent [app] |
| MLM | Multilevel models |
| MoCA | Montreal Cognitive Assessment |
| NMS | Non-motor symptoms |
| PD | Parkinson's disease |
| PDA | Personal Digital Assistant |
| UPDRS | Unified Parkinson's Disease Rating Scale |
| WMS-III | Wechsler Memory Scale 3rd edition |
| WP | Within Person [variability] |

# CHAPTER ONE: GENERAL INTRODUCTION

Thinking occurs dynamically. From day to day and moment to moment, our ability to hold information in mind or regulate attention is constantly in a state of flux. The sources of these fluctuations are varied—whether coming in the wake of mid-afternoon sleepiness, an anxiety-provoking presentation, or the confusion of navigating a crowded store, cognitive performance is influenced by the contexts in which a person operates. Despite this inherent variability, researchers and clinicians often assume that cognitive performance reflects internal processes that are constant and stable. This assumption leads to characterizations of cognitive performance as static, treating the "mean" of an individual or group's cognitive performance as the output of internal processes and dismissing variance in performance as "noise." A body of research has found that these within-person changes in performance, commonly referred to as *intraindividual cognitive variability*, can serve as highly sensitive markers of cognitive dysfunction (Sonuga-Barke & Castellanos, 2007), such as in ADHD (Kofler et al., 2014) and cognitive decline (Roalf et al., 2016) or dementia (Christ et al., 2018; Hultsch et al., 2000). Intraindividual variability has been used to describe change or dispersion across domains of function (e.g., memory vs. attention) (Hilborn et al., 2009), variability or inconsistency in response within a single measurement timepoint (Hultsch et al., 2002), or across a number of time-points of varying scales (e.g., moment to moment fluctuations) (Ram & Gerstorf, 2009). In this paper we seek to identify the factors that give rise to changes in attention and working memory (given these domains' susceptibility to fluctuation) (Braver, 2012) *within individuals*, *across time* (e.g., 1-2 weeks) and *contexts* (e.g., physical environment, emotional state, etc.).

To validly measure how contexts shape attention and working memory, dense sampling of cognitive performance across time and in various settings is essential. Measuring performance across a series of time-points in traditional lab or clinic settings is arduous and expensive, and represents a limited and unnatural context for cognitive function. Instead, by using a an experience sampling method (ESM) or ecological momentary assessment (EMA) of frequent testing over the course of days or weeks (Csikszentmihalyi & Larson, 1987), high resolution measurements can be obtained, which give rise to a rich picture of the dynamics and patterns of an individual's cognitive functioning within a period of time (Sliwinski, 2014). Until recently, methods of studying frequency and patterns of behavior in daily life relied on pen-and-paper response or frequent calls to participants that were burdensome and intrusive. However, the increasing ubiquity of mobile devices (e.g., smartphones, tablets) allows for feasible acquisition of high frequency data (Sliwinski et al., 2018; Trull & Ebner-Priemer, 2013), which can be collected in vivo, without the limitations of retrospective reporting. Mobile assessment data, in turn, produces a “digital phenotype” (Torous et al., 2016) – a reflection of a person’s moment-to-moment cognitive and behavioral function in the context of everyday life. The goal of digital phenotyping aligns with the advancement of precision medicine, where personalized models predict function based on relevant internal and environmental variables. The use of this method has become increasingly common in the fields of clinical psychology and psychiatry to understand the temporal patterns of symptoms (Torous et al., 2019; Trull & Ebner-Priemer, 2013), and how contextual factors, such as a person’s physical and social environment, influence important clinical features such as mood dysregulation (Sandstrom et al., 2017). Digital phenotyping is just beginning to be validated within the field of neuropsychology as digital cognitive assessments are being developed and implemented (Germine et al., 2019; Parsey & Schmitter-Edgecombe, 2013; Sliwinski et al., 2018).

Based on the research available to date, we propose that significant strides could be made in predicting meaningful health outcomes by assessing the everyday factors that influence cognition via smartphone or other mobile technologies. In this narrative review we begin by describing the importance of measuring intraindividual cognitive variability, focusing on the domains of attention and working memory. To better understand the factors that give rise to this variability, we explore how internal state-based variables (affect, motivation) and external-contextual variables (time of day, surrounding noise, activity) affect cognitive performance. We propose using mobile assessment as a means towards integrating ambulatory cognitive assessment tasks with contextual and environmental data to produce digital phenotypes that could aid in diagnostic precision.

**Intraindividual variability in cognition**

Intraindividual variability can be observed at multiple time scales (Nesselroade, 2004), and yet the majority of literature has focused on trial-to-trial response variability within a single timepoint (Braver, 2012; Dinstein et al., 2015; Kelly et al., 2008; Stuss et al., 2003; Vasquez et al., 2017). There is some evidence to suggest that intraindividual variability within and across days is positively correlated, signaling potentially similar processes, although the extent of this generalization is unclear (Rabbitt et al., 2001).

Broadly defined, variability in attention and inhibition of behavioral response is thought to be supported by the prefrontal cortex, which serves to differentially activate the strength of specific networks depending on factors of context, state, and task demands (Kelly et al., 2008). One study found the regulation of task performance and adaptive modulation of attention to be localized to the right dorsolateral frontal regions that exert top-down attentional control (Stuss et al., 2003). Cognitive variability (i.e., performance across course of a day) in attention and memory is known to increase in late-adulthood (Hertzog et al., 1992; Hultsch et al., 2002). Although the function and mechanisms of this increase are not fully understood (Braver, 2012), it seems to occur in both normal aging and various disease processes (Troyer et al., 2016), resulting in decreased frontal-executive function (Diehl et al., 2014), and also an observed decrease in dopamine response within extrastriatal regions (MacDonald et al., 2009). Given that a non-negligible amount of response variability can occur from time-point to time-point (Rabbitt et al., 2001) this may limit the generalizability of scores created from aggregate measures that are interpreted to reflect global functioning across many time-points (Fisher et al., 2018). Instead of dismissing the error or noise in these estimates, it is valuable to understand contributors that give rise to the variability seen within one’s performance - specifically, how internal state-based and environmental factors shape one’s cognitive performance from moment to moment.

**Modifying Factors of Moment-to-Moment Cognition**

Traditional approaches to measuring cognition seek to remove "noise" that might impact an individual's performance, such as time of day and distractors in the physical environment. However, in the context of daily life, these factors likely exert meaningful influence on one's overall function. With the emergence mobile technology, we now have the opportunity to measure these factors in real-time and quantify their influence on cognitive function. With few prior studies measuring the influence of contextual factors on repeat cognitive performance (Allard et al., 2014; A. A. M. Bielak et al., 2017; Dirk & Schmiedek, 2016; Salthouse & Berish, 2005; Waters et al., 2014), whether these factors produce reliable and meaningful patterns of variability remains equivocal. If consistent patterns of within-person variability at either the nomothetic or idiographic level are evident, this could signal opportunities for uniquely tailored real-time intervention points. The potential modifiers of cognitive performance that we will focus on in the following section include internal state-based factors (e.g., affect, motivation) and external contextually-based factors (e.g., time of day, surrounding noise, and recent activity level)(see Table 1). Additionally, we will discuss how each of these modifiers could be more optimally measured in relation to cognitive performance using mobile assessment.

Internal State-Driven Modifiers of Cognitive Performance

In this section we will cover how a handful of previous studies have measured individuals' state affect, either through self-report or experimental-manipulation, and how it relates to cognitive performance on specific tasks of attention, working memory, and recall. Secondly we examine how momentary motivation, measured either from self-report or via reward manipulations, influences how well people perform on tasks of decision making and working memory. The studies described in this section are listed in Table 1.

**Table 1. Internal and External Modifiers of Cognitive Variability**

| Author/year | Year | Moderator | Sample | Test Location and Task Domain | Length | Result |
|---|---|---|---|---|---|---|
| Jefferies et al. | 2008 | Mood | 100 Younger adults | In-lab; visual attention | 1 day | Low arousal with negative affect associated with best performance |
| Ellis et al. | 1984 | Mood | 160 Younger adults | In-lab; semantic recall | 1 day | Depressed mood reduced semantic processing, interaction btwn depressed mood and task difficulty |
| Brose et al. | 2012 | Mood & Motivation | 101 Younger adults | In-lab; working memory | 100 days | Negative affect and low task-motivation reduced working memory performance |
| Brose et al. | 2014 | Mood & Motivation | 101 Younger adults | In-lab; working memory | 100 days | Positive affect and high-task motivation improved working memory performance |
| Sliwinski et al. | 2006 | Mood | 184 Younger & older adults | In-lab; working memory | 1-2 wks | Higher stress ratings associated with slower response time on working memory tasks, greater effect in older adults |
| Salthouse & Berish. | 2005 | Mood | 420 Adults | Palm-pilot; reaction time | 7 days | No relation btwn mood and reaction time scores |
| Krawczyk | 2007 | Motivation | 16 Younger adults | In-scanner; working memory | 1 day | Modulation of reward potential correlates with response time and fMRI BOLD response |
| Yeo & Neal | 2004 | Motivation | 99 Younger adults | In-lab; executive function | 1 day | Motivation had strongest influence on multi-step task performance once task was learned and familiar |
| van der Heijden et al. | 2010 | Time of Day | 2,167 Children | In-lab; sustained attention | 1 day | Performance on sustained attention slower but more accurate in morning |
| Manly et al. | 2002 | Time of Day | 10 Younger adults | In-home; sustained attention | 4 days | Correlation btwn errors and sleepiness rating |
| West et al. | 2002 | Time of Day | 40 Younger & older adults | In-lab; computer task | 4 days | Younger adults performed best on WM tasks in evening, older adults best in morning |
| Schmidt et al. | 2007 | Time of Day | Adults | Review paper | | Time of day has impact on memory, executive functioning, and attention; effects vary btwn individuals |
| Banbury et al. | 2001 | Noise | Adults | Theoretical review paper | | Short-term memory most affected by sound irregularity compared to other audio dimensions |
| Lange | 2005 | Noise | 34 Younger adults | In-lab; working memory | 1 day | Noise disrupted visuospatial but not verbal working memory performance |
| Bell & Buchner | 2007 | Noise | 182/193 Younger & older adults | In-lab; working memory | 1 day | Same level of impairment on visual working memory from noise vs. silence in younger and older adults |
| Ljungberg & Neely | 2007 | Noise | 24 Adults | In-lab; reasoning & working memory | 1 day | No sig. effect of noise on performance; but higher levels of subjective task difficulty and stress ratings |
| Sibley & Beilock | 2007 | Activity | 48 Younger adults | In-lab; working memory | 2 days | Cardiovascular exercise significantly improves working memory |
| Whitbourne et al. | 2008 | Activity | 59 Younger adults | Daily diary; subjective complaints | 8 days | Older adults report fewer memory failures on days with exercise |
| Phillips et al. | 2016 | Activity | 51 Older adults | In-lab; reasoning and processing speed tasks | 5 days | Physical activity accounted for sig. within-person variance in cognition esp. processing speed |
| Allard et al. | 2014 | Activity | 60 Older adults | PDA; semantic memory task | 7 days | Intellectually-stimulating activities improved semantic memory performance measured later in same-day |
| Bielak et al. | 2017 | Activity | 146 Older adults | Web-based; processing speed, memory, and reasoning | 7 days | Faster memory and processing speed on days with individual/small-group social activities |

**Affect**. The valence (i.e., positive or negative) of one's momentary affect has been shown to impact cognitive performance in the domains of attention, working memory, and recall. Ellis and colleagues' (1984) seminal work demonstrated that when negative affect was induced in healthy undergraduates, using self-referential negative statements, word recall was reduced compared to those in a neutral affect condition, particularly on challenging trials. The investigators proposed that in states where strong affect is present, attention is diverted and fewer cognitive resources are allocated for the task at hand; the effects of this become most apparent on difficult tasks requiring greater cognitive resources (Ellis, 1984). A separate study of cognition and affect found that visual attention (target detection) varied in relation to the combination of level of arousal and affect valence in a sample of healthy adults. The primary finding was that low arousal-negative affect (e.g., sadness) was associated with greater accuracy on second but not first target detection, indicating that sadness enhanced attention prioritization, but not overall improvement of attention (Jefferies et al., 2008). It may be that affect serves to shape the strength of attentional control or allocation (Stuss et al., 2003).

In a study in which self-reported affect and working memory were assessed daily over a period of 6 months, Brose and colleagues (2012) found poorer working memory performance on days with greater negative affect. Conversely, on days when positive affect was higher, working memory performance was improved, which also related to higher task-related motivation (Brose et al., 2014). In line with these findings, Sliwinski et al. (2006) identified a within-person association between higher daily stress ratings and slowed response time on a working memory task; this effect was particularly pronounced in older adults. However, one of the first studies to use repeat cognitive assessment of attention with a reaction time task using Palm Pilots (prompted 100x over 7 days) found no association between momentary affect ratings and reaction times (Salthouse & Berish, 2005).

Taken together, there is evidence from the existing literature that suggests negative affect may reduce working memory and semantic recall abilities, at least in controlled settings. However, considerable heterogeneity in main effects and interaction effects suggest that there is still much to be learned about the role of affect in cognitive performance, particularly when examined outside of the lab context and in clinical populations where affect dysregulation may be a primary symptom. In this new digital era, the consistency with which people carry smartphones could be useful in allowing for quick, momentary probes about an individual’s current affect as it changes in daily life. Such assessments of affect could then be paired with smartphone cognitive assessments of various types (attention, working memory, recall) to provide a more detailed picture how people’s affect varies with task performance.

**Motivation**. Motivation drives the allocation of effort, which in turn affects cognitive performance (Katzell & Thompson, 1990). Measuring the influence of motivation has long been a core feature of clinical neuropsychological assessment with the inclusion of measures designed to assess effort and engagement as checks for validity in testing (Bigler, 2012). However, while these standard measures of effort capture motivation in a given moment in a lab or clinic, in daily life motivation is varied and depends on factors such as affect, the given rewards or costs of a task, and practice. Ellis et al. (1984) theorized that strong affect decreased performance by diverting attentional resources to the affective experience vs. the task at hand. Adding nuance to this explanation, Pessoa (2009) proposed a dual-competition framework of affect, motivation, and executive control. In this framework, both affect and motivation are hypothesized to enhance or impair executive control depending on whether the emotion or locus of reward is aligned or divergent from the task objective. This conceptual model was exemplified through a study by another group, and showed how the modulation of reward potential (i.e., money) on each trial was positively correlated with both behavioral response time and BOLD response (i.e., fMRI-measured neural activation), with the hypothesized mechanism being increased motivation to perform at one's best when there is a relatively high payout (Krawczyk et al., 2007). Another factor underlying motivation's influence on cognitive performance is practice, or task familiarity (Yeo & Neal, 2004). When accuracy was measured across 30 trials of a complex multi-step judgement task, a person's self-reported motivation had a greater positive effect on performance on later trials. In other words, when a task is well-learned, differences in motivation level have more of an effect on accuracy compared to when a task is unfamiliar (Yeo & Neal, 2004). One study examining self-reported motivation and cognitive performance serially over 100 days found that higher daily motivation ratings were positively correlated with higher scores on verbal and spatial working memory tasks (Brose et al., 2014).

While daily retrospective self-reports of motivation have been assessed in relation to working memory (Brose et al., 2012, 2014), what remains unknown is the extent to which naturally-occurring fluctuations in state-based motivation throughout a given day might influence performance on tasks of working memory or other cognitive domains (e.g., processing speed, recall, etc.). Using smartphone assessment, one could probe for real-time self-reported motivation before and/or after a test of cognitive function. Alternatively, tests could include passive or built-in measures of effort that determine the level of engagement in a given task. Simultaneously, sensing metrics, such as time of day and GPS location, could be used to determine the contextual factors that correlate with task engagement. Motivation and cognition could be better studied in tandem via smartphone assessment through the gamification (i.e., point-scoring, performance visualization, in-app rewards) of cognitive tasks that are modeled after smartphone games that participants may have prior familiarity with. In contrast to traditional tests of cognition that provide no immediate reward for the participant, smartphone tasks can easily incorporate points systems, potentially tied to monetary or other rewards, that could allow for observational or experimentally-induced modulation of task engagement.

External Contextual Modifiers of Cognitive Performance

In this section we have selected studies that measure the influence that factors outside of the person have on cognitive performance in various domains including attention, working memory, executive function, and memory. Below I describe studies that seek to understand the impact of a person's environment, specifically time of day (morning, afternoon, evening), the quality and amplitude of environmental noise, and the impact of recent social and physical activity on individuals' task performance. The studies included in this section are also listed in Table 1.

**Time of Day**. Internal circadian rhythms, driven by time of day, have been shown to impact cognitive performance. One study of healthy 10-12 year-olds found that time of day impacted performance on challenging trials of visual working memory and processing speed tasks (van der Heijden et al., 2010). Another study measured working memory in the morning and at night across multiple days, in both younger and older adults. Rather than time of day, working memory performance varied with self-reported alertness, which was higher for younger adults in the evening and older adults in the morning. A review of circadian rhythms and cognitive performance suggested that time of day had a significant effect on a wide range of cognitive tasks including those of attention, executive functioning, and memory; notably, performance fluctuation was linked to individual differences in peak circadian arousal (Schmidt et al., 2007). In sum, this literature suggests there is a relation between cognitive function and time of day, which seems to rely mostly on time of day's connection to intraindividual alertness. Mobile assessment can be particularly useful in relating time of day to cognition using time-stamps captured with assessments that can be scheduled at specific times or in random intervals throughout the course of a day. Further, subjective alertness could be captured through momentary reports of wakefulness at the time of the cognitive assessment. Eventually, passive measurement of alertness may be collected using touch-screen latencies when typing or through smartwatch sensors that measure heart rate and other biometrics.

**Surrounding Noise**. Distractions are inherent in a noisy or chaotic environment, and intuitively impact cognitive performance. Looking to the literature on auditory distractions, the primary area of study has been in the domain of working memory, where a series of findings suggests that the irregularity of sound is most impairing to cognitive performance, rather than the absolute volume of the noise in the environment (Banbury et al., 2001). Of all types of cognitive tasks, auditory working memory tasks involving maintenance of a series or order of information appear to be most affected by auditory distraction (Banbury et al., 2001). Similar to the proposed mechanism for affect and motivation, noise is thought to impair performance when attention is pulled away from the task at hand and towards task-irrelevant stimuli, particularly an irregular non-habituated stimulus (Cowan, 1998). One study demonstrated a particularly large impact of sound on attentional abilities when the distractor and task were similar (e.g., auditory-verbal), as this requires more resources for differentiation and suppression than when the task is more unique from the distractor (e.g., visuospatial) (Lange, 2005). Another study found that while distracting ambient noise did not affect objective performance on working memory and reasoning tasks, the addition of environmental noise was related to significantly higher subjective stress ratings and greater perceived difficulty of a given task (Ljungberg & Neely, 2007). While most studies of noise and cognitive performance have used younger adults in experiments, there has been some investigation into age as moderator of the influence of irrelevant noise on working memory (Bell & Buchner, 2007). When recorded ambient office-noise was played during a visual working memory task, there was a similar level of impairment for both younger and older adults (Bell & Buchner, 2007). All the studies above experimentally constructed a “noise condition” to study its effect on cognitive performance, and yet noise is a product of unpredictable real-world environments that can be challenging to replicate in lab settings. Here, mobile assessment could allow for real-time capture of ambient noise via use of a phone's microphone, or real-time self-reports of the characteristics of a given environment. Moving forward, sensing technology using advanced data analytics (e.g., machine learning algorithms for sound detection) could be helpful in determining the type of audio in a given environment; for example, whether the sounds are of a noisy subway train versus a conversation. These categories of sound could then be classified, measured in duration, and linked to cognitive function outcomes providing individualized assessment results in the context of a person's unique set of daily environments.

**Recent Physical and Social Activity**. The frequency and recency of physical and social activities appear to impact intraindividual variability in cognitive performance. For example, in one study of healthy adults, those with lower baseline working memory capacity performed significantly better on working memory tasks when they engaged in 30 minutes of cardiovascular exercise immediately before the tasks (Sibley & Beilock, 2007). Further, in an 8-day daily-diary study in young, middle-aged, and older adults, older adults reported fewer memory failures on the day of and day after physical activity; there was no effect of physical activity in younger adults (Whitbourne et al., 2008). In another study examining relationships between physical activity and cognitive performance in older adults, no significant correlation was found between same-day or day-to-day average physical activity and cognition, though, physical activity did explain significant variance in within-person performance in certain cognitive domains, most notably processing speed (Phillips et al., 2016). Furthermore, in mental illnesses that largely affect cognition, such as schizophrenia, a recent meta-analysis showed that aerobic exercise in individuals with this condition was associated with improved attention and working memory (Firth et al., 2015). Using smartphone assessment, physical activity can be measured objectively and unobtrusively via accelerometer. This passive measurement could be enhanced with heart rate detection in associated devices such as smartwatches. Recording physical activity using mobile devices has the potential to add precision to our understanding of how duration, intensity, and recency of physical activity may influence subsequent cognitive performance.

Few studies have looked at the impact of one's recent recreational or social activities on moment-to-moment cognitive performance, and existing studies are generally specific to samples of healthy older adults. In one study, 60 older adults were prompted via a personal digital assistant (PDA) five times per day for one week to answer questions about their location and to choose a category of their most recent activity. Additionally, two out of these five daily assessments included a measure of semantic reasoning, where an overall category was selected in relation to a list of words. Results indicated that when participants reported having recently engaged in "intellectual activities" (reading, crossword puzzles, etc.), scores on the semantic reasoning task were significantly higher (Allard et al., 2014). A seven-day study of healthy older adults examined short-term associations between a number of daily activities and performance on online tests of processing speed, memory, and reasoning. Intellectual activities were not a significant correlate of cognitive scores, as seen in the study described above; instead, same-day individual or small group social interaction was associated with higher memory scores and response times. Of note, recent physical activity in this study was not associated with better cognitive performance (A. A. M. Bielak et al., 2017). These findings suggest a need to explore the mechanisms by which recent social interaction improves cognitive performance, and whether or not this finding is limited to older adults or to other specific individual differences.

Social activity could be measured through subjective reports of recent or current activity from smartphone assessments administered throughout the day. Identification of activities could be enhanced and less burdensome through GPS-tagging of specific centers of activity such as school, work, friends' houses, or other locations where socialization takes place (e.g., church, gym, restaurants etc.). Smartphone assessments of cognition can be paired not only with identification of recent activity, but also important subjective ratings such as how important or enjoyable a person found their most recent activity, as this may be even more essential to cognitive performance than the activity type itself.

Limitations of Research on Modifiers of Cognitive Performance

In sum, research to date suggests that both internal (affect and motivation) and external (time of day, surrounding noise, recent activity) variables can impact cognitive performance at any given moment. Although this research has shed light on the question of what gives rise to real-life variability in cognitive function, many of the above studies were limited by traditional methods of data collection. For example, testing cognition in a laboratory or clinical setting, across several time-points, ignores the influence of one's real-world environment, activity levels, and social engagement. Performance was often measured in one cognitive domain or in one age group, which does not allow us to extrapolate results to other cognitive constructs and to people across the lifespan. Some studies relied on retrospective report of contextual factors or subjective cognition, which are vulnerable to biases. To better understand the unique impact that potential internal and external modifying factors can have on intraindividual variability in cognition over time, we will need to improve the quality and ecological validity of our assessments, including gathering data at a higher temporal frequency and in real-time.

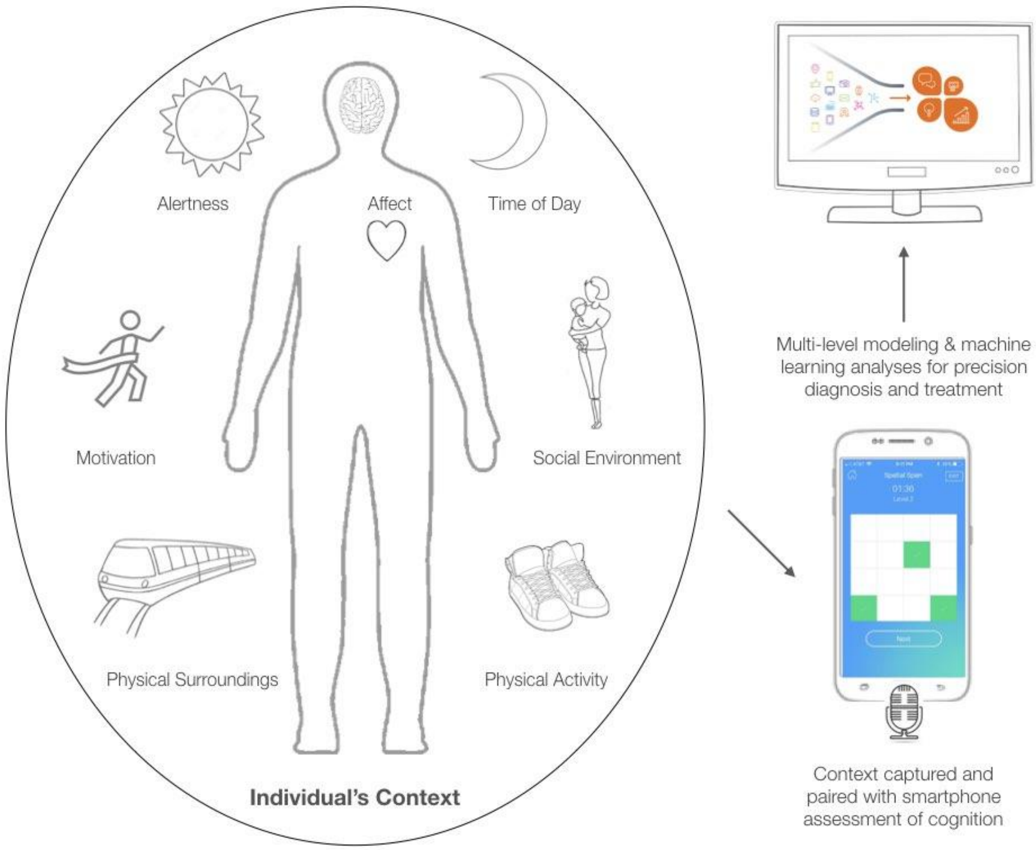
**Studying Intraindividual Variability in Cognition via Mobile Technology: Current Evidence**

From the existing body of research, it is clear that internal and external factors of one's environment contribute to cognitive performance. A number of the studies described above suggest that individual differences play a role in determining which, and to what extent, contextual factors affect cognitive performance. There are relatively few studies that have used repeated measurements across days or weeks to establish temporal relationships between cognitive performance and internal and external modifiers of cognition. Limited research in this area is, at least in part, due to constraints accompanying traditional laboratory or clinic-based research designs. By employing repeated measurement via mobile assessment, there is an opportunity to better understand how cognition and its modifiers temporally co-vary in real-world environments.

Building off research using Ecological Sampling Methods/Ecological Momentary Assessment (Csikszentmihalyi & Larson, 1987; Shiffman et al., 2008), mobile devices have been increasingly used in mental health assessment. Smartphone and tablet applications have the capacity to capture a person's unique fingerprint of response in different domains, ranging from establishing new behavioral habits to symptoms of severe mental illness or progressive neurological disorders. Mobile app-based assessments allow researchers and clinicians to reach people who face barriers to participating in lab-based studies or attending regular mental health care appointments (Bakker et al., 2016). One of the greatest benefits to using mobile devices, specifically smartphones, for ambulatory assessment is capitalizing on their ubiquity: owners carry them on-person most of the time, and check them up to hundreds of times in a given day (Wilcockson et al., 2018).

The incorporation of smartphone-based sensors, such as accelerometer, GPS, and microphone, is a new application of digital assessment that specifically allows for passive measurement of a person’s real-world context and environment. For example, accelerometers capture a proxy for physical activity, which can then be characterized into activity types (sedentary, walking, running), and correlated with metrics of cognitive variability. By identifying the patterns within a person's location data from GPS, we can observe how people interact with and traverse their environments through metrics such as distance and visit frequency over time. By taking advantage of a phone's built in microphone there lies potential to better understand the environmental context in which a person is thinking and behaving from moment to moment (Figure 1.).

**Figure 1. Analyzing cognition in context via smartphone assessment**




Mobile Assessment in Neuropsychology

In the field of neuropsychology, traditional pencil and paper tests have only just begun to be translated to computerized versions of these tests and utilization rates for computerized assessment remain below 10% (Rabin et al., 2014). A review of studies comparing computerized neuropsychological tests with their analog versions found reliability was similar if not improved on computerized versions. However, normative data for traditional tests could not be used interchangeably with the digital versions, given the smaller sample size for digital tests and differences in the sample characteristics (Parsey & Schmitter-Edgecombe, 2013).

Beyond replicating existing tests, computerized versions of neuropsychological tests have yet to capitalize on additional passive or process-based data that is possible with mobile technology. For example, response latencies and adaptive testing methods could be used to identify a precise and unique range of within-person cognitive performance. To date, real-time cognitive data collection via ESM/EMA has yet to be fully embraced by the field of neuropsychology, but it has the potential to inform diagnosis and intervention efforts by capturing patients' fluctuation in cognition (Casaletto & Heaton, 2017). Although digital cognitive assessment is growing exponentially with the use of computers and mobile devices, some have argued that the field of neuropsychology is in the midst of "a technology crisis," and suggest that high-frequency data capture is one of the key methods the field should be leveraging (Miller & Barr, 2017) Koo & Vizer, 2019; Sternin et al., 2019).

A review by Moore et al., 2017 identified 12 studies that used self-administered mobile cognitive assessments ranging from 1 to 5 times per day over 1-14 days, across various age groups and populations. This review reported high adherence rates (80% on average) and strong construct validity, mostly in the domains of attention, working memory, and executive functioning. Notably, very few of these studies reviewed combined smartphone-based cognitive assessment with real-time measurement of other temporally-dynamic contextual variables; and a number of the studies only sampled cognition 1x/day or less. In order for this methodology to assess potential links between variability in cognitive performance as it relates to context and state variables, high frequency of measurement is key.

## Review of High-Frequency Measurement Mobile Assessment Studies

To highlight the existing literature that has deployed high-frequency assessment, 15 studies were selected that used mobile paradigm-based assessments of cognition at a frequency of 2 or more times per day for at least one week (See Table 2). The majority of studies included here were conducted between 2014 and 2019. Six out 15 included a non-clinical sample of young or middle-aged adults, two centered specifically on healthy older adults, and one on children. Five of the 15 studies included samples of substance users or those at risk of smoking or alcohol abuse. Others featured clinical samples including adults with depression and also Parkinson's disease. Eight out of 15 studies, typically the more recent, used a smartphone device for mobile assessment, and five out of the 15 studies used a Personal Digital Assistant (PDA) device, one used a Nokia flip phone and the other used a smartwatch. The majority of studies used working memory test scores as the primary cognitive outcome measure, although other cognitive domains such as verbal and visual memory, attention, processing speed, and motor speed were also measured alone or alongside working memory in these studies. Of the 15 studies, 12 deployed repeat mobile tests for 1-2 weeks, and other durations included 4, 6 and 24 weeks. Six studies prompted cognitive assessments 2-3x per day, seven studies prompted 4-6x per day, and two prompted at an even higher frequency. In addition to cognitive performance, 4 out of 15 studies reported data collection related to momentary affect or mood, three out of 15 recorded reports of recent activity types or social setting, and 3 out of 15 examined time of day or fatigue in relation to cognitive performance. Generally, studies found high levels of concordant validity between mobile cognitive assessments and traditional in-lab measures.

**Table 2. Studies of Repeat Mobile Assessment of Cognition**

| Author | Date | Sample | Assessment Tool | Length | Daily Frequency | Cognitive domain | State/Context Variables | Result |
|---|---|---|---|---|---|---|---|---|
| Allard | 2014 | 60 older adults | PDA | 1 wk | 2x | Semantic reasoning, memory | Location, social setting, recent activities/ behaviors | Cognitive performance improved following intellectually stimulating activities |
| Cormack et al. | 2019 | 30 adults with depression | Apple watch | 6 wks | 3x | Working memory | Mood | High adherence, moderate concordance, no relationship between momentary mood and cognition trajectories |
| Dagum | 2018 | 27 young adults | Smartphone | 1 wk | continuous | Working memory, executive function, languages | Not specified | Digital biomarkers (e.g. taps, swipes) highly correlated with traditional in-lab neuropsych test scores |
| Dirk & Schmiedek | 2015 | 110 8-11 yr olds | Smartphone | 4 wks | 3x | Working memory | Motivation, affect, sleep, physical activity/accelerometer | Greater WM variability measured by phone correlated with lower performance on in-lab cognitive and academic tests |
| Lipmeister et al. | 2019 | 44 Parkinson's pts; 35 controls | Smartphone | 24 wks | 6x | Motor speed | Motor symptoms | Phone tests of motor speed correlated with questionnaire measures and differentiated patients from controls |
| Mendelson et al. | 2014 | 12 meth addicts; 20 controls | in lab computer & smartphone | 2 wks | 2x | Working memory | Not specified | N-Back & Stop Signal on iPhone comparable results to lab test; speech detection on stoop did not work; no btwn group differences |
| Price et al. | 2017 | 21 young adults | Smartphone | 2 wks | 3x | Working memory, attention, and processing speed | Mental fatigue | Fatigue ratings positively correlated with longer reaction times on attention task |
| Salthouse & Berish | 2005 | 420 adults | PDA | 1 wk | 15x | Reaction time | Time of day, mood | Large within-person variability; no sig. relation between time of probe or mood and reaction time |
| Sliwinski et al. | 2018 | 219 adults | Smartphone | 2 wks | 5x | Processing speed and working memory tasks | Not specified | High construct validity, reliability, and within-person variance |
| Schweitzer et al. | 2017 | 114 older adults | Smartphone | 1 wk | 5x | Memory and executive function | Physical environment, social interaction | High adherence and concordance with traditional neuropsych test scores |
| Schuster et al. | 2015 | 39 high risk smoker young adults | PDA | 1 wk | 5-7x | Working memory | Not specified | High feasibility/compliance and construct validity |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Timmers et al. | 2014 | 26 young adults | Smartphone | | 4x | Short-term memory task | Not specified | High concordance btwn smartphone and lab-based tasks |
| Tiplady et al. | 2009 | 38 adults freq. ETOH | Cell phone | 2 wks | 2x | Attention and working memory tasks | Alcohol consumpt ion | Greater errors on phone and lab-based tasks after alcohol consumption |
| Waters et al. | 2008 | 22 smokers, 22 controls | PDA | 1 wk | 4x | Working memory | State anxiety | High adherence, high reliability |
| Waters et al. | 2014 | 119 smokers | PDA | 1 wk | 4x | Attentional Bias | Not specified | Btwn subject craving and lab attn bias associated w/ PDA Stroop attn bias |

Just over half of the studies selected for Table 2 examined additional contextual or internal state-based variables in relation to cognitive performance, and those that did produced mixed findings. Smartphone semantic reasoning and memory scores were greater after recent intellectually stimulating activity (Allard et al., 2014), slower smartphone-recorded reaction times were associated with greater mental fatigue (Price et al., 2017), and increased error rates were seen on phone-based working memory and attention (Tiplady et al., 2009). By contrast, two studies conducted almost 15 years apart found no within-person associations between momentary mood ratings and working memory (Cormack et al., 2019) and  reaction time performance (Salthouse & Berish, 2005). It should be noted that studies of repeat mobile cognitive assessment have used a variety of mobile platforms, different populations, and different ways of measuring the same contextual variables (e.g., scales or dimensions of mood); as such, findings require replication using standardized methods.

Several studies commented on the psychometrics of reliable within-person fluctuations and the need to understand the drivers of these patterns of fluctuations. Dirk & Schmiedek (2016) examined the psychometrics of repeated mobile assessment of cognition in children, and found that moment-to-moment and day-to-day performance had reliable amounts of intraindividual variability that was indicative of individual differences and patterns of response. More recent work by (Sliwinski et al. (2018) suggested that within-person fluctuation in processing speed and working memory across the course of a day was reflective of meaningful, but unknown contextual moderators of the intraindividual variability observed. This signals the need to explore and identify the unknown contextual variables that influence variability in within-person cognitive functioning. Nomothetic approaches to analyzing within-person contextually-based modifiers of cognition may fail to produce reliable or meaningful findings, as the influence of particular contextual factors (e.g. time of day) may be dependent on individual differences. In addition to statistical analysis that uses aggregate measures, idiographic approaches could be used for understanding the temporal dynamics between context, state, and cognitive function of a unique individual, which may be reliably different from another individual (Fisher et al., 2017). Through the use of idiographic analyses such as individualized time-lagged modeling and network analyses, there is an opportunity to develop both personally targeted and ecologically generalizable interventions (Fisher et al., 2018).

**Developing More Sensitive Methods of Assessment in Parkinson's disease**

The rates of neurodegenerative disorders are rising as the population ages, as are costs of assessment and treatment (Hurd et al., 2013). The motor impairment endemic to Parkinson's disease (PD) is an impediment to clinic and lab visits, settings that may not be easily accessible to all (Bradford et al., 2009), and yet are the settings required for current methods of assessment of PD symptoms including non-motor symptoms such as cognitive impairment. These limitations present time and access barriers, resulting in diagnosis for some individuals long after symptoms emerge (Breen et al., 2013) and suboptimal treatment (Hogan et al., 2008). Neuropsychological testing to characterize these cognitive deficits requires hours of time from persons with PD, who often experience daytime fatigue (Friedman et al., 2007), and current methods of testing capture cognitive performance only at a single time-point. This is particularly problematic for persons with PD, which is a disorder marked by cognitive difficulties that fluctuate over time, as well as changes in mood, motivation, and arousal level (Cronin-Golomb et al., 2019). Even within a testing session, our lab recently documented cognitive variability in PD, using a spatial judgment task (Salazar et al., 2019). There is additional evidence to suggest that increased intraindividual variability in processing speed differentiates a sample of people with PD from healthy older adults above and beyond within-person means (Cheon et al., 2009).

Digital technology has the potential to addresses the problem of limited access to neuropsychological assessment in clinical populations by decreasing burden associated with traditional testing and potentially improving rates of early detection (Parsey & Schmitter-Edgecombe, 2013). Newly emerging digital assessments use computerized versions of the lengthy analog tests that provide a snapshot of cognition in a single session. There is evidence that variability in cognitive performance using a repeated assessment, or "burst" design (assessing many times over 1-2 weeks), may be a robust and sensitive marker of cognitive decline (Vaughan et al., 2013). Repeated testing of cognition and state-based variables via a mobile assessment device such as a smartphone allows us to understand disease symptoms' dynamics and determinants, which can inform our ability to predict changes in PD disease state or progression. This novel method of assessment offers the potential of measuring fluctuations in real-world activities of daily living as they relate to the individual. And the ubiquity of smartphones introduces the capability of putting evidence-based assessment into the hands of those who may most benefit from it.

There are few studies to date that have used repeated smartphone assessment in PD, all of them quite recent. Our lab demonstrated the feasibility of conducting a smartphone EMA design in PD, specifically related to sleep (Wu & Cronin-Golomb, 2019). Another study established proof-of-concept using smartphone technology to track intraindividual fluctuations in PD motor symptoms as they related to mood and context through smartphone EMA, and then modeled these fluctuations using network analysis, albeit in one participant (van der Velden et al., 2018). A third open-data study used a brief repeated measures smartphone design to assess PD motor symptoms including objective motor and reaction time testing over six months (Bot et al., 2016; Lipsmeier et al., 2018). Another mobile assessment in study in PD used responses to develop an independent and valid disease severity score using a supervised and weighted machine learning algorithm (Zhan et al., 2018). A limitation of this latter study was that all assessments occurred within the clinic setting and no measures of neuropsychological function beyond reaction time were administered.

To date, no studies have examined patterns of objective neuropsychological testing using smartphone technology in a sample of individuals with PD. It is unknown whether remote, repeated-measures of smartphone assessment of working memory and executive function is feasible in individuals with PD. This is unfortunate given the effects of impairment in cognition on quality of life (Lawson et al., 2016), and the difficulties that individuals experience in accessing traditional methods of neuropsychological assessment (Breen et al., 2013).

## Specific Aims and Hypotheses

**AIM 1: In Chapter 2 I describe the feasibility and reliability of a remotely conducted smartphone assessment of intraindividual cognitive variability using an ecological sampling design in young adults**

*Design:* A sample of 200 undergraduates remotely downloaded and used the mindLAMP app; participants were prompted to complete smartphone assessments of mood, alertness, motivation, and cognition 5x/day for 10 days.

*Hypothesis 1a Feasibility:* A meta-analysis of both one-time and repeated mobile cognitive assessment has shown average adherence to be 79.2% (Moore et al., 2017). I hypothesized that with clear instructions, responsive technological support, and incentives for completing smartphone assessments, a repeated-measures design of cognition using a smartphone app would be feasible. I operationalized “feasible” as a yield of greater than a 10% response rate to smartphone surveys, and hypothesized that the response rate would be closer to 80% as noted in above. A maximum of 90% missing data has been the suggested as the upper limit for estimation using dynamic structural equation modeling (DSEM) of intensive longitudinal survey data (Asparouhov, 2018).

*Hypothesis 1b Reliability:* A study of a similar design found that between-person reliability, averaged across time-points, was high and within-person reliabilities indicated systematic fluctuations across occasions (Sliwinski et al., 2018). I hypothesized I would be able to reliably measure the amount of between- and within-person variability. I expected that between-person variability would be greater than within-person variability.

**AIM 2: In Chapter 2, I evaluated construct validity and assessed associations between context and state variables (measured via smartphone survey of mood, alertness, motivation) and cognition using the mindLAMP app.**

*Design:* A sample of 200 undergraduates completed one-time online questionnaires and smartphone assessments of mood, alertness, motivation, and cognition 5x/day for 10 days.

*Hypothesis 2a Construct validity of smartphone cognitive assessment:* The brief cognitive assessments included in the mindLAMP app that are used in the study, Trails-B and Backward Spatial Span, are designed to measure elements of executive function that include set-shifting, self-monitoring, and working memory. Studies have shown high levels of construct validity with mobile assessments of similar constructs (i.e., working memory and processing speed) (Moore et al., 2017; Sliwinski et al., 2018). The measures used in my study had not yet been psychometrically evaluated for construct validity, and I conducted this needed validity analysis. I hypothesized that participant performance on the mindLAMP app’s cognitive assessments would be significantly correlated with participant self-reports of executive function using the *Behavior Rating Inventory of Executive Functioning (BRIEF-A SR)*, a one-time self-report measure, particularly on the subscales of set-shifting, self-monitoring, and working memory.

**AIM 3: In Chapter 3, I describe how I translated the smartphone assessment of cognition study design to a sample of participants with Parkinson's disease, and examined how clinical disease features relate to design feasibility and to intraindividual cognitive variability as assessed by smartphone**

*Design:* A sample of 27 participants with PD was administered a one-time online set of questionnaires, a one-hour study visit with neuropsychological testing, assessment of disease symptoms, and review of smartphone app features. The following active study period included 10 days of remote smartphone assessments of mood, alertness, motivation, and cognition 5x/day.

*Hypothesis 3a Feasibility:* I hypothesized that the majority of participants would exhibit a completion rate similar to that of the young adult sample (mean >60% completion of EMA assessments). I hypothesized that those with lower baseline cognitive function scores on the global cognitive screening measure, the Montreal Cognitive Assessment (MoCA), may have lower rates of response due to difficulty navigating the smartphone app interface without the assistance of a researcher or clinician.

*Hypothesis 3b Reliability:* As this is a novel design, there are no previous studies that comment on the reliability and quality of data collected via smartphone assessments of cognition and state-based variables in PD. I hypothesized that in PD there would be between-person and within-person variability in scores indicating the need for multi-level modeling to account for non-independent measures in regression analyses.

*Hypothesis 3c Convergent validity*: An emerging literature has demonstrated that smartphone assessments of cognition in clinical samples show strong convergent validity with their analog counterparts that are used in traditional neuropsychological assessment (Dagum, 2018; Pal et al., 2016; Schuster et al., 2016; Sliwinski et al., 2018; Timmers et al., 2014). While the cognitive measures included in the mindLAMP app are similar to tests used in the above studies, they have not been measured alongside the traditional neuropsychological tests (Trail Making Test and WMS-III Spatial Span) from which they are based. I hypothesized that there would be a significant association between scores on the traditional pencil-and-paper testing and smartphone cognitive assessments.

*Hypothesis 3d Clinical features related to intraindividual variability:* I hypothesized that cognitive score would be related to disease symptom burden as assessed in the disease rating scale (UPDRS) interview and on baseline questionnaires (i.e., BRIEF). I also anticipated smartphone cognitive task performance to be related to self-reported ON-OFF periods given that off-periods are associated with an increase in motor symptoms including tremor as well as other non-motor symptoms such as pain and anxiety (Cheon et al., 2009). In those who do have these type of fluctuations associated with medication efficacy; I hypothesized that off-periods will be associated with worsened cognitive performance at that timepoint. Finally, I hypothesized that more severe disease stage and higher symptom burden would moderate the relations between cognitive performance on mobile assessments and social context and state-based variables (mood, motivation, alertness).

Specifically, I expected that those with greater symptom burden, Unified Parkinson's Disease Rating Scale (UPDRS) score, and disease severity, Hoehn & Yahr (H&Y) score, would experience higher levels of fluctuation in their cognitive performance in relation to contextual factors; this hypothesis would be testable only if the disease severity range was not limited to mild-moderate cases, as are often seen in lab-based studies.

# CHAPTER TWO: MODELING THE RELATION BETWEEN MOMENTARY CONTEXTUAL FACTORS AND SMARTPHONE COGNITIVE PERFORMANCE IN YOUNG ADULTS

## Introduction

Traditionally, researchers and clinicians alike have assessed cognition in tightly controlled settings to eliminate the variance, and literal noise, in each person's everyday life. While it is essential to measure cognitive performance in a manner that produces clear and comparable results, we must also work to understand how humans' ability to process, inhibit, retain, and recall information is affected by a number of internal and environmental contextual factors. Remembering necessary grocery items in a crowded supermarket after a night of poor sleep, with a screaming child in tow, may be remarkably more difficult than performing a memory task with superficially similar demands but in a more optimized context, such as a quiet testing room. Further, there may be individual differences (e.g., baseline cognitive ability, age, trait anxiety) that differentiate those whose function is impaired in a chaotic supermarket from others whose performance is unaffected by that particular context.

**Cognitive Variability and Context:** For the last several decades, psychologists have carefully measured how cognition relates to internally and environmentally driven contextual factors. Across healthy and clinical populations, the cognitive domains of attention, inhibition, processing speed, and working memory are frequently studied in relation to a person's context (Weizenbaum et al., in press). These particular domains of cognition contribute to one's ability to quickly and efficiently solve a problem at hand (Duncan et al., 2017) and are predictive of both level of independent daily function in older adults (Kievit et al., 2016), risk of Alzheimer's disease at the preclinical stage (Huntley & Howard, 2010), and executive dysfunction could also indicate risk for developing a range of different psychopathologies in younger adults (Bowie et al., 2006; Lieshout et al., 2013; Luck & Gold, 2008). Attention, working memory, and processing speed are thought to be particularly susceptible to moment-to-moment changes in context because their functions rely on top-down control mechanisms of the prefrontal cortex, which helps to regulate behavior in response to environmental demands (Braver, 2012).

Lab-based experimental studies support the hypothesis that external environmental factors (e.g., environmental sounds, time of day) can affect cognitive performance. For example, intermittent sound was more disruptive to performance on a working memory task than high volume or no sound (Banbury et al., 2001). Other studies have measured cognitive performance within subjects at multiple times per day in lab settings and found that time of day has a significant effect on attention (Manly et al., 2002; van der Heijden et al., 2010) and on working memory (West et al., 2002). Notably, these findings are marked by individual differences (i.e., age and subjective alertness) in the time of day that leads to optimal performance on tasks. The literature has also established that internal factors that change from moment to moment, such as mood and motivation, can also affect performance on cognitive tests. For example, one study found that working memory performance was poorer on days when participants reported high levels of negative affect and low task motivation (Brose et al., 2012). For a more thorough review of the influence of environmentally- and internally-driven contextual variables on cognitive performance in a number of domains, see Weizenbaum et al., in press.

**Mobile Assessment of Cognition:** Ecologically valid measurement of a person's cognitive function in context requires assessment in the real-life environment (Ladouce et al., 2017). To observe change in function across time and place, the experience sampling method (ESM) allows for repeated assessment in everyday settings (Trull & Ebner-Priemer, 2013). ESM began in the form of daily diary studies to gather information about a person's state (e.g., mood, social interaction) in a given moment without the bias of retrospective reporting (Csikszentmihalyi & Larson, 1987). ESM has since evolved from handwritten diaries to Personal Digital Assistants (PDAs) to now-ubiquitous smartphones (Poushter, 2016), which, in young adults, are checked on average over 80 times per day (Andrews et al., 2015). The field of neuropsychological assessment has begun to take advantage of this new technology (Parsons et al., 2018) with the use of smartphones to measure cognition with strong metrics of reliability, validity, and feasibility (Moore et al., 2017). While there is an opportunity

While mobile assessment of cognition is being piloted in a number of clinical populations including those with HIV and substance use (Paolillo et al., 2018), Alzheimer's disease (Hassenstab et al., 2018), and Parkinson's disease (Bot et al., 2016), the method is also being used in healthy individuals to shed light on basic science questions of human cognitive processes measured in real-world settings. We know that working memory and attention fluctuate in response to state-based contextual conditions (largely induced in laboratory settings) but until now, due to methodological constraints, it has been difficult to understand how normal cognition operates in people's real-world environments. One study of using mobile assessment of cognition and *in situ* contextual variables have revealed that self-reported fatigue ratings in young adults are positively correlated with longer reaction times on an attention task (Price et al., 2017). In a smartphone study of over 700 participants of all ages (but primarily young adults), people performed better on processing speed and working memory games when alone than when with others, but there was no direct effect of mood on performance (von Stumm, 2018). Another recent study of healthy adults (age 19-73) also found that participants' smartphone visual working memory accuracy was higher when alone than when with others . In contrast to the study of von Stumm, Daniëls and colleagues found that working memory accuracy was lower when participants reported high positive affect. On a separate processing speed game, accuracy was lower when participants reported being tired, but higher with greater positive affect (Daniëls et al., 2020).

It comes as no surprise that variables such as fatigue, affect, and being alone vs. with others differentially affect our ability to hold information in mind and complete tasks quickly. Establishing which contexts most affect which domains of cognition is an important pursuit, not only as a means of improving our basic knowledge of human cognition, but also by providing a normative benchmark to assess data from clinical samples who may experience even greater performance variability in various contexts.

**Objective and Hypotheses:** In the present study, we examined the extent to which contextual variables shape thinking. We also examined individual differences in these associations. We addressed these questions using a smartphone app to test cognition (i.e., working memory and executive function) and context 5 times each day for 10 days in 177 young adults. The two smartphone cognitive tests were modeled after traditional neuropsychological measures of visual working memory and executive function, specifically complex attention and task-switching. Given the mixed prior findings on momentary affect and cognitive performance, we hypothesized that higher ratings of either positive or negative affect, and higher self-reported anxiety (nervousness), would negatively impact task accuracy. Second, we hypothesized that higher momentary sleepiness ratings would correlate negatively with cognitive performance. Third, we predicted that higher endorsement of “motivation to do well on the following games” would predict higher accuracy. We examined how individual differences such as baseline anxiety, trait-mood, daytime sleepiness, and subjective executive functioning, as well as demographics (age, education, sex), might moderate any significant relations between contextual variables and working memory and executive function assessed by smartphone.

## Methods

***Participants.***

204 young adults were recruited from the Boston University Department of Psychological and Brain Sciences' SONA system, a platform where undergraduates can sign up to participate in research in exchange for course credit for introductory psychology. Students were asked if they met inclusion criteria (English speaking adults, 18 years or older, own a smartphone with iOS or Android capability) on Qualtrics, an online questionnaire platform, before they were able to proceed with the initial questionnaires and smartphone study instructions. Of the 204 young adults who completed the initial set of online questionnaires, 177 proceeded to download and actively respond to smartphone notifications. The sample descriptive statistics in Table 3 reflect the data from these 177 participants.

**Table 3. Young Adults Participant Characteristics (N = 177)**

| **Years of Age** ***Mean (SD)*** | **Men, Women** | **Years of Education** ***Mean (SD)*** | **Race & Ethnicity** |
|---|---|---|---|
| 19.2 (1.3) | 77 Men<br>100 Women | 13.4 (1.7) | 50% White<br>• 13% Hispanic<br>• 8% Middle eastern<br>40% Asian<br>4% African American<br>6% Other |

**Measures.**

**Online Pre-Study Survey**

- *Mood and Anxiety Symptoms Questionnaire (MASQ-short)* - assess trait level of low mood and anxiety
- *Positive and Negative Affect Scale (PANAS-SF)* – assess tendency trait towards positive and negative affect
- *Social Provisions Scale* - assess level of social activity and social support in person’s life
- *Social Phobia Scale* – assess level of social anxiety
- *Motivation and Pleasure Scale (MAP-SR)* – assess degree of motivation and tendency towards pleasure-seeking behavior
- *Behavior Rating Inventory of Executive Functioning (BRIEF-A SR)* – assess self-perceived difficulties in a variety of domains of executive functioning
- *Epworth Sleepiness Scale (ESS)* – assess degree of typical daytime sleepiness

**mindLAMP Smartphone Assessments –** brief assessments of context, mood state, and cognition using the Mind Learn-Assess-Manage-Prevent app developed at Beth Israel Deaconness Medical Center Department of Digital Psychiatry (Torous et al., 2019). See screenshots of the app interface in Figures 2-5 below.

**Figure 2.**

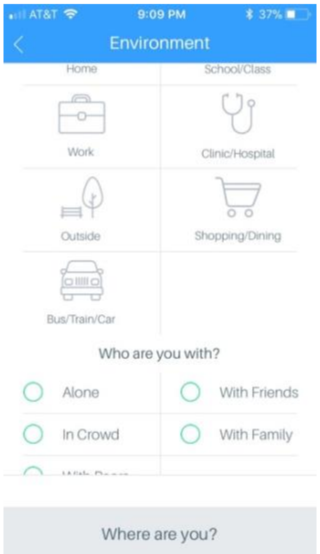



**Figure 3.**

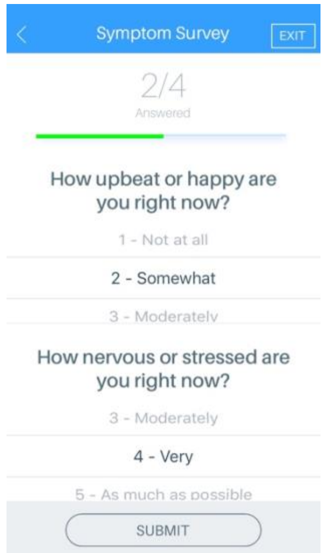



**Figure 4.**

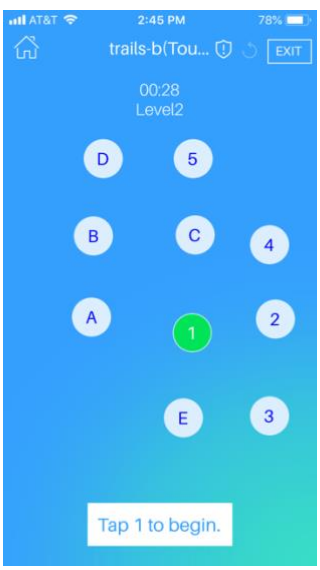
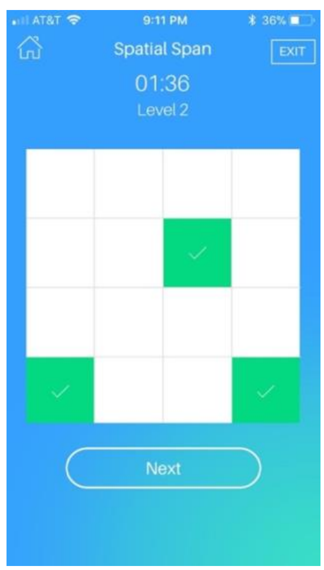



**Figure 5.**



- *Location & Social Context Survey* – indication of type of current location and social context (Fig. 2)
    - Location options: home, work/school, someone else's home, public place indoors, public place outdoors, transportation
    - Social context options: alone, strangers, classmates/co-workers, acquaintances, close friends, family members, romantic partner
- *Mood & Alertness Survey* – 5 questions relating to mood (happy/sad), anxiety, alertness, and level of motivation to try one's hardest on subsequent cognitive games (Fig. 3)
- *Trails-B* – four increasingly lengthy trials of a task requiring one to tap numbers and letters on the screen in a sequential and alternating manner (Fig. 4)
- *Backwards Spatial Span Test* – four increasingly challenging trials of a task requiring one to tap the sequences of boxes that appeared on the screen during that trail but in reverse order (Fig. 5)

**Study Procedures**

*Remote online questionnaires:* The study lasted 10 days and was conducted remotely with no in-lab component. The first part of the study involved completing questionnaires online via Qualtrics to assess trait affect, motivation, sleepiness, and executive dysfunction.

*Remote app study:* Following the initial online questionnaires, participants were emailed an instructions guide on how to install the study app on their smartphone and respond to prompts for surveys and cognitive games. Participants were shown in-app instructions on how to play the study's two cognitive games, and were asked to complete the surveys and games in their present environment within 2 hours of the pre-set reminder. Notification prompts were uniformly sent out at 9am, 12pm, 3pm, 6pm, and 9pm. The smartphone surveys included a multiple-choice question of where they were and whom they were with; four questions on mood state, alertness, and motivation; and the two tests of executive function (*Trails B)* and working memory performance (*Backwards Spatial Span)*. The whole survey with all questions and tests required less than 5 minutes to complete. Participants were sent emails on Day 2 and Day 5 of the study reminding them of how to complete surveys, and offering support in the case of any difficulties encountered using the app. Participants received 0.5 course credits for the online questionnaires, and 0.5 credits for every six smartphone assessments completed for a total of 3.0 possible credits, the total required for their course.

**Data analyses.**

Participant response rate was calculated as the number of completed smartphone assessments divided by the total possible number of assessments for each person (50), and then within-person means were averaged to create a sample response rate mean. To determine whether there was reliable and systematic between-person variability that was greater than within-person variability in participants' cognitive tests scores (i.e., Trails-B and Backwards Spatial Span), a multilevel modeling approach was taken using the software MPlus Version 8.1.7. Below is the formula for reliability where Var(BP) is the total variance in scores that is between persons, Var(WP) is the total variance in scores that is within persons, and *n* is the number of assessments (Sliwinski et al., 2018).

$$BP\ Reliability = \frac{Var(BP)}{\left(Var(BP) + \frac{Var(WP)}{n}\right)}$$

Scores were aggregated across time-points to find the sample's average within-person means and standard deviations across the full study period and by time-point and day. In addition to whole-sample analyses, within-person means were calculated along with within-person standard deviation to identify whether a person's level of variability in score was associated with the overall performance on the task. Pearson correlations were used to assess potential convergence between within-person smartphone cognitive test means and self-reported executive dysfunction (BRIEF-A). Because of the repeated nature of smartphone tests compared to the one-time BRIEF-A questionnaire, a multi-level model with maximum likelihood estimation was used with the BRIEF-A as a Level 2 fixed effect predictor of smartphone test performance.

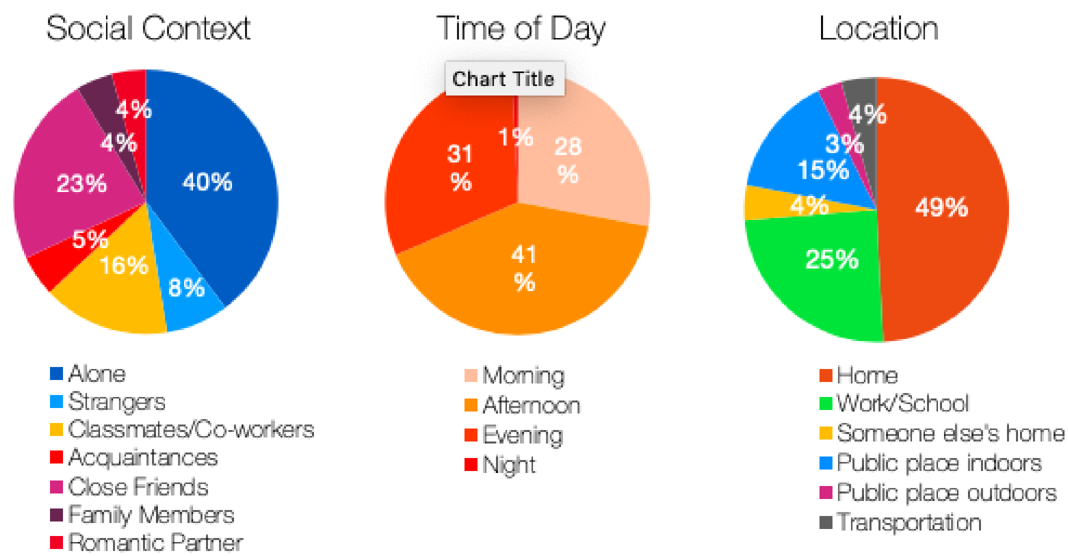
Multi-level model regressions were used to assess how smartphone test performance on either the Trails-B or Backwards Spatial Span task relate to context and state variables collected at the same timepoint. In this model, each contextual variable was entered as an independent Level-1 predictor of the outcome: Trails-B or Backwards Spatial Span. Survey timepoint (1-50) and day of study were independently analyzed as Level-1 predictors to determine whether smartphone cognitive test performance improved over the course of the study period as a result of practice and familiarity with the task. Finally, a two-level regression model was used to determine the extent to which Level-2 fixed effects (between-person measures) moderated the within-person relations between context and smartphone score. These fixed effects included demographics and self-reported trait affect, anxiety, daytime sleepiness, and motivation collected at baseline.

## ***Results***

*Feasibility.* Adherence rate in young adults was 59% (SD = 9.9), meaning that on average young adults responded to roughly 30 out of the 50 possible assessments over the course of the 10-day study period. Responses were given in a variety of social contexts, although the majority of the time participants reported being alone when completing the smartphone surveys and games (40% of the time), followed by close friends and classmates (Figure 6). Response was fairly evenly distributed across the course of the day, with the majority of responses occurring in the afternoon (41%) and evening (31%), which is to be expected as two prompts occurred in the afternoon and two in the evening. Responses were most likely to occur within the physical context of being home (49%)

followed by work and school (25%), which is sensible given that the sample was composed of undergraduates, specifically in their first or second years, who spend the majority of their time on campus.

**Figure 6. Summary of Smartphone Survey Responses in Young Adults**




*Smartphone cognitive test psychometrics*

*Trails-B:* On average, when scores were collapsed across time-points within individual participants, Smartphone Trails-B accuracy was 97.1% (SD = 1.5), indicating a ceiling effect with little variability around score from moment to moment within individuals (Table 4). Within-person means were significantly negatively correlated with standard deviation ($r = -0.79$, $p<.001$), meaning that high accuracy was associated with less intraindividual variability around the mean. Trails-B had a between-person reliability value of 0.66 which indicates a low-moderate amount of between-person variance in Trails-B score. Reliability was found by dividing between-person variance by total between- and within-person variance while accounting for the average number of measurement time-points (29.5 mean responses). Day of the study (1-10), and timepoint survey number (1-50) of the study were included as Level 1 predictors of Trails-B score. There was a direct effect of Day ($B = 0.16$, $SE = 0.02$, $p < .001$) and Timepoint ($0.02$, $SE = 0.003$, $p < .001$) on Trails-B score. This indicates the presence of a learning effect, such that performance improved with practice of the measure. When looking at aggregated averages across the sample by day, performance increased by roughly half a percentage point on days 1-3, and varied, on average, by only 0.1 to 0.2 on days 4-10. Convergent validity was assessed between the self-reported executive dysfunction questionnaire scores (BRIEF-A total and subscale scores) and smartphone Trails-B performance, a test of executive function. BRIEF-A total and subscale scores did not correlate with Trails-B within-person means or standard deviations. Further, the BRIEF-A total score was not a significant predictor of an individuals' Trails-B performance when entered as a Level 2 predictor in a multi-level model with the outcome as Level 1 Trails-B scores ($B = 0.07$, $SE = 0.34$, $p = 0.84$).

*Backwards Spatial Span*: Smartphone Backwards Spatial Span within-person mean accuracy was 87.8% ($SD = 10.4$) (See Table 4). Within-person means on Backwards Spatial Span were significantly negatively correlated with standard deviation ($r = -0.60$, $p<.001$), meaning that high accuracy was associated with less intraindividual variability. Backwards Spatial Span had a between-person reliability value of 0.908, which indicates a high level of between-person variance in score. There was not a significant direct effect of Day ($B = 0.169$, $SE = 0.091$, $p = 0.09$) or Time-point ($B= 0.02$, SE = 0.01, p = 0.11) on Backwards Spatial Span score. When aggregate averages by day were compared, scores from day 1 to day 2 increased on average by 4 percentage points, and for the rest of the study days the mean score varied less than 1 percentage point. While for the majority of participants there was not a significant learning or practice effect, the relation between day or time-point and Backwards Spatial Span score varied significantly from person to person as indicated by the between-person residual variance; day of study (B = 0.45, SE = 0.13, p = 0.001), timepoint (B = 0.01, SE = 0.002, p = 0.001). In regard to convergent validity, an individuals' executive dysfunction questionnaire (BRIEF-A) total and subscale scores did not correlate with Backwards Spatial Span within-person means or standard deviations; nor was it a significant predictor of smartphone Backwards Spatial Span performance in a multi-level model when entered as a Level 2 predictor (B = -0.10, SE = 1.9, p = 0.96).

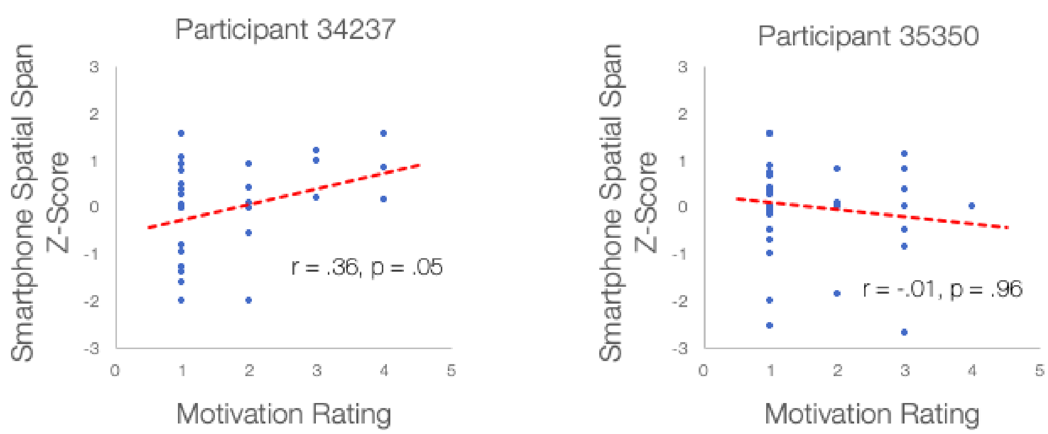
**Table 4. Smartphone Cognitive Test Psychometrics in Young Adults**

| | Within-Person Mean Accuracy | Correlation between Within-Person Mean and SD | Between-Person Reliability Value |
|---|---|---|---|
| **Trails-B** | 97.1% (SD=1.5) | r = -0.79 (p<.001) | 0.6629 |
| **Backwards Spatial Span** | 87.8 (SD=10.4) | r = -0.60 (p<.001) | 0.908 |

*ESM contextual variables and smartphone test performance:* We sought to understand how context, including state-based mood, anxiety, alertness, and motivation, as well as social and physical location, may relate to performance on smartphone tests completed at that same time-point. For Trails-B, none of the surveyed contextual or state-based variables predicted performance (Table 5 in Appendix A). For Backwards Spatial Span, there were no direct effects of surveyed contextual variables (Table 6 in Appendix A). By contrast, there was a significant level of between-person variability in the relation between momentary motivation (i.e., "How important is it to you do well on the following games?") and Backwards Spatial Span performance, residual variance (B = 2.31, SE = 1.15, p = 0.04) (Figure 7). We next conducted an exploratory analysis to identify any between-person variables (e.g., trait-based questionnaire score) that may have helped to explain for whom the association between momentary motivation and smartphone score was most salient. Exploratory Pearson correlations revealed associations between within-person mean motivation and years of education (r = -0.18, p

= 0.02) and PANAS negative affect subscale (r = -.173, p = 0.03). These variables were individually entered as moderating Level 2 predictors into the multi-level model with the Level 1 predictor, momentary motivation, and the Level 1 outcome, Backwards Spatial Span score; neither Level 2 predictors (education or PANAS score) were significant predictors of the between-person variance in motivation and Backwards Spatial Span.

**Figure 7. Examples of between-person variability in the association between momentary motivation and smartphone working memory performance**




## Discussion

### *Feasibility*

This study establishes the feasibility of using the mindLAMP app in a non-clinical sample of young adults. Unlike other studies of repeated smartphone assessment of cognition, this study was completed entirely remotely with no “in-lab” component. Of the 204 participants who initially completed online questionnaire, 177 successfully downloaded the study app and responded, on average, to 59% of the ESM prompts. This response rate was higher than another remotely conducted smartphone assessment survey where participants on average responded to 23 of the 3x/day surveys over the course of several months (von Stumm, 2018), but was lower than other studies of smartphone cognitive assessment for which response rates were typically in the 70-80% range (Daniëls et al., 2020; Moore et al., 2017); however, the latter all required an in-lab training session. Regardless, response from participants produced over 5,000 data points linking an individual's context and cognitive performance at that time-point.

One important factor relevant to feasibility is the incentive structure for survey response. In our study, participants received course credit for study participation, and fulfilled the baseline requirement if at least 30 of the 50 surveys were completed. Unsurprisingly, a large number of participants stopped participating after completing 30, leading to a precise sample average of 29.5/50 ESM prompts. In future it would potentially be valuable to build gamification (i.e., point-scoring, performance visualization, in-app rewards) into the app platforms that would drive intrinsic reward for completing surveys and games. Future studies should also consider a reward structure that remunerates participants throughout the study period, which would incentivize completing larger volumes of surveys through bonuses. Another option would be to incorporate variable reinforcement that would motivate a participant to continue responding in search of a "jackpot" survey reward.

*Trails-B*

This study included two smartphone cognitive assessments tasks. The first task given at each survey point was Trails-B, inspired by the traditional neuropsychological measure Trail Making Test B, a widely-used measure of executive function requiring switching accuracy and processing speed. We found that average performance was very high on the Trails-B task with little between or within person variability. The lack of variability associated with this ceiling effect made it difficult to measure unique relations between score and other variables. If this measure is to be used in the future it would be necessary to increase the difficulty of the task overall, potentially by adjusting the scoring such that time, as opposed to simply accuracy, was incorporated into the trial score. Similarly, quantifying the latencies between finger taps on the screen could allow for increasing the sensitivity of the measure and increasing variability indices. It would be necessary to consider that with finer scale measures of performance comes the challenge of accounting for differences in measurement between types of phones and operating systems needed for standardization (Germine et al., 2019).

The BRIEF-A questionnaire of self-reported executive dysfunction was not significantly related to either of the smartphone tests. This finding brings up the question of whether the tests measured two different constructs, or if the two assessment modalities, one a subjective self-reported questionnaire and the others objective smartphone performance-based tasks, are too disparate of methods to show convergence.

Despite the ceiling effect, there was significant between-person variability in scores, such that multi-level modeling was useful to account for the non-independence of the repeated measures. To answer our final, and central question, we sought to determine whether any of the contextual predictors of location, affect, alertness, or motivation may predict smartphone cognitive performance at that time. For Trails-B, there were no significant direct effects or significant between-person residuals indicating significant individual differences in the effect. This null finding was expected given the lack of variability in score seen with the ceiling effect for this measure. Instead of making the assumption that context has no effect on executive functioning, or specifically switching accuracy, it is more accurate to assume that contextual changes could not be observed using this particular measure of executive function.

*Backwards Spatial Span*

The second smartphone test was the Backwards Spatial Span task, which was inspired by the traditional WMS-III Backwards Spatial Span task of visual working memory. Compared to the Trails-B test, there was significantly more variability both within- and between-persons on the Backwards Spatial Span task. This indicates that it was the more sensitive of the two measures, and was a challenging enough visual working memory task that it allowed for differentiation to be seen in accuracy from moment to moment. The sample average accuracy was still quite high (above 85%) throughout the study period, however.

One important relation that was observed on both measures was that higher within-person mean accuracy was associated with lower within-person variability. This is an important finding that showed that those who overall were performing more poorly were also less consistent in performance. This aligns with recent work in a large healthy sample comprised of individuals across the lifespan showing that intraindividual variability was negatively related to baseline cognitive performance, most profoundly in young adults (A. Bielak & Anstey, 2019). Increased intraindividual variability has also been shown to distinguish clinical samples from healthy control samples in a number of clinical groups ranging from ADHD (Adams et al., 2011; Kofler et al., 2014), mood disorders (Gallagher et al., 2015), and mild cognitive impairment and Alzheimer's disease (Christ et al., 2018; Roalf et al., 2016) These findings have implications for traditional one-time testing, which may not always be sufficient to capture the variability that those with greater cognitive challenges may experience via fluctuations in day-to-day life.

While there was no direct effect of time-point or day-of-study found for the Backwards Spatial Span task, the significant residual variance showed that the learning effect was highly variable, such that some did experience this effect, while others did not. Similar to Trails-B When examining the relation between context and visual working memory as measured by the Backwards Spatial Span task, there were no direct effects observed. Individual differences were observed via significant between-person residual variance in the relation between momentary motivation ratings and backwards spatial span score. None of the between-person variables (demographics or questionnaire scores) were significant predictors of the motivation-backwards spatial span association. This indicates that there are likely variables that were not measured or analyzed as a part of our study that could shed light on whose self-report of feeling motivated to do well actually relates to variability in their smartphone cognitive test score.

*Conclusion*

This study revealed that a remotely deployed ESM study of cognitive performance via participants' own smartphones is feasible. In order to successfully measure variability from moment to moment, tasks need to be sufficiently challenging so that analysis is not limited by ceiling effects. Our smartphone measure of visual working memory, Backwards Spatial Span, showed that individual differences exist in learning rate or degree of practice effect. Individual differences were also evident in the relation between self-reported motivation and working memory performance at that given moment. Future directions include investigating variables that may moderate these effects and help to explain the variance from person to person. We recommend that in a non-clinical population, in order for cognitive tasks to be highly sensitive to variability, measures may need to include metrics such as finger-tap latencies to detect differences in score as they relate to one's state or context.

In these early stages of developing mobile assessment of cognition, these findings help guide the field of neuropsychology towards improved diagnostic sensitivity. Much of the research on smartphone cognitive assessment to date has been proof-of-concept studies focusing on feasibility and validity of primary outcome measures in a specific clinical sample or in healthy older adults (Moore et al., 2017). Given the emergence of new measures on a plethora of mobile platforms, continued work is needed in order to ensure strong foundational psychometrics of this new form of assessment (Bauer et al., 2012). One branch of this body of foundational research should be to improve the psychometrics and sensitivity of cognitive tests using smartphones in real-world settings. By establishing a body of data in non-clinical samples, we will be able to produce normative data for clinical use and a broader understanding of the dynamics and determinants of normative human cognitive performance across time and context.

## Appendix A.

**Table 5. Multi-level Model Results of Level 1 Predictors of Trails-B**

| Trails-B Level 1 | Direct Effect | | | BP residual | |
|---|---|---|---|---|---|
| | Unstd. Beta | SE | P-value | Unstd. Beta (SE) | P-value |
| **Day** | **0.159** | **0.024** | **<.001** | 0.016 (0.10) | 0.102 |
| **Timepoint** | **0.020** | **0.003** | **<.001** | <.001 (<.001) | 0.111 |
| **Time of Day** | -0.070 | 0.079 | 0.38 | 0.045 (0.154) | 0.771 |
| **Home** | -0.137 | 0.403 | 0.735 | 0.039 (0.290) | 0.892 |
| **Alone** | -0.209 | 0.281 | 0.456 | 0.016 (0.046) | 0.720 |
| **Upbeat** | -0.532 | 0.291 | 0.068 | 0.031 (0.086) | 0.717 |
| **Nervous** | -0.204 | -0.84 | 0.400 | 0.028 (0.095) | 0.769 |
| **Sad** | -0.367 | 0.347 | 0.291 | 0.270 (0.157) | 0.086 |
| **Alert** | -0.070 | 0.196 | 0.722 | 0.023 (0.034) | 0.500 |
| **Motivation** | -0.428 | 0.347 | 0.218 | 0.258 (0.149) | 0.083 |

**Table 6. Multi-level Model Results of Level 1 Predictors of Backwards Spatial Span**

| Backwards Spatial Span Level 1 | Direct Effect | | | BP residual | |
|---|---|---|---|---|---|
| | **Unstd. Beta** | **SE** | **P-value** | **Unstd. Beta (SE)** | **P-value** |
| **Day** | 0.156 | 0.091 | 0.086 | **0.449 (0.131)** | **0.001** |
| **Timepoint** | 0.018 | 0.011 | 0.114 | **0.007 (0.002)** | **0.001** |
| **Time of Day** | -0.166 | 0.254 | 0.515 | 1.366 (1.096) | 0.213 |
| **Home** | 0.834 | 1.251 | 0.507 | 1.345 (2.436) | 0.581 |
| **Alone** | 0.922 | 0.897 | 0.304 | 2.502 (2.795) | 0.371 |
| **Upbeat** | 0.032 | 1.033 | 0.975 | 1.377 (.777) | 0.076 |
| **Nervous** | -1.253 | 0.863 | 0.147 | 1.185 (0.736) | 0.109 |
| **Sad** | -0.431 | 1.277 | 0.735 | 0.39 (0.947) | 0.68 |
| **Alert** | -0.575 | 0.675 | 0.398 | 0.512 (0.338) | 0.13 |
| **Motivation** | 0.885 | 1.102 | 0.422 | **2.306 (1.147)** | **0.044** |

# CHAPTER THREE: FEASIBILITY, VALIDITY, AND CONTEXTUALLY-DRIVEN VARIABLITY IN COGNTION: A STUDY OF SMARTHPONE-BASED NEUROPSYCHOLOGICAL ASSESSMENT IN PARKINSON'S DISEASE

## Introduction

The rates of neurodegenerative disorders are rising as the population ages, and costs of assessment continue to stand as a barrier to timely diagnosis and treatment (Hurd et al., 2013). Persons with Parkinson's disease (PD), the most prevalent of the neurodegenerative movement disorders, experience the usual barriers to treatment found with other neurodegenerative disorders but their motor impairment is an additional impediment to clinic and lab visits (e.g., slow gait and risk of falls). Although these settings may not be easily accessible to all (Bradford et al., 2009), they are required for current methods of assessment of PD symptoms, including non-motor symptoms such as cognitive impairment. These limitations present time and access barriers, resulting in delayed diagnosis for some individuals long after symptoms emerge (Breen et al., 2013) and suboptimal treatment (Hogan et al., 2008).

Neuropsychological testing to characterize cognitive deficits requires hours of time from persons with PD, who often experience daytime fatigue (Friedman et al., 2007), and current methods of testing capture cognitive performance only at a single time-point. This is particularly problematic for persons with PD, which is a disorder marked by fluctuations in symptoms over time, as well as changes in mood, motivation, and arousal level (Cronin-Golomb et al., 2019). Even within a testing session, our lab recently documented cognitive variability in PD, using a spatial judgment task (Salazar et al., 2019). There is additional evidence to suggest that increased intraindividual variability in processing speed, even more so than within-person means, differentiates a sample of people with PD from healthy older adults (Burton et al., 2006; de Frias et al., 2007).

Digital technology has the potential to addresses the problem of limited access and the one-time nature of neuropsychological assessment in clinical populations such as PD by decreasing burden associated with traditional testing (Parsey & Schmitter-Edgecombe, 2013). Newly emerging digital assessments use computerized versions of the lengthy analog tests that only provide a snapshot of cognition in a single session (Bauer et al., 2012; Rentz et al., 2016). Instead, digital forms of these tests can now be completed in participant' and patients' homes at whatever frequency makes most sense for tracking cognitive function.

Another benefit to remote assessment via digital technology is the ability to easily assess cognitive variability across timepoints. Evidence suggests that variability in cognitive performance using a repeated assessment, or "burst" design (assessing many times over 1-2 weeks), may be a robust and sensitive marker of cognitive decline (Vaughan et al., 2013). Repeated testing of cognition and state-based variables via a mobile assessment device such as a smartphone allows us to understand disease symptoms' dynamics and determinants, which can inform our ability to predict changes in PD disease state or progression. This novel method of assessment offers the potential of measuring fluctuations in real-world activities of daily living as they relate to the individual. The ubiquity of smartphones introduces the capability of putting evidence-based assessment into the hands of those who may most benefit from it.

There are few studies to date that have used repeated smartphone assessment in PD, all of them quite recent. Ecological sampling method (ESM) and ecological momentary assessment (EMA) are two terms that describe frequent assessment of people in their real world environment with the objective of better understanding how targeted constructs change from moment to moment in daily life (Csikszentmihalyi & Larson, 1987). Our lab demonstrated the feasibility of conducting a smartphone ecological momentary assessment (EMA) study specifically designed to measure sleep in PD (Wu & Cronin-Golomb, 2019). Another study established proof-of-concept using smartphone technology to track intraindividual fluctuations in PD motor symptoms as they related to mood through smartphone EMA, and then modeled these fluctuations using network analysis, albeit in a single participant; increased anxiety was prospectively associated with increased tremor and increased cheerfulness was associated with decreased tremor (van der Velden et al., 2018). An open-data study used a brief repeated measures smartphone design to assess PD motor symptoms including objective motor and reaction time testing over six months (Bot et al., 2016; Lipsmeier et al., 2018). All elements of the smartphone assessment were correlated with at least one aspect of Unified Parkinson's Disease Rating Scale (UPDRS), for example finger tapping (motor speed) was significantly associated with ability to independently dress oneself without difficulty (Lipsmeier et al., 2018). Another mobile assessment study in PD used responses to develop an independent and valid disease severity score using a supervised and weighted machine learning algorithm; the score was significantly associated with the UPDRS and other common tests in PD such as Timed Up and Go (Zhan et al., 2018). A limitation of this latter study was that all assessments occurred within the clinic setting and no measures of neuropsychological function beyond reaction time were included.

To date, no studies have examined patterns of objective neuropsychological testing using smartphone technology in a sample of individuals with PD. It is unknown whether remote, repeated smartphone assessment of cognition is feasible in individuals with PD. This is unfortunate given the detrimental effects of cognitive impairment on quality of life (Lawson et al., 2016), and the difficulties of individuals in accessing traditional methods of neuropsychological assessment (Breen et al., 2013).

**Objectives and Hypotheses:**

The purpose of this study was to use a smartphone ESM study design to measure executive function and working memory, which are among the earliest cognitive dysfunctions to arise in PD (Cronin-Golomb et al., 2019) and have shown to be related to a wide range of other symptoms including anxiety (Reynolds et al., 2017) and impairments in spatial judgement (Salazar et al., 2019). An additional objective was to examine how PD symptoms and individual differences may relate to design feasibility and to intraindividual variability in cognitive performance. We hypothesized that the majority of participants would exhibit a completion rate similar to that seen in other studies of mobile cognitive assessment, roughly 70-80% completion of surveys (Moore et al., 2017). We hypothesized that those with lower baseline cognitive function scores on a global cognitive screening measure, the Montreal Cognitive Assessment (MoCA), would have lower rates of response due to difficulty navigating the smartphone app interface without the assistance of a researcher or clinician. Those with greater symptom burden (higher UPDRS score) were expected to have lower rates of response due to disease impairments such as pain or motor difficulties.

An emerging literature has demonstrated that smartphone assessments of cognition in clinical samples show strong convergent validity with their analog counterparts that are used in traditional neuropsychological assessment (Dagum, 2018; Pal et al., 2016; Schuster et al., 2016; Sliwinski et al., 2018; Timmers et al., 2014). While the cognitive measures included used in the present study are similar to tests used in the above-cited studies, performance on the app-based tests had not been measured alongside performance on the traditional neuropsychological tests Spatial Span [working memory]) and (Trail Making Test [executive function] upon which they are based. It was hypothesized that there would be a significant association between scores on the traditional neuropsychological tests and participant overall mean score (across timepoints) on the respective smartphone cognitive assessments.

Intraindividual variability is a sensitive marker of function and risk for decline in neurodegenerative disorders (Christ et al., 2018). Specifically in PD, we know that greater intraindividual variability in processing speed is associated with later stages of the disease (de Frias et al., 2007). To date, no studies have characterized within-person fluctuations in cognition across time-points and real-world contexts in a sample of participants with PD. It is hypothesized that cognitive score will be related to disease features such as level of motor impairment and subjective impairment in executive function.

In regard to state and context, we expected that factors such as recent exercise may be associated with improved cognitive performance, as a review of previous work has shown the widespread benefit of exercise to cognition in people with PD (Murray et al., 2014). For example, a dance program, motor training, and treadmill training were all associated with preservation or improvement of cognitive function in PD patients (da Silva et al., 2018). We also anticipated that smartphone cognitive performance would be related to self-reported on-off periods of medication, given that off-periods are associated with an increase in motor symptoms including tremor as well as other non-motor symptoms such as pain and anxiety (Cheon et al., 2009). In those without medication-related fluctuations, it was predicted that off-periods would be associated with poorer cognitive performance at that timepoint. It was also hypothesized that disease severity (motor symptoms and impairments in gait and cognition) would moderate the relations on smartphone assessments between cognitive performance and both social context and state-based variables (mood, motivation, alertness). Specifically, it was expected that those with higher disease stage and symptom burden would experience higher levels of fluctuation in their cognitive performance in relation to contextual factors.

## Methods

*Participants.*

Participants included thirty community-dwelling individuals who met the clinical criteria for mild to moderate idiopathic Parkinson's disease, following the United Kingdom Parkinson's Disease Society Brain Bank diagnostic criteria (Hughes et al., 1992). They were recruited from the Boston University Center for Neurorehabilitation and the Parkinson's Disease and Movement Disorders Center at Boston Medical Center. The study protocols were approved by the Boston University Institutional Review Board, with consent obtained according to the Declaration of Helsinki. All participants were native-English speaking adults who owned a smartphone with iOS or Android capabilities. Each needed to obtain a score of 20 or above on a cognitive impairment screening measure (Montreal Cognitive Assessment), and could not have a motor impairment (e.g., significant tremor or dyskinesia) that prohibited them from regularly using a smartphone. Out of the 30 enrolled, 27 completed the remote smartphone assessments in addition to the in-lab initial assessment. One dropped out due to other time commitments, and the other two did not complete the smartphone portion of the study due to technical failures with their phone and the study app. Of the 27 participants, all but one participant were taking medication to treat PD; we calculated levodopa equivalent daily dosages (LEDD) based on convention (Tomlinson et al., 2010). LEDD was not significantly correlated with any cognitive measures (in-lab or smartphone) ($r$'s $< 0.20$, $p$'s $>0.33$). See Table 7 for descriptive statistics of the final sample.

**Table 7. Parkinson's Disease Participant Characteristics (N = 27)**

| Age Mean (SD) | Men:Women | Years of Education Mean (SD) | Race & Ethnicity | PD Stage Hoehn & Yahr Median (Range) | MoCA Score Mean (SD) | UPDRS Total Score (SD) |
|---|---|---|---|---|---|---|
| **63.2 (8.7)** | 14 Men<br>13 Women | 17.7 (3.2) | 26 White<br>1 Middle Eastern<br>1 Native American | 2.0<br>(1.0 – 3.0) | 27.7<br>(2.5) | 27.0<br>(11.3) |

## <u>Study Measures</u>

### Online Pre-Study Survey

- *Mood and Anxiety Symptoms Questionnaire (MASQ-short)* - assess trait level of low mood and anxiety
- *Positive and Negative Affect Scale (PANAS-SF)* – assess tendency trait towards positive and negative affect
- *Motivation and Pleasure Scale (MAP-SR)* – assess degree of motivation and tendency towards pleasure-seeking behavior
- *Behavior Rating Inventory of Executive Functioning (BRIEF-A SR)* – assess self-perceived difficulties in a variety of domains of executive functioning
- *Epworth Sleepiness Scale (ESS)* – assess degree of typical daytime sleepiness

### In-Lab Baseline Assessment

- *Trail Making Test A & B* – traditional neuropsychological measure of processing speed and complex attention; will analyze scores in relation to the smartphone adaptation of this test

- *Spatial Span Test (WMS-III)* – traditional neuropsychological measure of visual working memory; will analyze scores in relation to the smartphone adaptation of this test
- *Unified Parkinson's Disease Rating Scale (UPDRS)* – standard assessment of PD symptoms, which includes in-person interview of symptoms as well as a motor examination; will analyze scores in relation to participants' ability to successfully use the smartphone app (as indicated by assessment compliance rate), PD-specific EMA question regarding ON-OFF periods, and smartphone cognitive assessment scores

**mindLAMP Smartphone Assessments** – brief assessments of context, mood state, and cognition using the Mind Learn-Assess-Manage-Prevent app developed at BIDMC Dept. of Digital Psychiatry. (Torous et al., 2019). See screenshots of the app interface in Figures 2-5 below.

*Figure 2.*

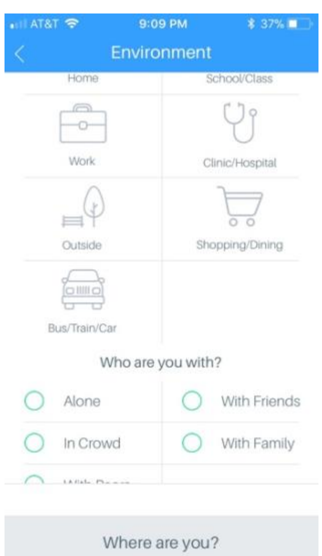
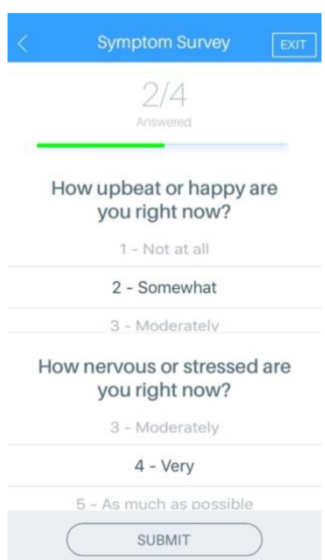



*Figure 3.*



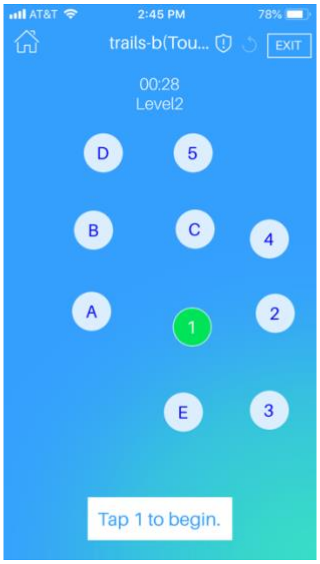



*Figure 4.*

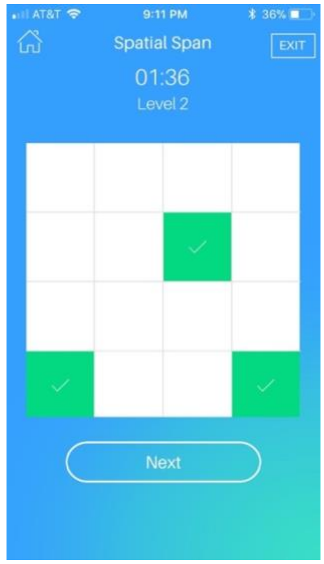



*Figure 5.*

- *Location & Social Context Survey* – indication of type of current location and social context (Fig. 2)
    - Location options: home, work/school, someone else's home, public place indoors, public place outdoors, transportation
    - Social context options: alone, strangers, classmates/co-workers, acquaintances, close friends, family members, romantic partner
- *Mood & Alertness Survey* – 5 questions relating to mood (happy/sad), anxiety, alertness, and level of motivation to try one's hardest on subsequent cognitive games (Fig. 3)
- *Trails-B* – four increasingly lengthy trials of a task requiring one to tap numbers and letters on the screen in a sequential and alternating manner (Fig. 4)

- *Backwards Spatial Span Test* – four increasingly challenging trials of a task requiring one to tap the sequences of boxes that appeared on the screen during that trail but in reverse order (Fig. 5)

**Study procedure**

*In-lab Baseline Assessment:*

After enrollment by phone, participants were emailed a link to complete a set of online questionnaires at home before coming into the lab for an in-person assessment. At the start of the in-lab session, a brief cognitive screen was administered (MoCA) to confirm eligibility for participation with a score above the cognitive impairment cut-off (score of 20 or below); all participants met this requirement. During the in-lab assessment, the Unified Parkinson’s Disease Rating Scale (UPDRS) was administered as well as two brief traditional neuropsychological measures: the Trailmaking Test A & B and the Wechsler Memory Scale-III Spatial Span Test. Participants were then assisted in downloading the study app on to their own Android or Apple smartphones. They were guided through a slideshow presentation on how to use the app and practiced completing surveys and games with the researcher present. Immediately following the study visit, participants were emailed the slideshow presentation of instructions, and were emailed on Day 2 and Day 6 of the study to check-in and troubleshoot any problems encountered using the study app. Participants received $15 per hour for the in-lab portion of the study, and $1 for every smartphone assessment they completed, with the opportunity to earn an additional $15 bonus if they completed at least 80% (40/50) of all assessments.

**Data Analysis**

ESM protocol adherence, meaning participant response rate, was calculated as the number of completed smartphone assessments divided by the total possible number of assessments for each person (50), and then within-person means were averaged to create a sample response rate mean. In order to be able to determine whether there was reliable and systematic between-person variability that was greater than within-person variability in participants' cognitive tests scores (i.e., Trails-B and Backwards Spatial Span), a multilevel modeling approach was taken using the software MPlus Version 8.1.7 that uses between- and within-person variance to find between-person reliability. Below is the formula for reliability where Var(BP) is the total variance in scores that is between persons, Var(WP) is the total variance in scores that is within persons, and *n* is the number of assessments (Sliwinski et al., 2018).

$$BP\ Reliability = \frac{Var(BP)}{\left(Var(BP) + \frac{Var(WP)}{n}\right)}$$

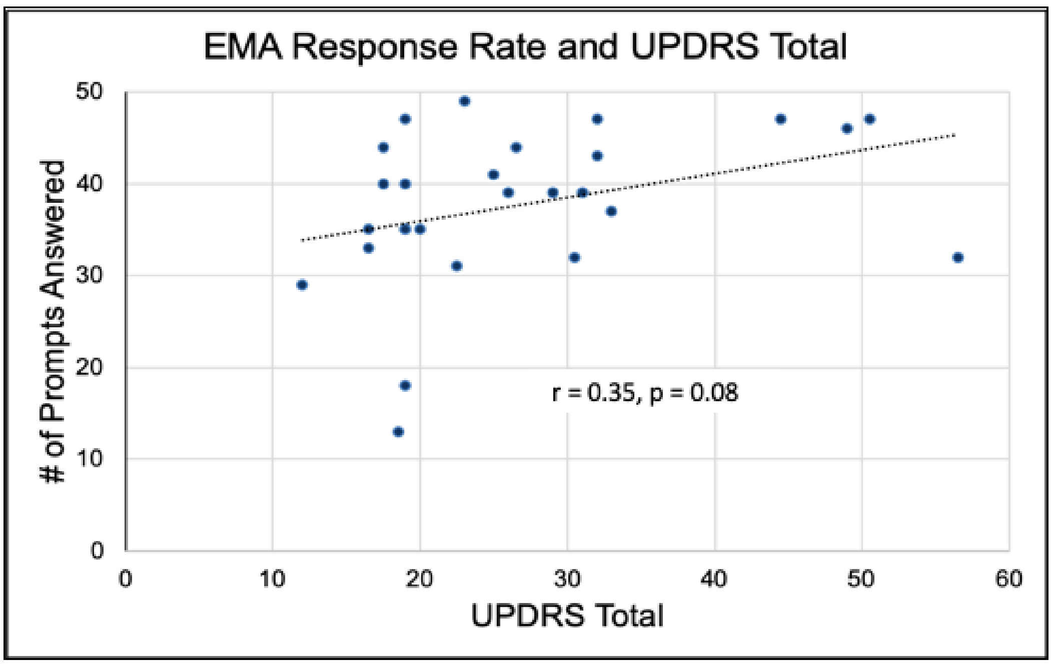
Scores were aggregated across time-points to find the sample's average within-person means and standard deviations across the full study period and by time-point and day. In addition to whole-sample analyses, within-person means were calculated along with within-person standard deviation to identify whether a person's level of variability in score was associated with the overall performance on the task. Pearson correlations were used to assess potential convergence between within-person smartphone cognitive test means and self-reported executive dysfunction (BRIEF-A). Because of the repeated nature of smartphone tests compared to the one-time BRIEF-A questionnaire, a multi-level model maximum likelihood regression was used with the BRIEF-A as a Level 2 fixed effect predictor of smartphone test performance as a Level 1 random effect. In order to assess convergent validity with traditional neuropsychological measures, in-lab tests scores were entered as Level 2 fixed effect predictors Trails-B and Backwards Spatial Span performance, Level 1 outcome variables, each in their own separate model.

Multi-level model regressions were used to assess how smartphone test performance on either the Trails-B or Backwards Spatial Span task relate to context and state variables collected at the same timepoint. In this model, each contextual variable was entered as an independent Level-1 predictor of the Level-1 outcome: Trails-B or Backwards Spatial Span. Survey timepoint (1-50) and day of study were independently analyzed as Level-1 predictors to determine whether smartphone cognitive test performance improved over the course of the study period as a result of practice and familiarity with the task. Finally, a two-level regression model was used to determine whether Level-2 fixed effects (between-person measures) may moderate the within-person relations between context and smartphone score. These fixed effects include demographics, PD-specific measures (UPDRS, H&Y score), baseline cognitive score (MoCA), and online questionnaires of trait affect, anxiety, sleepiness tendency, and motivation.

## Results

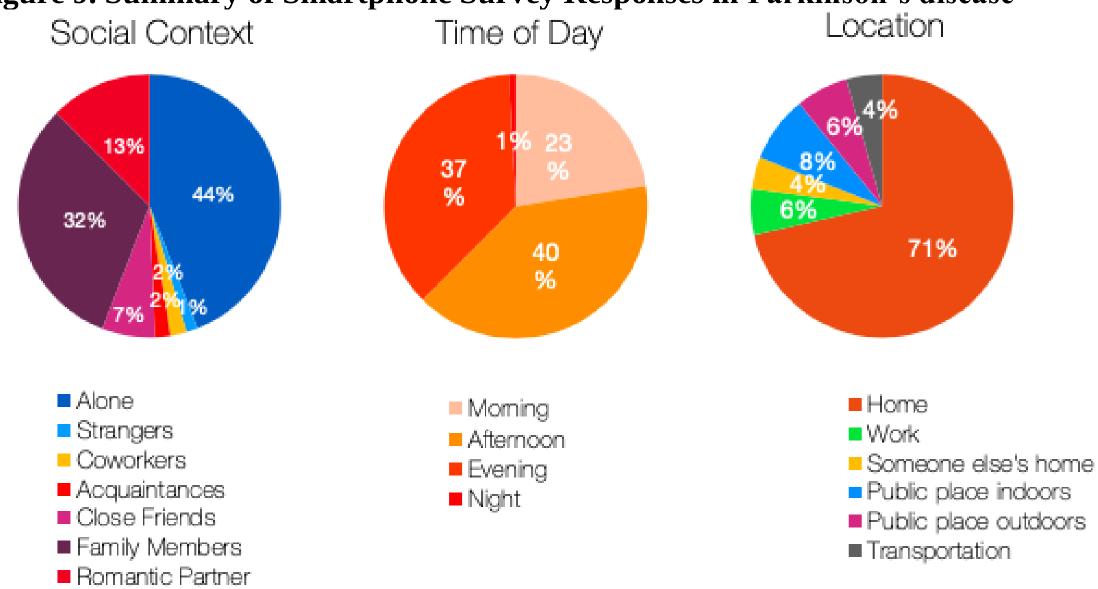
*Feasibility:* Adherence was 73% (SD = 10.0), meaning that on average participants responded to roughly 37 out of the 50 possible assessments over the course of the 10-day study period. We had hypothesized that age, PD symptoms (as measured by the UPDRS total), and baseline MoCA score would be negatively correlated with response rate to ESM prompts. We found that UPDRS total score was *positively* correlated, at the trend-level, with response rate ($r = 0.35$, $p = 0.08$), meaning that those with higher disease burden were more likely to respond to ESM prompts (Figure 8). However, age ($r = 0.21$, $p = 0.30$) and MoCA score ($r = -0.06$, $p = 0.77$) were not significantly correlated with adherence.

**Figure 8. Correlation Between Survey Response Rate and UPDRS Score**




Responses were given in a variety of social contexts, although the majority of the time participants reported being alone (44% of the time) when completing the surveys and games, followed by family members (32% of the time) (Figure 9). Response was fairly evenly distributed across the course of the day, with the majority of responses occurring in the afternoon (40%) and evening (37%), which is to be expected as two prompts occurred in the afternoon and two in the evening. Responses were most likely to occur within the physical context of being home (71%), with the remaining 29% relatively evenly distributed across other locations including public places indoors, outdoors, work, someone else's home, and transportation

**Figure 9. Summary of Smartphone Survey Responses in Parkinson's disease**




*Smartphone cognitive test psychometrics*

Trails-B. On average, when scores were collapsed across time-points, but within individual participants, Smartphone Trails-B accuracy was 96.2% (SD = 3.3), indicating a ceiling effect with little variability around score from moment to moment within individuals (Table 8). Within-person means were significantly negatively correlated with standard deviation ($r = -0.93$, $p<.001$), meaning that high accuracy was associated with less intraindividual variability around the mean. Trails-B had a between-person reliability value of 0.874. Day of the study (1-10) was analyzed as a predictor of Trails-B score. There were no direct effects of Day (B = 0.13, SE = 0.08, p 0.12) on Trails-B score, nor was there significant residual variance between persons. This indicates the absence of a learning effect, but likely is a reflection of the high and unvaried score accuracy across participants and across the study period.

Backwards Spatial Span. Smartphone Backwards Spatial Span within-person mean accuracy was 85.9% (SD = 5.2) (Table 8). Within-person means were significantly negatively correlated with standard deviation ($r = -0.70$, $p<.001$), meaning that high accuracy was associated with less intraindividual variability on Backwards Spatial Span score as well as Trails-B. Backwards Spatial Span had a between-person reliability value of 0.885. Day of the study (1-10) was analyzed as a predictors of Backwards Spatial Span score. Unlike Trails-B, there was a significant direct effect of Day (B = 0.33, SE = 0.13, p = 0.01) on Backwards Spatial Span score. This indicates the presence of learning effect, such that performance improved with practice of the measure; however, there was no significant residual variance indicating no significant individual differences in learning effect. Specifically, when scores were aggregated by day, the greatest amount of variability occurred on Day 1 mean score = 84.0 (SD = 12.1), after which SD rapidly narrowed to the study average of 85.9 (SD = 5.2).

**Table 8. Smartphone Cognitive Test Psychometrics in Parkinson's Disease**

| | Within-Person Mean Accuracy | Correlation between Within-Person Mean and SD | Between-Person Reliability Value |
|---|---|---|---|
| **Trails-B** | 96.2% (SD=3.3) | r = -0.93 (p<.001) | 0.874 |
| **Backwards Spatial Span** | 85.9 % (SD=5.2) | r = -0.70 (p<.001) | 0.885 |

*Smartphone cognitive test convergent validity*

The smartphone Trails-B task accuracy was not predicted by any of the traditional neuropsychological measures, including a baseline cognitive screening test (MoCA), the WMS-III Spatial Span Test, Trailmaking Test A & B Time, or the BRIEF-A Total Score (Table 9). By contrast, smartphone Backwards Spatial Span was significantly predicted by participants' performance on the MoCA, WMS-III Spatial Span Test (including total score backwards and forwards spatial spans and sub-test scores), the Trailmaking Test (time on A & B) (Table 9). Neither smartphone test scores were predicted by the BRIEF-A questionnaire score, nor was the questionnaire correlated with any of the traditional in-lab measures, all of which were objective task-based measures in contrast to a subjective self-report.

*Individual Differences as independent predictors of smartphone test scores*

Using multi-level regression models of between-person (Level 2) predictors of repeated (Level 1) measurements, we found that higher scores on the UPDRS tremor subscale (B = -0.77, SE = 0.27, p = 0.005), motor subscale (B = -0.25, SE = 0.07, p =

0.001), and total score (B = -0.13, SE = 0.05, p = 0.009) predicted lower overall smartphone Trails-B scores in participants, but had no effect on Backwards Spatial Span Scores. Older age predicted lower overall Backwards Spatial Span Scores, but not Trails-B (Table 14 in Appendix B).

**Table 9. Convergent Validity between Smartphone and Analog Cognitive Tests**

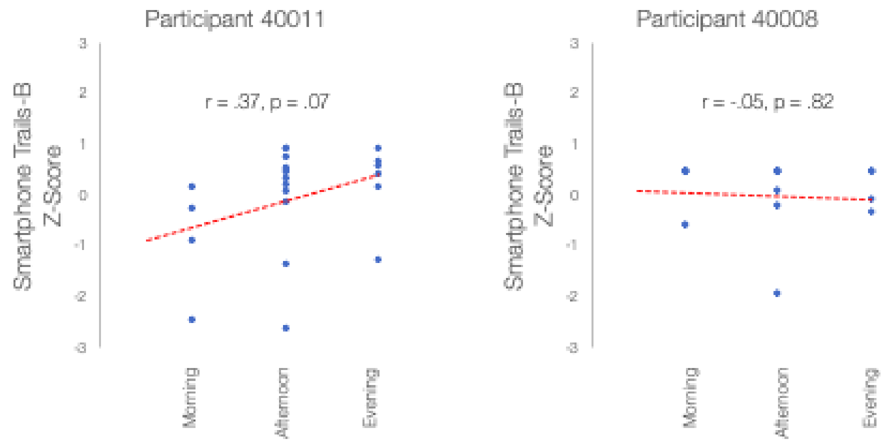
| | **Smartphone Trails-B** Beta Coefficient (SE) | **Smartphone Backwards Spatial Span** Beta Coefficient (SE) |
|---|---|---|
| **MoCA Total** | 0.128 (0.27) p = 0.637 | **0.91 (0.33) p = 0.012** |
| **Trailmaking Test B Time** | -0.035 (0.03) p = 0.186 | **-0.14 (0.02) p = <0.001** |
| **WMS-III Spatial Span Total** | 0.41 (0.22) p =0.066 | **1.08 (0.23) p = <0.001** |
| **BRIEF-A Total** | -1.281 (3.47) p = 0.712 | -3.41 (1.04) p = 0.148 |

*ESM contextual variables and smartphone test performance*

See Tables 10 and 11 in Appendix B. for a full listing of contextual variable results. For Trails-B, there were no significant direct effects of contextual ESM variables on performance. However, there was a significant level of between-person variability, given the residual variance values, in the relation between time of day and Trails-B score (B = 3.75, SE = 1.65, p = 0.02) (Figure 9) and being at home versus another location and Trails-

B score (B = 57.76, SE = 23.49, $p = 0.01$). Based on exploratory Pearson correlations between individual differences and contextual variables we entered significant correlates as Level 2 or between-person predictors in multi-level models (See Table 12 in Appendix B for full results). We found that participants with a UPDRS total score <23 were more likely to have a higher Trails-B performance at later time-points in the day. Also, those who reported lower baseline daytime sleepiness were more likely to have a higher Trails-B performance at later time-points in the day. No individual difference correlates of being at home predicted the strength of relation between being at home and Trails B performance.

**Figure 10. Examples of between-person variability in the association between momentary motivation and smartphone working memory performance**




For Backwards Spatial Span, being at home (vs. other location) predicted an increase in Backwards Spatial Span score at the same time-point (B = 7.19, SE = 3.01, p = 0.02); there was no significant between-person residual variance in this association. Having reported exercising in the last 3 hours also predicted an increase in Backwards Spatial Span score at the same time-point (B = 2.16, SE = 0.78, p = 0.01); there was significant between person residual variance associated with this direct effect (B = 1.26, SE = < 0.001, p = < 0.001), meaning that individual differences exist in the strength of the exercise and Backwards Spatial Span Score association. When entered as a Level 2 predictor we found that a person's mean exercise frequency score (the percent of times a person endorsed having exercised in the past 3 hours) predicted the strength of the relationship between exercise and Backwards Spatial Span Score. Specifically, when those who exercise less frequently *do* endorse having exercised in the last 3 hours, they are most likely to have the greatest increase in Backwards Spatial Span score ($B = -2.42$, $SE = 0.10$, $p < 0.001$). No other individual differences predicted the exercise and spatial span association (see Table 13. in Appendix B).

## ***Discussion***

### *Feasibility*

This study establishes the feasibility of using a repeat smartphone assessment to measure cognition and context in a sample of persons with idiopathic PD. There is a paucity of data on smartphone cognitive assessment in PD, and while a visual working memory task has been incorporated into other apps used to measure symptoms in PD (Bot et al., 2016), to our knowledge this is the first study of its kind to report any data on remote repeat measurement of working memory and executive functioning in PD. Of the 30 participants who initially completed an in-lab assessment, 27 successfully used the study app from home and responded, on average, to 73% of the ESM prompts. This response rate is similar to other studies of mobile cognitive assessment in a variety of clinical samples, which found average study response to be 79% (Moore et al., 2017), and is higher than the 61% response rate in a study of smartphone assessment of PD symptoms (Lipsmeier et al., 2018).

In our analysis of between-person variables that may contribute to feasibility and response rate, we found that a higher total UPDRS score was associated with a higher response rate to surveys. One possible interpretation of this finding is that those with greater disease burden, as indicated by a higher UPDRS score, were more often at home (vs. another location) during the day than those with lower UPDRS scores ($r = 0.44$, $p = 0.02$). Participants with more symptoms may also have fewer household obligations as a result of their disease symptoms, which may have allowed for more feasible response to smartphone notifications compared to others with less disease burden and greater engagement in other tasks.

*Psychometrics*

This study included two smartphone cognitive assessments. The first task given at each survey point was Trails-B, inspired by the traditional neuropsychological measure Trailmaking Test B, a measure of executive function. We found that average performance was very high on the smartphone Trails-B task indicating a ceiling effect with this measure. In the future, increasing the difficulty or scoring of this measure would improve psychometrics. Regardless, the measure was found to be reliable in its ability to differentiate between participants above and beyond within-person fluctuation when taking into account repeated measurement.

Our second task, Backwards Spatial Span, was a challenging visuospatial working memory task that had notable spread in accuracy scores both within and between people. On both smartphone cognitive measures, it was found that higher within-person mean accuracy was associated with lower within-person variability. This indicates that those who are performing more poorly are likely to be less consistent. An implication is that traditional one-time testing may not always be sufficient to capture the extent of variability (the good days and bad days) that people with cognitive impairment experience in day-to-day life. On Backwards Spatial Span, there was a positive effect of time such that later study timepoint predicted higher score, though this was largely restricted to the first day of study in which there was a large amount of variability, indicated by aggregated standard deviation, with variability decreasing substantially after the first study day.

*Convergent Validity*

Performance on Trails-B was not related to performance on any of the traditional neuropsychological tests that were administered in-lab. This does not come as a surprise in light of the ceiling effect elicited by the smartphone measure. Further, the smartphone version of Trails-B was a measure of switching accuracy versus speed, whereas completion time, or speed, is the primary metric of performance in the traditional Trailmaking Test B. Based on the constraints of the in-app task, a measure of trial speed was not available. In the future, scoring based on speed or screen tap latency could make this task a more sensitive measure of processing speed and executive function. For example, it may be that the microsecond latency or response time in between alternating numbers in letters in the task is what differentiates people more than the accuracy of switches, and this kind of sensitive timing ability is what differentiates digital assessment from current analog versions of this task.

In contrast, the smartphone Backwards Spatial Span task was related to a number of the traditional in-lab neuropsychological measures. Most notably, in-lab WMS-III Spatial Span score (including total, forward, and backwards subscores) strongly predicted an individual's performance on the smartphone task. This provides evidence that this smartphone task likely measures a similar construct as the traditional in-lab neuropsychological measure. Before this study, smartphone tasks had not been validated against traditional measures. The present finding bolsters the argument for use of smartphone cognitive assessment as a feasible and valid adjunct to traditional neuropsychological measures, especially because smartphone tasks can completed quickly, administered remotely, and repeated across time-points without a significant practice effect.

When considering the use of smartphone cognitive assessment in clinical populations with motor impairments, such as in PD, it is essential to understand any relations between disease symptoms and performance on the tasks themselves. For example, in our study the Trails-B task required quick and spatially precise taps on small on-screen circles. We found that higher UPDRS motor and tremor subscale scores were associated with lower scores on the task itself. If the objective of the assessment is to measure changes in processing speed and motor accuracy as a result of disease, this could be valuable to capture in a smartphone cognitive measure. When the goal is to understand complex attention, switching, and executive function, however, it is clear that a participant's motor impairments may introduce a confound that obscures the utility of the cognitive measure. By contrast, the Backwards Spatial Span task was not associated with PD symptoms as measured by the UPDRS, which signals its utility in measuring cognitive performance via smartphone in those with motor speed impairments.

*ESM contextual variables and smartphone test performance*

One of the core theoretical questions in our study of repeat smartphone assessment was whether performance varied across different contexts. While we did not find any direct effects between contextual predictors and Trails-B, we did find that the tendency to perform better later in the day was associated with less symptom burden (UPDRS total score) and lower daytime sleepiness (ESS score). While there is reason to be cautious about these findings given the ceiling effect and motor confounds seen with Trails-B, it is worth following up on the finding that those with greater disease burden may show *less* sensitivity to varying contexts, alternatively they may be more homebound and thus have less exposure to varying physical and social contexts. One potential future question is how variability in context and one's response to these different contexts relate to stage of a disorder; could a decrease in the variety of contexts or one's differential response to them as measured by repeated mobile testing help to predict advancing disease state? Alternatively, mobile testing may only be efficacious for those in early stages of a disease, but may help to identify contexts in which performance is lower and targeted interventions could be of benefit.

Looking to our other task, Backwards Spatial Span, it was found that being home rather than in another environment and endorsing recent exercise predicted higher scores on the smartphone measure. On a measure of working memory, it is unsurprising that being in a more familiar and perhaps more controlled or less distracting environment such as one's home would be conducive to stronger working memory performance. There was no indication of significant between-person differences in this effect.

In interpreting the direct effect of exercise on visual working memory performance, it is important to note that this is an observational finding only. There is a possibility that exercise and its physiological effects could be a mechanism towards improved cognitive performance in PD.  Indeed, a recent systematic review showed that a variety of exercise programs can lead to improvements in global cognitive function, processing speed, attention, and mental flexibility in people with PD (da Silva et al., 2018). A number of studies together point towards increased levels of blood-derived neurotrophic factor (BDNF) as the mechanism of action in the relation between exercise and higher cognitive performance in PD (Hirsch et al., 2018). It is also quite possible that on days when people are feeling better, broadly, they may be more likely to perform better on a working memory task and also may feel more motivated to exercise. We found significant residual variance in this association indicating individual differences in how much recent exercise influences Backwards Spatial Span score from one person to the next. Specifically, those who least frequently endorsed recent exercise were the ones most likely to have a higher score when they *did* report having exercised in the last 3 hours. This finding may further support the correlational interpretation that people with PD who exercise more frequently may be exercising unconditionally on "good" days and "bad" days or despite other factors of variability in their daily lives. By contrast, those who endorse exercising less frequently may have a more conditional relationship with exercise, such that choosing to exercise may be associated with other factors (e.g., less physical discomfort, a less busy day, etc.) that could also be conducive to better working memory.

*Conclusion*

This pilot study revealed that repeated smartphone assessment of cognitive performance in persons with PD is a feasible method of assessment. Our smartphone measure of executive functioning and visual working memory both displayed strong between-person reliability. There was convergent validity for the smartphone test of visual working memory, with its results predicted by traditional neuropsychological tests administered in-lab. The visual working memory task performance was also unrelated to PD motor impairment and tremor score, which demonstrates feasibility even in those with notable motor-based symptoms of PD. Lastly, we found that certain contexts and conditions (being at home and recent exercise) can differentially predict performance on a smartphone test of visual working memory at that same time-point. However, these results should be taken as initial signals of the context-cognition relation in PD rather than conclusive results given the study's small sample size composed of individuals with mild-moderate disease severity, thus generalization to a larger of people with more severe PD must be limited. Together these findings lend support for further investigation of smartphone cognitive assessment in PD as a means of understanding idiographic patterns of symptoms with the potential to better predict disease progression and provide targeted and individualized interventions.

# Appendix B.

**Table 10. Multi-level Model Results of Contextual Predictors of Smartphone Trails-B Score**

| Trails-B Level 1 | Direct Effect | | | BP residual | |
|---|---|---|---|---|---|
| | Unstd. Beta | SE | P-value | Unstd. Beta (SE) | P-value |
| **Day** | 0.125 | 0.08 | 0.117 | 0.005(.038) | 0.893 |
| **Timepoint** | 0.016 | 0.01 | 0.121 | <.0.000 (.001) | 0.607 |
| **Time of Day** | 0.279 | 0.474 | 0.556 | **3.753 (1.646)** | **0.023** |
| **Home** | 1.259 | 2.114 | 0.552 | **57.761 (23.494)** | **0.014** |
| **Alone** | 0.946 | 1.274 | 0.458 | 0.178(1.057) | 0.866 |
| **Upbeat** | -1.294 | 2.862 | 0.651 | 4.762(2.636) | 0.071 |
| **Nervous** | 0.121 | 2.303 | 0.958 | 3.863 (3.144) | 0.219 |
| **Sad** | 2.333 | 2.887 | 0.419 | 0.147(0.366) | 0.688 |
| **Alert** | 0.568 | 0.756 | 0.452 | 0.756(0.401) | 0.059 |
| **Motivation** | -2.274 | 9.747 | 0.816 | 1.630(2.300) | 0.554 |
| **ON-Med Period** | 1.21 | 3.789 | 0.749 | 31.02 (22.265) | 0.164 |
| **Caffeine** | 0.587 | 7.806 | 0.94 | 1.544(7.056) | 0.827 |
| **Exercise** | 0.195 | 0.966 | 0.84 | 1.048(2.668) | 0.694 |

**Table 11. Multi-level Model Results of Contextual Predictors of Smartphone Backwards Spatial Span Score**

| Backwards Spatial Span Level 1 | Direct Effect | | | BP residual | |
|---|---|---|---|---|---|
| | **Unstd. Beta** | **SE** | **P-value** | **Unstd. Beta (SE)** | **P-value** |
| **Day** | **0.329** | **0.129** | **0.011** | 0.080(0.151) | 0.597 |
| **Timepoint** | **0.042** | **0.016** | **0.008** | 0.001(.002) | 0.595 |
| **Time of Day** | 0.683 | 0.43 | 0.112 | 0.176(0.061) | 0.951 |
| **Home** | **7.188** | **3.012** | **0.017** | 3.421(5.951) | 0.565 |
| **Alone** | 0.778 | 2.069 | 0.715 | 0.368(8.413) | 0.965 |
| **Upbeat** | -0.367 | 2.804 | 0.896 | .391(1.391) | 0.779 |
| **Nervous** | -2.017 | 2.435 | 0.408 | 0.950(0.782) | 0.224 |
| **Sad** | 1.651 | 9.24 | 0.858 | 0.629(7.473) | 0.933 |
| **Alert** | 1.066 | 1.045 | 0.308 | 0.262(1.023) | 0.798 |
| **Motivation** | -8.075 | 13.71 | 0.556 | 1.522(11.482) | 0.895 |
| **ON-Med Period** | 6.538 | 6.046 | 0.286 | 8.836(18.118) | 0.626 |
| **Caffeine** | 0.857 | 7.806 | 0.940 | 1.544(7.056) | 0.827 |
| **Exercise** | **2.157** | **0.778** | **0.006** | **1.262(<.001)** | **<.001** |

**Table 12. Multi-level Model Results of Between-Person Predictors of Context-Trails-B Performance Association**

| | Smartphone Trails-B | | | |
|---|---|---|---|---|
| **Level 1 Direct Effect** | **Level 2 Predictor** | **Unstd. Beta** | **SE** | **P-value** |
| **Time of Day** | UPDRS Total | -0.022 | 0.016 | 0.163 |
| | UPDRS Total >23 | -0.635 | 0.327 | 0.052 |
| | UPDRS Motor | -0.3 | 0.023 | 0.2 |
| | UPDRS Motor >11 | -0.577 | 0.318 | 0.07 |
| | Time since dx | -0.2 | 0.037 | 0.58 |
| | H&Y score | 0.124 | 0.298 | 0.678 |
| | Age | -0.012 | -0.618 | 0.536 |
| | ESS | -0.666 | 0.249 | 0.008 |
| **Home** | UPDRS Total | 0.027 | 0.038 | 0.478 |
| | Time since dx | -0.104 | 0.093 | 0.26 |
| | Age | 0.027 | 0.055 | 0.62 |
| | MoCA | 0.05 | 0.149 | 0.739 |
| **ON-OFF Meds** | UPDRS Total | -0.096 | 0.075 | 0.199 |
| | UPDRS Total >23 | -2.243 | 1.694 | 0.185 |
| | Time since dx | -0.091 | 0.252 | 0.72 |

**Table 13. Multi-level Model Results of Between-Person Predictors of Context-Backwards Spatial Span Performance Association**

| | Smartphone Backwards Spatial Span | | | |
|---|---|---|---|---|
| **Level 1 Direct Effect** | **Level 2 Predictor** | **Unstd. Beta** | **SE** | **P-value** |
| | UPDRS Total | 0.062 | 0.052 | 0.467 |
| **Exercise** | UPDRS Total >23 | 0.102 | 0.081 | 0.389 |
| | Age | -0.037 | 0.14 | 0.792 |
| | MoCA | 0.135 | 0.386 | 0.726 |

**Table 14. Multi-level Model Results of PD-Specific Individual Differences and Smartphone Cognitive Test Scores**

| | Trails-B | | | Backwards Span Span | | |
|---|---|---|---|---|---|---|
| | **Unstd. Beta** | **SE** | **P-value** | **Unstd. Beta** | **SE** | **P-value** |
| **H&Y Score** | -0.333 | 1.298 | 0.797 | 2.41 | 1.81 | 0.183 |
| **UPDRS Total** | **-0.132** | **0.051** | **0.009** | -0.025 | 0.084 | 0.767 |
| **UPDRS Tremor** | **-0.768** | **0.271** | **0.005** | 0.122 | 0.456 | 0.789 |
| **UPDRS MMB** | 0.225 | 0.312 | 0.471 | 0.126 | 0.457 | 0.784 |
| **UPDRS ADL** | -0.28 | 0.149 | 0.06 | -0.301 | 0.228 | 0.186 |
| **UPDRS Motor** | **-0.246** | **0.074** | **0.001** | 0.026 | 0.129 | 0.841 |
| **Age** | -0.096 | 0.075 | 0.202 | **-0.349** | **0.089** | **<.001** |

# CHAPTER FOUR: GENERAL DISCUSSION

*Conclusion: Overview of Dissertation Findings*

Together, Study 1, in non-clinical young adults, and Study 2, in people with Parkinson's disease, demonstrated the wide-ranging feasibility of smartphone cognitive assessment. When further considering the central results of these studies, several key takeaways emerge:

1) An effective incentive or remuneration schedule is important in intensive longitudinal studies such as these where high levels of survey response are needed to fully evaluate theoretical questions of context and performance. Payment per survey in a highly motivated self-selected clinical sample (PD participants) led to higher survey and game response rates than students who required to complete at least 60% of surveys required course credit; had this required percentage been higher than 60, compliance likely would have risen as well.
2) Smartphone measures of cognition must be psychometrically evaluated and calibrated to avoid a restricted score range, such as seen with a ceiling effect. In both studies scores tended to stay above 80%, and little variability in range was seen particularly for the test of executive function, or switching accuracy, Trails-B.
3) In a sample of non-clinical young adults attending a competitive university, there was not significant variability in cognitive performance as a function of

differing real-time state and context variables. These largely null findings indicate that either the amount of fluctuation in executive function and working memory in a non-clinical sample of young adults is negligible in relation to the contextual variables assessed, or the smartphone measures of cognitive function used in these studies were not sensitive enough to capture meaningful fluctuations.

4) In persons with PD, there were several direct effects of context on working memory performance. Further, individual differences, such as disease severity, identified the groups of individuals for which certain contextual or state-based variables led to significant changes in their smartphone cognitive performance at that time. This finding indicates the potential of this method to precisely identify when and where certain groups of people experience the most cognitive difficulty, and from there be able to personalize real-time interventions to the moments where they most matter.

*Future Steps for the Field*

Although new innovative forms of smartphone assessment are becoming increasingly common, there are several likely reasons for a dearth of research to date using mobile technology to study cognition and context. First is the financial barrier to developing app-based assessment tools. In addition, there are significant time demands and logistical challenges of establishing feasibility, reliability, and validity of these new instruments. In traditional neuropsychological assessment, the objective has been to measure cognitive function in a quiet setting with few distractions to obtain the best possible performance. When measuring cognitive function with a mobile device, by contrast, there is opportunity for ample variability and distractions. This contextual "noise" can in fact be a strength, reflecting a more ecologically valid environment, but it also means that in order to form an accurate picture of a person's functioning across contexts, a large volume of data is needed. Analysis of acquired mobile assessment data will need to take into account the high frequency of measurements within individuals and between groups in order to produce clinically useful normative comparison data (Cohen, 2019). This brings forth the challenge of localizing the driving factors of within-person variance, while also quantifying practice effects.

Some have raised concerns about the use of computerized neuropsychological assessment due to the lack of standardization and normative data, which when applied to patient care could ultimately lead to poor clinical decisions and errors in diagnosis (Gates & Kochan, 2015). There have been several calls to the field to address this concern by focusing resources and investigation into the psychometrics of mobile cognitive assessment (Bakker et al., 2016; Cohen, 2019; Germine et al., 2019; Parsons et al., 2018) and also centralizing the development of a toolbox of standardized mobile neuropsychological assessments that can be validated efficiently through a network clinical researchers across the field (Gershon et al., 2020).

Inherent in pioneering mobile health technology comes the challenge of developing measurement tools that are useable and study designs that are informative. Unlike ever before, there is immense potential to study familiar neuropsychological constructs outside the laboratory setting and in the environments in which they naturally operate. In order to move beyond proof of concept validation of ambulatory assessments, the field is faced with the challenge of creating new paradigms that are not mere replications of existing neuropsychological measures but are designed to be independently administered across time, contexts, and devices. For example, moving beyond the gold-standard neuropsychological measures of cognition, some work has explored how smartphone activity, such as swipes and taps, could serve as a proxy for working memory, episodic memory, executive function, language, and intelligence in a pilot study of healthy adults (Dagum, 2018). The accessibility and availability of smartphones sets the stage for the emergence of new scalable research efforts that could easily recruit participants from both healthy and clinical samples and record vast quantities of new data. Highlighted in Figure 2 are the ways in which some of the key ecologically-situated variables can be measured using smartphones and connected to digital neuropsychological tasks performed on digital devices. The momentary and state-related variables can be measured passively (e.g., GPS, microphone) or actively (e.g., surveys). Together, data streams can be combined and analyzed using computational and machine learning methods to identify relationships between cognitive performance and contextual momentary factors.

Ultimately, the goal of mobile assessment methods is to refine understanding of clinical impairments in cognition and give rise to preventative interventions for mitigating cognitive dysfunction and decline. New work is being done to develop a smartphone app designed to improve clinical care based on the needs of the patient and clinician, in addition to the researcher. The objective of this app development is to create an open platform that can be used across a variety of research and clinical settings (Torous et al., 2019). Unlike other existing mobile health apps developed for a specific population or function, the mindLAMP (Learn-Assess-Manage-Prevent) has been created to allow for customizable surveys, sensors, cognitive tests, and schedules of notification-prompted assessment. The app seeks to integrate active assessments such as surveys and cognitive games with passive sensing data (GPS, pedometer, microphone) and even phone metadata, such as the number of times other social media or communication apps are used on the phone. With the rich dimensionality of smartphone data from such apps, multilevel modeling and machine learning analyses would be logical directions to take analysis in order to elucidate the casual links and predictive relations between cognitive performance and state-based and environmental factors.

Like any new method, repeat assessment of cognition and context via mobile technology must be carefully employed. Given the complexity and sheer data volume of temporarily dense, longitudinal, and dynamic smartphone and sensor data, it will be important for theory-driven hypotheses to guide this work instead of only data-driven models that will likely produce spurious as well as meaningful correlations. Moreover, given the nature of this new data, care must be taken to ensure its privacy protections and ethical uses (Kane & Parsons, 2017). Looking back at the history of genetics and neuroimaging, it is clear that the greatest progress with this new tool will emerge from interdisciplinary collaborations. As health care moves towards personalized medicine, the field also has an opportunity to move towards personalized assessment of cognition. While population level screening tests and individualized in-office neuropsychological evaluations will remain important, they cannot fully capture the social, physical, and environmental variance that each individual experiences. Fortunately, mobile technologies, like smartphones, can collect such data and so present the opportunity for a paradigm shift with personal devices capturing personal measurements to generate personal cognitive profiles and treatment directions.

**REFERENCES**


Adams, Z. W., Roberts, W. M., Milich, R., & Fillmore, M. T. (2011). Does response variability predict distractibility among adults with attention-deficit/hyperactivity disorder? *Psychological Assessment*, *23*(2), 427–436. https://doi.org/10.1037/a0022112

Allard, M., Husky, M., Catheline, G., Pelletier, A., Dilharreguy, B., Amieva, H., Pérès, K., Foubert-Samier, A., Dartigues, J.-F., & Swendsen, J. (2014). Mobile technologies in the early detection of cognitive decline. *PloS One*, *9*(12), e112197. https://doi.org/10.1371/journal.pone.0112197

Andrews, S., Ellis, D. A., Shaw, H., & Piwek, L. (2015). Beyond Self-Report: Tools to Compare Estimated and Real-World Smartphone Use. *PLoS ONE*, *10*(10). https://doi.org/10.1371/journal.pone.0139004

Asparouhov, T. (2018). *Dynamic Structural Equation Modeling of Intensive Longitudinal Data Using Mplus Version 8*. http://www .statmodel.com/download/Part%205%20Asparouhov.pdf

Bakker, D., Kazantzis, N., Rickwood, D., & Rickard, N. (2016). Mental Health Smartphone Apps: Review and Evidence-Based Recommendations for Future Developments. *JMIR Mental Health*, *3*(1). https://doi.org/10.2196/mental.4984

Banbury, S. P., Macken, W. J., Tremblay, S., & Jones, D. M. (2001). Auditory Distraction and Short-Term Memory: Phenomena and Practical Implications. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *43*(1), 12–29. https://doi.org/10.1518/001872001775992462

Bauer, R. M., Iverson, G. L., Cernich, A. N., Binder, L. M., Ruff, R. M., & Naugle, R. I. (2012). Computerized Neuropsychological Assessment Devices: Joint Position Paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. *The Clinical Neuropsychologist*, *26*(2), 177–196. https://doi.org/10.1080/13854046.2012.663001

Bell, R., & Buchner, A. (2007). Equivalent irrelevant-sound effects for old and young adults. *Memory & Cognition*, *35*(2), 352–364. https://doi.org/10.3758/BF03193456

Bielak, A. A. M., Mogle, J., & Sliwinski, M. J. (2017). What Did You Do Today? Variability in Daily Activities is Related to Variability in Daily Cognitive Performance. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*. https://doi.org/10.1093/geronb/gbx145

Bielak, A., & Anstey, K. (2019). Covariation of intraindividual variability in cognitive speed and cognitive performance across young, middle, and older adulthood. *Developmental Psychology*, *55*(5), 994–1004. http://dx.doi.org/10.1037/dev0000688

Bigler, E. D. (2012). Symptom Validity Testing, Effort, and Neuropsychological Assessment. *Journal of the International Neuropsychological Society*, *18*(4), 632–640. https://doi.org/10.1017/S1355617712000252

Bot, B. M., Suver, C., Neto, E. C., Kellen, M., Klein, A., Bare, C., Doerr, M., Pratap, A., Wilbanks, J., Dorsey, E. R., Friend, S. H., & Trister, A. D. (2016). The mPower study, Parkinson disease mobile data collected using ResearchKit. *Scientific Data*, *3*(1), 160011. https://doi.org/10.1038/sdata.2016.11

Bowie, C. R., Reichenberg, A., Patterson, T. L., Heaton, R. K., & Harvey, P. D. (2006). Determinants of Real-World Functional Performance in Schizophrenia Subjects: Correlations With Cognition, Functional Capacity, and Symptoms. *American Journal of Psychiatry*, *163*(3), 418–425. https://doi.org/10.1176/appi.ajp.163.3.418

Bradford, A., Kunik, M. E., Schulz, P., Williams, S., & Singh, H. (2009). Missed and Delayed Diagnosis of Dementia in Primary Care Prevalence and Contributing Factors. *Alzheimer Disease & Associated Disorders*, *23*(4), 306–314.

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113. https://doi.org/10.1016/j.tics.2011.12.010

Breen, D., Evans, J., Farrell, K., Brayne, C., & Barker, R. (2013). Determinants of delayed diagnosis in Parkinson's disease. *Journal of Neurology*, *260*(8), 1978–1981. https://doi.org/10.1007/s00415-013-6905-3

Brose, A., Lövdén, M., & Schmiedek, F. (2014). Daily Fluctuations in Positive Affect Positively Co-Vary With Working Memory Performance. *Emotion*, *14*(1), 1–6. https://doi.org/10.1037/a0035210

Brose, A., Schmiedek, F., Lövdén, M., & Lindenberger, U. (2012). Daily Variability in Working Memory is Coupled With Negative Affect: The Role of Attention and Motivation. *Emotion*, *12*(3), 605–617. https://doi.org/10.1037/a0024436

Burton, C. L., Strauss, E., Hultsch, D. F., Moll, A., & Hunter, M. A. (2006). Intraindividual variability as a marker of neurological dysfunction: A comparison of Alzheimer's disease and Parkinson's disease. *Journal of Clinical and Experimental Neuropsychology*, *28*(1), 67–83. https://doi.org/10.1080/13803390490918318

Casaletto, K. B., & Heaton, R. K. (2017). Neuropsychological Assessment: Past and Future. *Journal of the International Neuropsychological Society*, *23*(9–10), 778–790. https://doi.org/10.1017/S1355617717001060

Cheon, S.-M., Park, M. J., Kim, W.-J., & Kim, J. W. (2009). Non-Motor Off Symptoms in Parkinson's Disease. *Journal of Korean Medical Science*, *24*(2), 311. https://doi.org/10.3346/jkms.2009.24.2.311

Christ, B. U., Combrinck, M. I., & Thomas, K. G. F. (2018). Both Reaction Time and Accuracy Measures of Intraindividual Variability Predict Cognitive Performance in Alzheimer's Disease. *Frontiers in Human Neuroscience*, *12*, 124. https://doi.org/10.3389/fnhum.2018.00124

Cohen, A. S. (2019). Advancing ambulatory biobehavioral technologies beyond "proof of concept": Introduction to the special section. *Psychological Assessment*, *31*(3), 277–284. https://doi.org/10.1037/pas0000694

Cormack, F., McCue, M., Taptiklis, N., Skirrow, C., Glazer, E., Panagopoulos, E., van Schaik, T. A., Fehnert, B., King, J., & Barnett, J. H. (2019). Wearable Technology for High-Frequency Cognitive and Mood Assessment in Major

Depressive Disorder: Longitudinal Observational Study. *JMIR Mental Health*, *6*(11), e12814. https://doi.org/10.2196/12814

Cowan, N. (1998). *Attention and Memory: An Integrated Framework*. Oxford University Press.

Cronin-Golomb, A., Reynolds, G. O., Salazar, R. D., & Saint-Hilaire, M.-H. (2019). Parkinson's Disease and Parkinson-Plus Syndromes. In M. L. Alosco & R. A. Stern (Eds.), *The Oxford Handbook of Adult Cognitive Disorders* (pp. 599–630). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190664121.013.28

Csikszentmihalyi, M., & Larson, R. (1987). Validity and Reliability of the Experience-Sampling Method. *Journal of Nervous*, *175*(9), 526–536.

da Silva, F. C., Iop, R. da R., de Oliveira, L. C., Boll, A. M., de Alvarenga, J. G. S., Gutierres Filho, P. J. B., de Melo, L. M. A. B., Xavier, A. J., & da Silva, R. (2018). Effects of physical exercise programs on cognitive function in Parkinson's disease patients: A systematic review of randomized controlled trials of the last 10 years. *PLOS ONE*, *13*(2), e0193113. https://doi.org/10.1371/journal.pone.0193113

Dagum, P. (2018). Digital biomarkers of cognitive function. *Npj Digital Medicine*, *1*(1), 10. https://doi.org/10.1038/s41746-018-0018-4

Daniëls, N. E. M., Bartels, S. L., Verhagen, S. J. W., Van Knippenberg, R. J. M., De Vugt, M. E., & Delespaul, Ph. A. E. G. (2020). Digital assessment of working memory and processing speed in everyday life: Feasibility, validation, and lessons-learned. *Internet Interventions, 19,* 100300. https://doi.org/10.1016/j.invent.2019.100300

de Frias, C. M., Dixon, R. A., Fisher, N., & Camicioli, R. (2007). Intraindividual variability in neurocognitive speed: A comparison of Parkinson's disease and normal older adults. *Neuropsychologia*, *45*(11), 2499–2507. https://doi.org/10.1016/j.neuropsychologia.2007.03.022

Diehl, M., Hooker, K., & Sliwinski, M. J. (2014). *Handbook of Intraindividual Variability Across the Life Span*. Routledge.

Dinstein, I., Heeger, D. J., & Behrmann, M. (2015). Neural variability: Friend or foe? *Trends in Cognitive Sciences*, *19*(6), 322–328. https://doi.org/10.1016/j.tics.2015.04.005

Dirk, J., & Schmiedek, F. (2016). Fluctuations in elementary school children's working memory performance in the school context. *Journal of Educational Psychology*, *108*(5), 722–739. https://doi.org/10.1037/edu0000076

Duncan, J., Chylinski, D., Mitchell, D. J., & Bhandari, A. (2017). Complexity and compositionality in fluid intelligence. *Proceedings of the National Academy of Sciences, 114*(20), 5295–5299. https://doi.org/10.1073/pnas.1621147114

Ellis, I. A., Henry C. Thomas, Roger L. Rodriguez. (1984). Emotional mood states and memory: Elaborative encoding, semantics processing, and cognitive effort. *Ournal of Experimental Psychology: Learning, Memory, and Cognition, 10*(3), 470–482.

Firth, J., Cotter, J., Elliott, R., French, P., & Yung, A. R. (2015). A systematic review and meta-analysis of exercise interventions in schizophrenia patients. *Psychological Medicine*, *45*(07), 1343–1361. https://doi.org/10.1017/S0033291714003110

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(27), E6106–E6115. https://doi.org/10.1073/pnas.1711978115

Fisher, A. J., Reeves, J. W., Lawyer, G., Medaglia, J. D., & Rubel, J. A. (2017). Exploring the idiographic dynamics of mood and anxiety via network analysis. *Journal of Abnormal Psychology*, *126*(8), 1044–1056. https://doi.org/10.1037/abn0000311

Friedman, J. H., Brown, R. G., Comella, C., Garber, C. E., Krupp, L. B., Lou, J.-S., Marsh, L., Nail, L., Shulman, L., & Taylor, C. B. (2007). Fatigue in Parkinson's disease: A review. *Movement Disorders*, *22*(3), 297–308. https://doi.org/10.1002/mds.21240

Gallagher, P., Nilsson, J., Finkelmeyer, A., Goshawk, M., Macritchie, K. A., Lloyd, A. J., Thompson, J. M., Porter, R. J., Young, A. H., Ferrier, I. N., McAllister-Williams, R. H., & Watson, S. (2015). Neurocognitive intra-individual variability in mood disorders: Effects on attentional response time distributions. *Psychological Medicine*, *45*(14), 2985. https://doi.org/10.1017/S0033291715000926

Gates, N. J., & Kochan, N. A. (2015). Computerized and on-line neuropsychological testing for late-life cognition and neurocognitive disorders: Are we there yet?

*Current Opinion in Psychiatry*, *28*(2), 165. https://doi.org/10.1097/YCO.0000000000000141

Germine, L., Reinecke, K., & Chaytor, N. S. (2019). Digital neuropsychology: Challenges and opportunities at the intersection of science and software. *The Clinical Neuropsychologist*, 1–16. https://doi.org/10.1080/13854046.2018.1535662

Gershon, R., Weiner, M., Rentz, D., Varela, M., Kaat, A., Condon, D., Nowinski, C., Mangravite, L., Omberg, L., Wagster, M., King, J., Prtrap, A., & Kellen, M. (2020, February 7). *MobileToolbox for Mobile Monitoring of Cognitive Change* [Symposium: Cognitive Testing In The Wild: Turning Smartphones into Digital Neuropsychological Assessment Tools]. International Neuropsychological Society Meeting, Denver, CO. https://www.mobiletoolbox.org/mobiletoolbox

Hassenstab, J., Aschenbrenner, A. J., Balota, D. A., McDade, E., Lim, Y., Fagan, A. M., Benzinger, T. L. S., Cruchaga, C., Goate, A., Morris, J. C., & Bateman, R. J. (2018). COMPARING SMARTPHONE-ADMINISTERED COGNITIVE ASSESSMENTS WITH CONVENTIONAL TESTS AND BIOMARKERS IN SPORADIC AND DOMINANTLY INHERITED ALZHEIMER DISEASE. *Alzheimer's & Dementia*, *14*(7), P224–P225. https://doi.org/10.1016/j.jalz.2018.06.2350

Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1992). Intraindividual change in text recall of the elderly. *Brain and Language*, *42*(3), 248–269. https://doi.org/10.1016/0093-934X(92)90100-S

Hilborn, J. V., Strauss, E., Hultsch, D. F., & Hunter, M. A. (2009). Intraindividual variability across cognitive domains: Investigation of dispersion levels and performance profiles in older adults. *Journal of Clinical and Experimental Neuropsychology*, *31*(4), 412–424. https://doi.org/10.1080/13803390802232659

Hirsch, M. A., van Wegen, E. E. H., Newman, M. A., & Heyn, P. C. (2018). Exercise-induced increase in brain-derived neurotrophic factor in human Parkinson's disease: A systematic review and meta-analysis. *Translational Neurodegeneration*, *7*(1), 7. https://doi.org/10.1186/s40035-018-0112-1

Hogan, D., Bailey, P., Black, S., Carswell, A., Chertkow, H., Clarke, B., Cohen, C., Fisk, J. D., Forbes, D., Man-Son-Hing, M., Lanctot, K., Morgan, D., & Thorpe, L. (2008). Diagnosis and treatment of dementia: 5. Nonpharmacologic and pharmacologic therapy for mild to moderate dementia. *Canadian Medical Association Journal*, *179*(10), 1019–1026. https://doi.org/10.1503/cmaj.081103

Hughes, A. J., Daniel, S. E., Kilford, L., & Lees, A. J. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease: A clinico-pathological study of 100 cases. *Journal of Neurology, Neurosurgery, and Psychiatry*, *55*(3), 181–184.

Hultsch, D. F., MacDonald, S. W. S., & Dixon, R. A. (2002). Variability in Reaction Time Performance of Younger and Older Adults. *The Journals of Gerontology: Series B*, *57*(2), P115. https://doi.org/10.1093/geronb/57.2.P101

Hultsch, D. F., MacDonald, S. W. S., Hunter, M. A., Levy-Bencheton, J., & Strauss, E. (2000). Intraindividual variability in cognitive performance in older adults:

Comparison of adults with mild dementia, adults with arthritis, and healthy adults. *Neuropsychology*, *14*(4), 588–598. https://doi.org/10.1037/0894-4105.14.4.588

Huntley, J. D., & Howard, R. J. (2010). Working memory in early Alzheimer's disease: A neuropsychological review. *International Journal of Geriatric Psychiatry*, *25*(2), 121–132. https://doi.org/10.1002/gps.2314

Hurd, M. D., Martorell, P., Delavande, A., Mullen, K. J., & Langa, K. M. (2013). Monetary costs of dementia in the United States. *The New England Journal of Medicine*, *368*(14), 1326–1334. https://doi.org/10.1056/NEJMsa1204629

Jefferies, L. N., Smilek, D., Eich, E., & Enns, J. T. (2008). Emotional valence and arousal interact in attentional control. *Psychological Science*, *19*(3), 290–295. https://doi.org/10.1111/j.1467-9280.2008.02082.x

Kane, R. L., & Parsons, T. D. (2017). *The Role of Technology in Clinical Neuropsychology*. Oxford University Press.

Katzell, R. A., & Thompson, D. E. (1990). An Integrative Model of Work Attitudes, Motivation, and Performance. *Human Performance*, *3*(2), 63. https://doi.org/10.1207/s15327043hup0302_1

Kelly, A. M. C., Uddin, L. Q., Biswal, B. B., Castellanos, F. X., & Milham, M. P. (2008). Competition between functional brain networks mediates behavioral variability. *NeuroImage*, *39*(1), 527–537. https://doi.org/10.1016/j.neuroimage.2007.08.008

Kievit, R. A., Davis, S. W., Griffiths, J., Correia, M. M., Cam-CAN, & Henson, R. N. (2016). A watershed model of individual differences in fluid intelligence.

*Neuropsychologia*, *91*, 186–198. https://doi.org/10.1016/j.neuropsychologia.2016.08.008

Kofler, M., Alderson, R., Raiker, J., Bolden, J., Sarver, D., & Rapport, M. (2014). Working memory and intraindividual variability as neurocognitive indicators in ADHD. *Neuropsychology*, *28*(3). http://diginole.lib.fsu.edu/islandora/object/fsu%3A209935/

Koo, B. M., & Vizer, L. M. (2019). Mobile Technology for Cognitive Assessment of Older Adults: A Scoping Review. *Innovation in Aging*, *3*(1). https://doi.org/10.1093/geroni/igy038

Krawczyk, D. C., Gazzaley, A., & D'Esposito, M. (2007). Reward modulation of prefrontal and visual association cortex during an incentive working memory task. *Brain Research*, *1141*, 168–177. https://doi.org/10.1016/j.brainres.2007.01.052

Ladouce, S., Donaldson, D. I., Dudchenko, P. A., & Ietswaart, M. (2017). Understanding Minds in Real-World Environments: Toward a Mobile Cognition Approach. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00694

Lange, E. B. (2005). Disruption of attention by irrelevant stimuli in serial recall. *Journal of Memory and Language*, *53*(4), 513–531. https://doi.org/10.1016/j.jml.2005.07.002

Lawson, R. A., Yarnall, A. J., Duncan, G. W., Breen, D. P., Khoo, T. K., Williams-Gray, C. H., Barker, R. A., Collerton, D., Taylor, J.-P., & Burn, D. J. (2016). Cognitive decline and quality of life in incident Parkinson's disease: The role of attention.

*Parkinsonism & Related Disorders*, *27*, 47–53. https://doi.org/10.1016/j.parkreldis.2016.04.009

Lieshout, M. van, Luman, M., Buitelaar, J., Rommelse, N. N. J., & Oosterlaan, J. (2013). Does neurocognitive functioning predict future or persistence of ADHD? A systematic review. *Clinical Psychology Review*, *33*(4), 539–560. https://doi.org/10.1016/j.cpr.2013.02.003

Lipsmeier, F., Taylor, K. I., Kilchenmann, T., Wolf, D., Scotland, A., Schjodt-Eriksen, J., Cheng, W.-Y., Fernandez-Garcia, I., Siebourg-Polster, J., Jin, L., Soto, J., Verselis, L., Boess, F., Koller, M., Grundman, M., Monsch, A. U., Postuma, R. B., Ghosh, A., Kremer, T., … Lindemann, M. (2018). Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial: Remote PD Testing with Smartphones. *Movement Disorders*, *33*(8), 1287–1297. https://doi.org/10.1002/mds.27376

Ljungberg, J. K., & Neely, G. (2007). Stress, subjective experience and cognitive performance during exposure to noise and vibration. *Journal of Environmental Psychology*, *27*(1), 44–54. https://doi.org/10.1016/j.jenvp.2006.12.003

Luck, S. J., & Gold, J. M. (2008). The Construct of Attention in Schizophrenia. *Biological Psychiatry*, *64*(1), 34–39. https://doi.org/10.1016/j.biopsych.2008.02.014

MacDonald, S. W. S., Cervenka, S., Farde, L., Nyberg, L., & Bäckman, L. (2009). Extrastriatal dopamine D2 receptor binding modulates intraindividual variability in episodic recognition and executive functioning. *Neuropsychologia*, *47*(11), 2299–2304. https://doi.org/10.1016/j.neuropsychologia.2009.01.016

Manly, T., Lewis, G. H., Robertson, I. H., Watson, P. C., & Datta, A. K. (2002). *Coffee in the cornflakes: Time-of-day as a modulator of executive response control* (Vol. 40). https://doi.org/10.1016/S0028-3932(01)00086-0

Miller, J. B., & Barr, W. B. (2017). The Technology Crisis in Neuropsychology. *Archives of Clinical Neuropsychology*, *32*(5), 541–554. https://doi.org/10.1093/arclin/acx050

Moore, R. C., Swendsen, J., & Depp, C. A. (2017). Applications for self-administered mobile cognitive assessments in clinical research: A systematic review. *International Journal of Methods in Psychiatric Research*, *26*(4), e1562. https://doi.org/10.1002/mpr.1562

Murray, D. K., Sacheli, M. A., Eng, J. J., & Stoessl, A. J. (2014). The effects of exercise on cognition in Parkinson's disease: A systematic review. *Translational Neurodegeneration*, *3*(1), 5. https://doi.org/10.1186/2047-9158-3-5

Nesselroade, J. (2004). Intraindividual Variability and Short-Term Change: Commentary. *Gerontology*, *50*(1), 44–47.

Onnela, J.-P., & Rauch, S. L. (2016). Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health. *Neuropsychopharmacology*, *41*(7), 1691–1696. https://doi.org/10.1038/npp.2016.7

Pal, R., Mendelson, J., Clavier, O., Baggott, M. J., Coyle, J., & Galloway, G. P. (2016). Development and Testing of a Smartphone-Based Cognitive/Neuropsychological

Evaluation System for Substance Abusers. *Journal of Psychoactive Drugs*, *48*(4), 288–294. https://doi.org/10.1080/02791072.2016.1191093

Paolillo, E., Obermeit, L., Tang, B., Depp, C., Vaida, F., & Moore, R. (2018). Smartphone-based ecological momentary assessment (EMA) of alcohol and cannabis use in older adults with and without HIV infection. *Addictive Behaviors*, *83*, 102–108. https://doi.org/10.1016/j.addbeh.2017.10.016

Parsey, C. M., & Schmitter-Edgecombe, M. (2013). Applications of Technology in Neuropsychological Assessment. *The Clinical Neuropsychologist*, *27*(8), 1328–1361. https://doi.org/10.1080/13854046.2013.834971

Parsons, T. D., McMahan, T., & Kane, R. (2018). Practice parameters facilitating adoption of advanced technologies for enhancing neuropsychological assessment paradigms. *The Clinical Neuropsychologist*, *32*(1), 16–41. https://doi.org/10.1080/13854046.2017.1337932

Pessoa, L. (2009). How do emotion and motivation direct executive control? *Trends in Cognitive Sciences*, *13*(4), 160–166. https://doi.org/10.1016/j.tics.2009.01.006

Phillips, C. B., Edwards, J. D., Andel, R., & Kilpatrick, M. (2016). Daily Physical Activity and Cognitive Function Variability in Older Adults. *Journal of Aging and Physical Activity*, *24*(2), 256–267. https://doi.org/10.1123/japa.2014-0222

Poushter, J. (2016). Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies. *Pew Research Center*, 45.

Price, E., Moore, G., Galway, L., & Linden, M. (2017). Validation of a Smartphone-Based Approach to In Situ Cognitive Fatigue Assessment. *JMIR MHealth and UHealth*, *5*(8), e125. https://doi.org/10.2196/mhealth.6333

Rabbitt, P., Osman, P., Moore, B., & Stollery, B. (2001). There are stable individual differences in performance variability, both from moment to moment and from day to day. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *54*(4), 981–1003. https://doi.org/10.1080/713756013

Rabin, L. A., Spadaccini, A. T., Brodale, D. L., Grant, K. S., Elbulok-Charcape, M. M., & Barr, W. B. (2014). Utilization rates of computerized tests and test batteries among clinical neuropsychologists in the United States and Canada. *Professional Psychology: Research and Practice*, *45*(5), 368–377. https://doi.org/10.1037/a0037987

Ram, N., & Gerstorf, D. (2009). Time-structured and net intraindividual variability: Tools for examining the development of dynamic characteristics and processes. *Psychology and Aging*, *24*(4), 778–791. https://doi.org/10.1037/a0017915

Rentz, D. M., Dekhtyar, M., Sherman, J., Burnham, S., Blacker, D., Aghjayan, S. L., Papp, K. V., Amariglio, R. E., Schembri, A., Chenhall, T., Maruff, P., Aisen, P., Hyman, B. T., & Sperling, R. A. (2016). The Feasibility of At-Home iPad Cognitive Testing For Use in Clinical Trials. *The Journal of Prevention of Alzheimer's Disease*, *3*(1), 8–12. https://doi.org/10.14283/jpad.2015.78

Reynolds, G. O., Hanna, K. K., Neargarder, S., & Cronin-Golomb, A. (2017). The Relation of Anxiety and Cognition in Parkinson's Disease. *Neuropsychology*, *31*(6), 596–604. https://doi.org/10.1037/neu0000353

Roalf, D. R., Quarmley, M., Mechanic-Hamilton, D., Wolk, D. A., Arnold, S. E., Moberg, P. J., & Alzheimer's Disease Neuroimaging Initiative. (2016). Within-Individual Variability: An Index for Subtle Change in Neurocognition in Mild Cognitive Impairment. *Journal of Alzheimer's Disease: JAD*, *54*(1), 325–335. https://doi.org/10.3233/JAD-160259

Salazar, R. D., Moon, K., Neargarder, S., & Cronin-Golomb, A. (2019). Spatial judgment in Parkinson's disease: Contributions of attentional and executive dysfunction. *Behavioral Neuroscience*, *133*(4), 350–360.

Salthouse, T. A., & Berish, D. E. (2005). Correlates of Within-Person (Across-Occasion) Variability in Reaction Time. *Neuropsychology*, *19*(1), 77–87. https://doi.org/10.1037/0894-4105.19.1.77

Sandstrom, G. M., Lathia, N., Mascolo, C., & Rentfrow, P. J. (2017). Putting mood in context: Using smartphones to examine how people feel in different locations. *Journal of Research in Personality*, *69*, 96–101. https://doi.org/10.1016/j.jrp.2016.06.004

Schmidt, C., Collette, F., Cajochen, C., & Peigneux, P. (2007). A time to think: Circadian rhythms in human cognition. *Cognitive Neuropsychology*, *24*(7), 755–789. https://doi.org/10.1080/02643290701754158

Schuster, R. M., Mermelstein, R. J., & Hedeker, D. (2016). Ecological momentary assessment of working memory under conditions of simultaneous marijuana and tobacco use. *Addiction*, *111*(8), 1466–1476. https://doi.org/10.1111/add.13342

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological Momentary Assessment. *Annual Review of Clinical Psychology*, *4*(1), 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Sibley, B. A., & Beilock, S. L. (2007). Exercise and Working Memory: An Individual Differences Investigation. *Journal of Sport and Exercise Psychology*, *29*(6), 783–791. https://doi.org/10.1123/jsep.29.6.783

Sliwinski, M. J. (2008). Measurement-Burst Designs for Social Health Research. *Social and Personality Psychology Compass*, *2*(1), 245–261. https://doi.org/10.1111/j.1751-9004.2007.00043.x

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

Sliwinski, M. J., Smyth, J. M., Hofer, S. M., & Stawski, R. S. (2006). Intraindividual coupling of daily stress and cognition. *Psychology and Aging*, *21*(3), 545–557. https://doi.org/10.1037/0882-7974.21.3.545

Sonuga-Barke, E. J. S., & Castellanos, F. X. (2007). Spontaneous attentional fluctuations in impaired states and pathological conditions: A neurobiological hypothesis. *Neuroscience & Biobehavioral Reviews*, *31*(7), 977–986. https://doi.org/10.1016/j.neubiorev.2007.02.005

Sternin, A., Burns, A., & Owen, A. M. (2019). Thirty-Five Years of Computerized Cognitive Assessment of Aging—Where Are We Now? *Diagnostics*, *9*(3), 114. https://doi.org/10.3390/diagnostics9030114

Stuss, D. T., Murphy, K. J., Binns, M. A., & Alexander, M. P. (2003). Staying on the job: The frontal lobes control individual performance variability. *Brain*, *126*(11), 2363–2380. https://doi.org/10.1093/brain/awg237

Timmers, C., Maeghs, A., Vestjens, M., Bonnemayer, C., Hamers, H., & Blokland, A. (2014). Ambulant Cognitive Assessment Using a Smartphone. *Applied Neuropsychology: Adult*, *21*(2), 136–142. https://doi.org/10.1080/09084282.2013.778261

Tiplady, B., Oshinowo, B., Thomson, J., & Drummond, G. B. (2009). Alcohol and Cognitive Function: Assessment in Everyday Life and Laboratory Settings Using Mobile Phones. *Alcoholism: Clinical and Experimental Research*, *33*(12), 2094–2102. https://doi.org/10.1111/j.1530-0277.2009.01049.x

Tomlinson, C. L., Stowe, R., Patel, S., Rick, C., Gray, R., & Clarke, C. E. (2010). Systematic review of levodopa dose equivalency reporting in Parkinson's disease. *Movement Disorders: Official Journal of the Movement Disorder Society*, *25*(15), 2649–2653. https://doi.org/10.1002/mds.23429

Torous, J., Kiang, M. V., Lorme, J., & Onnela, J.-P. (2016). New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research. *JMIR Mental Health*, *3*(2), e16. https://doi.org/10.2196/mental.5165

Torous, J., Wisniewski, H., Carpenter, E., David, G., Eduardo, E., Fulford, D., & Synthia, G. (2019). *Creating a digital health smartphone app and digital phenotyping platform for mental health and diverse healthcare needs: An interdisciplinary and collaborative approach.*

Troyer, A. K., Vandermorris, S., & Murphy, K. J. (2016). Intraindividual variability in performance on associative memory tasks is elevated in amnestic mild cognitive impairment. *Neuropsychologia*, *90*, 110–116. https://doi.org/10.1016/j.neuropsychologia.2016.06.011

Trull, T. J., & Ebner-Priemer, U. (2013). Ambulatory Assessment. *Annual Review of Clinical Psychology*, *9*(1), 151–176. https://doi.org/10.1146/annurev-clinpsy-050212-185510

van der Heijden, K., de Sonneville, L. M. J., & Althaus, M. (2010). TIME-OF-DAY EFFECTS ON COGNITION IN PREADOLESCENTS: A TRAILS STUDY. *Chronobiology International*, *27*(9–10), 1870–1894. https://doi.org/10.3109/07420528.2010.516047

van der Velden, R. M. J., Mulders, A. E. P., Drukker, M., Kuijf, M. L., & Leentjens, A. F. G. (2018). Network analysis of symptoms in a Parkinson patient using experience sampling data: An n = 1 study: Symptom Network Analysis in Parkinson's Disease. *Movement Disorders*, *33*(12), 1938–1944. https://doi.org/10.1002/mds.93

Vasquez, B. P., Binns, M. A., & Anderson, N. D. (2017). Response Time Consistency Is an Indicator of Executive Control Rather than Global Cognitive Ability. *Journal* *of the International Neuropsychological Society*, 1–10. https://doi.org/10.1017/S1355617717001266

Vaughan, L., Leng, I., Dagenbach, D., Resnick, S. M., Rapp, S. R., Jennings, J. M., Brunner, R. L., Simpson, S. L., Beavers, D. P., Coker, L. H., Gaussoin, S. A., Sink, K. M., & Espeland, M. A. (2013). Intraindividual variability in domain-specific cognition and risk of mild cognitive impairment and dementia. *Current Gerontology and Geriatrics Research*, *2013*, 495793. https://doi.org/10.1155/2013/495793

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

Waters, A. J., Szeto, E. H., Wetter, D. W., Cinciripini, P. M., Robinson, J. D., & Li, Y. (2014). Cognition and craving during smoking cessation: An ecological momentary assessment study. *Nicotine & Tobacco Research*, *16*(Supp2), S118. https://doi.org/10.1093/ntr/ntt108

Weizenbaum, E., Torous, J., & Fuford, D. (in press). Cognition in Context: Understanding the everyday predictors of cognitive performance in a new era of measurement. *JMIR MHealth and UHealth*.

West, R., Murphy, K. J., Armilio, M. L., Craik, F. I. M., & Stuss, D. T. (2002). Effects of Time of Day on Age Differences in Working Memory. *The Journals of Gerontology: Series B*, *57*(1), P10. https://doi.org/10.1093/geronb/57.1.P3

Whitbourne, S. B., Neupert, S. D., & Lachman, M. E. (2008). Daily Physical Activity: Relation to Everyday Memory in Adulthood. *Journal of Applied Gerontology*, *27*(3), 331–349. https://doi.org/10.1177/0733464807312175

Wilcockson, T. D. W., Ellis, D. A., & Shaw, H. (2018). Determining Typical Smartphone Usage: What Data Do We Need? *Cyberpsychology, Behavior, and Social Networking*, *21*(6), 395–398. https://doi.org/10.1089/cyber.2017.0652

Wu, J. Q., & Cronin-Golomb, A. (2019). Temporal associations between sleep and daytime functioning in Parkinson's disease: A smartphone-based ecological momentary assessment. *Behavioral Sleep Medicine*.

Yeo, G. B., & Neal, A. (2004). A Multilevel Analysis of Effort, Practice, and Performance: Effects; of Ability, Conscientiousness, and Goal Orientation. *Journal of Applied Psychology*, *89*(2), 231–247. https://doi.org/10.1037/0021-9010.89.2.231

Zhan, A., Mohan, S., Tarolli, C., Schneider, R. B., Adams, J. L., Sharma, S., Elson, M. J., Spear, K. L., Glidden, A. M., Little, M. A., Terzis, A., Dorsey, E. R., & Saria, S. (2018). Using Smartphones and Machine Learning to Quantify Parkinson Disease Severity: The Mobile Parkinson Disease Score. *JAMA Neurology*, *75*(7), 876. https://doi.org/10.1001/jamaneurol.2018.0809

**CURRICULUM VITAE**

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]