

**Within-Person Dynamics between Lifestyle Factors and Cognitive Functioning using
Accelerometer-Determined Physical Activity and Mobile Cognitive Assessments**

by

Rebecca Vendittelli
B.A., York University, 2014
M.Sc., University of Victoria, 2017

A Dissertation Submitted in Partial Fulfilment of the Requirements for the Degree of
DOCTOR OF PHILOSOPHY
In the Department of Psychology

© Rebecca Vendittelli, 2023
University of Victoria

All rights reserved. This Dissertation may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.

We acknowledge and respect the ləkʷəŋən peoples on whose traditional territory the university stands, and the Songhees, Esquimalt and WSÁNEĆ peoples whose historical relationships with the land continue to this day.

Supervisory Committee

**Within-Person Dynamics between Lifestyle Factors and Cognitive Functioning using
Accelerometer-Determined Physical Activity and Mobile Cognitive Assessments**

by

Rebecca Vendittelli
B.A., York University, 2014
M.Sc., University of Victoria, 2017

Supervisory Committee

Dr. Scott Hofer, Supervisor
Department of Psychology

Dr. Colette Smart, Departmental Member
Department of Psychology

Dr. Jonathan Rush, Departmental Member
Department of Psychology

Dr. Sam Liu, Outside Member
School of Exercise Science, Physical & Health Education

Abstract

The Canadian population is fundamentally changing such that the proportion of seniors is expected to be one in four by 2030 (Government of Canada, 2014). This shift will undoubtedly be accompanied by a surge in the prevalence of age-related health issues, including cognitive decline and dementia. Compounded by increased life expectancy, this demographic change is expected to overwhelm the health care system (Wister & Speechley, 2015) and have grave economic impact (Wimo et al., 2013, 2017). As such, researchers have endeavoured to find innovative and efficient solutions that are preventative, rather than reactive. Lifestyle interventions, such as physical activity (PA) and stress reduction, have gained ample support for their role in protecting against cognitive decline. In tandem, digital cognitive assessment tools have also been developed to support the anticipated demand for efficient screening. Also known as mobile assessments, this state-of-the-art technology can simultaneously assess contextual, psychosocial, and lifestyle factors along with cognition. In this way, a nuanced understanding of the temporal association between cognition and lifestyles variables may be explored. To date, however, there is little research examining these relationships. Chapter 1 reviews psychometric evidence for mobile cognitive assessments and their efficacy in measuring cognitive functioning and daily variability, as well as provides results of a psychometric replication study. Both Chapters 2 (focusing on PA) and 3 (focusing on stress) look at between-person and within-person differences regarding how these lifestyle factors influence cognitive performance. More specifically, Chapter 2 presents results on the relationship between daily variation in PA and cognition, and Chapter 3 examines the relationship between momentary and daily stress and cognition. Chapter 4 provides a brief summary of potential clinical implications for advances in mobile and remote cognitive assessments and potential for lifestyle interventions.

Table of Contents

Supervisory Committee	ii
Abstract	iii
Table of Contents	iv
List of Tables	vii
List of Figures	viii
List of Appendices	ix
Acknowledgements	x
Chapter 1	1
Abstract	2
Background	4
Table 1.1 Summary of Findings Across Studies Employing DM, SS, and Grid	8
Table 1.2 Summary of Practice Effect Findings Across Studies Employing DM, SS, and Grid	12
Method	14
Procedure	14
Participants	15
Measures	16
Figure 1 Brain Games used in DASH	19
Analytical Approach	19
Results	22
Demographic and Study Characteristics	22
Table 1.3 Baseline Characteristics of Participants (N=66)	24
ICC and BP Reliabilities	24
Table 1.4 ICC's and Between-Person Reliabilities for all Cognitive Measures	25
WP Reliabilities	25
Table 1.5 Within-Person Reliabilities Across Sessions and Days for all Cognitive Measures	26
Figure 1.1a Dot Memory Accuracy Score Trajectories for a Random Subset of Participant Across Sessions	26
Figure 1.1b Stroop Reaction Time Trajectories for a Random Subset of Participant Across Sessions	27
Figure 1.1c Symbol Search Reaction Time Trajectories for a Random Subset of Participant Across Sessions	27

Figure 1.2a Dot Memory Accuracy Score Trajectories for a Random Subset of Participant Across Days	28
Figure 1.2b Stroop Reaction Time Trajectories for a Random Subset of Participant Across Days	28
Figure 1.2c Symbol Search Reaction Time Trajectories for a Random Subset of Participant Across Days	29
Practice Effects	29
Table 1.6 Multilevel Estimates of the Effects of Session, Effort, and Demographic Factors on Cognitive Tasks	30
Figure 1.3a Dot Memory Accuracy Scores Across Sessions Within the Study for Individual Participants with a Line of Best Fit.....	31
Figure 1.3b Stroop Reaction Time Scores Across Sessions Within the Study for Individual Participants with a Line of Best Fit.....	31
Figure 1.3c Symbol Search Reaction Time Scores Across Sessions Within the Study for Individual Participants with a Line of Best Fit.	32
Table 1.7 Multilevel Estimates of the Effects of Day, Effort, and Demographic Factors on Cognitive Tasks	33
Figure 1.4a Dot Memory Accuracy Scores Across Days Within the Study for Individual Participants with a Line of Best Fit.....	34
Figure 1.4b Stroop Reaction Time Scores Across Days Within the Study for Individual Participants with a Line of Best Fit.....	34
Figure 1.4c Symbol Search Reaction Time Scores Across Days Within the Study for Individual Participants with a Line of Best Fit.	35
Discussion and Limitations.....	35
References.....	41
Appendix 1.1 Count of Scores for each Cognitive Test	47
Chapter 2.....	49
Abstract	50
Background.....	51
Methods.....	58
Procedure	58
Measures	58
Analytical Approach	60
Results.....	60
Correlation Between Objective and Subjective PA	60

MLM Results for WP and BP Effects of PA on Cognition	61
Table 2.1 Multilevel Estimates of the Effects of Session, Effort, and Demographic Factors, and Physical Activity on Cognitive Tasks	61
Discussion	62
References	65
Appendix 2.1 Count Distribution of Participant’s Study Mean Fitbit Activity Level.....	72
Chapter 3.....	73
Abstract	74
Background.....	75
Method	79
Procedure	79
Measures	80
Analytical Approach	83
Results.....	84
Table 3.1 Multilevel Estimates of the Effects of Session, Effort, and Demographic Factors, and Momentary Subject Stress Appraisal on Cognitive Tasks (Model 1A).....	85
Table 3.2 Multilevel Estimates of the Effects of Session, Effort, and Demographic Factors, and Daily and Participant Mean Stress on Cognitive Tasks (Model 2).....	86
Discussion.....	86
References.....	92
Chapter 4.....	98
Abstract	99
Introduction.....	100
Large-Scale Research and Clinical Potential.....	103
WP Variability and Psychosocial/Contextual Factors	106
Summary and Conclusion.....	110
References.....	112

List of Tables

Table 1.1 Summary of Findings Across Studies Employing M2C2 Cognitive Tasks

Table 1.2 Summary of Practice Effect Findings Across Studies Employing M2C2 Cognitive Tasks

Table 1.3 Baseline Characteristics of Participants (N=66)

Table 1.4 ICC's and Between-Person Reliabilities for all Cognitive Measures

Table 1.5 Within-Person Reliabilities Across Sessions and Days for all Cognitive Measures

Table 1.6 Multilevel Estimates of the Effects of Session, Effort, and Demographic Factors on Cognitive Tasks

Table 1.7 Multilevel Estimates of the Effects of Day, Effort, and Demographic Factors on Cognitive Tasks

Table 2.1 Multilevel Estimates of the Effects of Session, Effort, and Demographic Factors, and Physical Activity on Cognitive Tasks

Table 3.1 Multilevel Estimates of the Effects of Session, Effort, and Demographic Factors, and Momentary Subject Stress Appraisal on Cognitive Tasks (Model 1)

Table 3.2 Multilevel Estimates of the Effects of Session, Effort, and Demographic Factors, and Daily and Participant Mean Stress on Cognitive Tasks (Model 2)

List of Figures

Figure 1.1a Dot Memory Accuracy Score Trajectories for a Random Subset of Participant Across Sessions

Figure 1.1b Stroop Reaction Time Trajectories for a Random Subset of Participant Across Sessions

Figure 1.1c Symbol Search Reaction Time Trajectories for a Random Subset of Participant Across Sessions

Figure 1.2a Dot Memory Accuracy Score Trajectories for a Random Subset of Participant Across Days

Figure 1.2b Stroop Reaction Time Trajectories for a Random Subset of Participant Across Days

Figure 1.2c Symbol Search Reaction Time Trajectories for a Random Subset of Participant Across Days

Figure 1.3a Dot Memory Accuracy Scores Across Sessions Within the Study for Individual Participants with a Line of Best Fit.

Figure 1.3b Stroop Reaction Time Scores Across Sessions Within the Study for Individual Participants with a Line of Best Fit.

Figure 1.3c Symbol Search Reaction Time Scores Across Sessions Within the Study for Individual Participants with a Line of Best Fit.

Figure 1.4a Dot Memory Accuracy Scores Across Days Within the Study for Individual Participants with a Line of Best Fit.

Figure 1.4b Stroop Reaction Time Scores Across Days Within the Study for Individual Participants with a Line of Best Fit.

Figure 1.4c Symbol Search Reaction Time Scores Across Days Within the Study for Individual Participants with a Line of Best Fit.

List of Appendices

Appendix 1.1 Count of Scores for each Cognitive Test

Appendix 2.1 Count Distribution of Participant's Study Mean Fitbit Activity Level

Acknowledgements

I would like to acknowledge how exceptionally privileged and fortunate I am to have had the opportunity to attend graduate school. I feel very lucky to have been able to choose a path which has allowed me to merge my passion for understanding the human experience, including my own, with my profession. My experience throughout this journey has been enriched by many inspirational and supportive individuals, to whom I owe the sincerest thank you. I would first like to acknowledge my research supervisor, Dr. Scott Hofer, for his support, enthusiasm, and generosity. Your professional guidance, cushioned with warmth and encouragement, has made working with you a delight. I would also like to thank my committee members, Dr. Colette Smart and Dr. Sam Liu for lending their expertise on this project to support my development as a researcher. I would like to extend an extra thank you to Dr. Smart for her mentorship as my clinical supervisor. You have played a critical role in my professional development as an aspiring neuropsychologist, and I will forever be grateful for the opportunity to have learned from you. I am sincerely grateful to have been part of the iLifespan laboratory and would like to thank my fellow lab members, Dr. Jonathan Rush, Dr. Tomiko Yoneda, Dr. Nathan Lewis, Dr. Jamie Knight, and Dr. Raquel Grahm for their ample contribution to this project, as well as the many ways in which they uniquely supported and advised me throughout. I am further grateful for Dr. Andrea Piccinin's mentorship early on as my Master's supervisor; the support you offered during my transition into graduate studies was invaluable. I would like to thank my fellow cohort members for making this journey an enjoyable one, and well as my endlessly supportive family and friends for their love and encouragement. Finally, this project would not have been possible without funding support from AGE-WELL. I thank this organization, as well as the Canadian Institutes of Health Research (CIHR), the Integrative Analysis of Longitudinal

Studies of Aging research network, and the University of Victoria for their financial assistance for my graduate research.

Chapter 1

Psychometric Properties of a Mobile Based Cognitive Assessment Tool to Detect Early Cognitive Changes

Abstract

Objective: Digitalized cognitive assessments are growing in popularity due to the numerous advantages they afford. However, the literature is vast in terms of tasks/batteries used, versions of tests, and modality (e.g., computerized versus mobile phone), among other factors. The purpose of this study is to summarize the literature on studies using a common smart-phone based cognitive battery, as well as replicate and extend previous findings by evaluating between-person (BP) and within-person (WP) reliability of three cognitive tasks (Dot Memory, Stroop, and Symbol Search), and examine practice effects across different time schedules.

Method: Healthy community-residing older adults ($n=66$; $M_{age}=70.75$ years) from Victoria, B.C. completed surveys up to 4 times per day for 14 days. These surveys concluded with the cognitive measures. Multi-level models (MLMs) were used to estimate intraindividual and interindividual variation in cognitive performance across various time scales (sessions and days) to capture BP and WP variability in cognitive performance, as well as practice effects.

Results: Findings show high BP reliabilities across the three cognitive tests based on aggregate scores ($>.96$). WP reliabilities were less robust, with Dot Memory showing the lowest systematic variation in individual test performance ($.10-.24$), and the two reaction time-based tasks (Symbol Search and Stroop) showing relatively higher WP reliabilities ($>.56$), across sessions and days. Test performance across all 3 tasks were significant for both linear and quadratic effects of time across sessions and days, meaning that practice effects are observed early in the study, and then level off. Demographic factors and self-reported effort were largely unrelated to cognitive outcomes, with few exceptions.

Conclusions: Findings from the present study support the BP reliability of the cognitive tasks, as well as elucidate important considerations for future studies. WP reliabilities were shown to be

moderate to strong on reaction-time based measures, supporting their potential in detecting systematic WP change due to time varying factors. Additionally, practice-effects observed early in the study have implications for future methodological design and analyses.

Background

Dementias are progressive and irreversible neurodegenerative disorders of varying etiologies and forms and include (but are not limited to) Alzheimer's Disease, vascular dementia, Frontotemporal dementia, Lewy body dementia, and mixed dementias (Adlimoghaddam, Roy, & Albeni, 2018). It has been well established that patients experience changes in their cognitive abilities well before being diagnosed. In fact, early changes have been observed up to 10 years prior to a diagnosis (Hassenstab et al., 2016; Karr et al., 2018; Rabin, Smart, & Amariglio, 2017), and approximately two thirds of patients screen positive for possible early dementia at the time of first assessment (Claveau, Presse, Kergoat, & Villalpando, 2018). The early stage of dementia is often characterized by mild cognitive impairment (MCI). However, patients diagnosed with MCI do not always go on to develop dementia. Individuals may not transition for a variety of reasons, one of those being appropriate intervention. For example, physical activity has been associated with a reduced risk of transitioning forward from mildly impaired to severely impaired, and impaired to death (Yoneda et al., 2021). As such, early detection is paramount and has been the focus of screening initiatives. Traditionally, multi-domain screening tests are administered within a physician's office. However, more recently, self-administered computerized cognitive assessments are burgeoning within the field of neuropsychology as traditional face-to-face approaches are becoming impractical in meeting the high demand for neurocognitive screening and assessment within healthcare settings.

mHealth Assessments

Mobile health (mHealth) assessments are typically completed on a computer or smartphone, with the benefit of being timelier and more cost-effective compared to traditional face to face assessments. Among mHealth assessments, study protocols, tests, and test batteries are vastly different. That is, testing may more closely resemble in-person testing in breadth and

duration on lengthier, and less frequently administered computerized measures (see Zygoris & Tsolaki, 2015 for a full review). These tests are sometimes administered with the support by a trained professional, while others are not. On the other hand, some researchers have embedded briefer and more frequent testing (e.g., assessments spanning minutes that are often completed hours apart, usually across days to weeks) into ecological-momentary-assessment protocols (EMA) that are completed autonomously (i.e., unsupervised, and outside of a research or clinical setting) using smartphones. In this way, measures of psychosocial, contextual, and lifestyle factors on relatively shorter times-scales have been simultaneously assessed along with cognition in some instances.

MHealth assessments are not only advantageous for the flexibility in measurement scheduling they afford (Cain, Depp, & Jeste, 2009), but also in that they are completed within the individuals every-day context, enhancing ecological validity compared to in-person testing, where the goal is to obtain the individuals best performance in a quiet and standardized testing environment (Sliwinski et al., 2018). Additionally, they are more amenable to repeated administration for monitoring purposes. In this way, the cognitive trajectory of an individual may be tracked over time so early, subtle changes may be detected for streamlined referral and intervention. Moreover, compared to single time point measurements, aggregating across repeated mHealth assessments on a shorter time scale results in a more accurate estimation of an individual's true score through "cancelling out" the effects of random and systemic time varying factors (Shiffman, Stone, & Hufford, 2008; Sliwinski et al., 2018). For example, Allard et al. (2014) found that aggregate scores from a repeatedly administered mobile semantic memory test were associated with early detection of hippocampal changes in healthy older adults, while the same test administered traditionally (i.e., one in-person administration in lab) showed no

association. This cancelling out effect is especially important in neurodivergent populations, as their cognitive performance can vary depending on the time of day (Wilks et al., 2021) and across days (Matar, Shine, Halliday, & Lewis, 2020).

At the same time, time-varying factors which can influence cognition may be simultaneously assessed within EMA protocols. As such, repeated assessments can facilitate an understanding of systematic (rather than random) intraindividual variability in cognitive performance across assessment occasions (Moore, Swendsen, & Deep, 2016; Sliwinski, 2009; Sliwinski et al., 2018). Tests that are sensitive to this within-person (WP) variance are reliable in that they capture fluctuations in cognition that may be due to factors like stress, sleep, physical activity, etc. Notably, Salthouse and Berish (2005) have demonstrated WP variability to be similar in magnitude to between-person (BP) variability (i.e., differences between individual means), meaning that on any single occasion, an individual may perform near the level of the best or worst performer in the sample. These findings have important implications for traditional, single time point assessments, by challenging the perception of cognition as static rather than dynamic (Salthouse & Berish, 2005). As an appropriate first step, the purpose of the majority of EMA mHealth cognitive assessment studies have been to determine feasibility and demonstrate psychometric integrity.

Psychometric Properties and Feasibility of EMA Cognitive Testing

Though limited in numbers, a variety of EMA assessment designs have been examined in the literature (See Moore, Swendsen, & Depp, 2017 for their review of 12 studies). Overall, findings across indicators of adherence, acceptance, and attrition have been promising (Brewster, Rush, Ozen, Vendittelli, & Hofer, 2021; Hassenstab et al., 2020; Jongstra et al., 2017; Lancaster et al., 2020; Mackin et al., 2018; Nicosia et al., 2022; Öhman, Hassenstab, Berron, Schöll, & Papp, 2021; Papp et al., 2021; Schweitzer et al., 2017). Reliability has been demonstrated across

various mHealth assessments that include tests of processing speed and working memory (Brewster et al., 2021; Brouillette et al., 2013; Nicosia et al., 2022; Sliwinski et al., 2018; Thompson et al., 2022) and the validity of mHealth tasks has been demonstrated against standardized neuropsychometric testing for tests of attention and processing speed (Brewster et al., 2021; Brouillette et al., 2013), memory (Brewster et al., 2021; Lancaster et al., 2020; Schweitzer et al., 2017), and executive functioning (Jongstra et al., 2017).

Although the emerging research on the psychometric integrity and feasibility of EMA mHealth tests are promising, it is essential to replicate and expand on findings from a common battery to build evidence for use within the clinical and research setting. In the aforementioned studies, Sliwinski et al. (2018) and Brewster et al. (2021) have both used the same visual working memory task called Dot Memory (DM), and Nicosia et al. (2022) and Thompson et al. (2022) used the same processing speed task as these authors, called Symbol Search (SS, where participants were required to quickly scan the screen and tap a card on the top row which contained matching symbols as one on the bottom row). Nicosia et al. also included a variation of DM, called Grid, where instead of having to recall where three dots were on a 5x5 grid, they used 3 common objects (keys, cellphone, and pen). Table 1 summarizes general findings from these studies regarding their overlapping cognitive tasks.

Table 1.1 Summary of Findings Across Studies Employing DM, SS, and Grid

Authors	Overlapping Cognitive Task	Demographics	Study Characteristics	Feasibility ^a	BP Reliability	WP Reliability
Brewster et al., 2021	DM, SS	Healthy older adults, N=17 (MA=73, SD=5.79)	5 days, 2 surveys/day, with 5 follow up bursts across the year.	94% retention and 87.5% compliance;	≥.80 across all bursts but one for DM, ≥.94 across all bursts for SS	NA
Sliwinski et al., 2018	DM, SS	Healthy adults, N=219 (MA=47, SD=10.74)	14 days, 5 surveys/day	85% compliance	≥.97 for DM and SS	0.41 to 0.53
Nicosia et al., 2022	Grid, SS	Healthy older adults, N=268 (MA=77, SD=5.7) and MI older adults, N=22 (MA=77, SD=6.1)	7 days, 4 surveys/day, with 6-month and 1-year follow up bursts	79-81% adherence, with drop-out rates below 5% for both groups	≥.90 for Grid and SS	NA
Thompson et al., 2022	SS	Healthy older adults, N=52 (61-80 years-old)	8 days, 3 surveys/day	93% compliance	.96 for SS	0.97

Notes. DM = Dot Memory, SS = Symbol Search, MA = Mean Age, SD = Standard Deviation, MI = Mildly Impaired
^aNot all feasibility results are summarized within the table.

Study Design and Feasibility Brewster et al.'s (2021) measurement burst design study consisted of five bursts over the course of 1 year, with each burst lasting 5 days each (containing two assessments per day) in a sample of seventeen healthy older adults living in Victoria, BC (mean age=73, SD=5.79). The Brewster et al. study demonstrated solid retention with 94% of participants completing the full-year (with one dropping out at month 9 due to concerns about COVID-19). Compliance was high, as 97.5% of all scheduled mobile assessments were completed within the allotted window of time. One hundred percent of participants reported that the intensive measurement design was manageable, and a mean of 78.63 and 85.64 was reported for user-friendliness and study enjoyment, respectively, with 100 being the highest possible positive rating for each. Sliwinski et al.'s (2018) study consisted of 219 healthy adults (mean

age=47, SD=10.74) recruited from the New York City Registered Voters List in the Bronx. Participants completed 14 days of ecological momentary assessment entailing five survey notifications per day. Similarly, participants in their study responded to an average of 85% of the surveys, with 79% of the surveys being completed within ten minutes of the prompt. Ninety-three percent compliance was observed in Thompson et al.'s study (2022), with 52-healthy older adults (61-80 years-old) asked to complete three surveys/day, for 8 days. Participants in Nicosia et al.'s (2022) study were healthy (N=268, mean age=77, SD=5.7) and mildly impaired older adults (N=22, mean age=77, SD 6.1) enrolled in ongoing studies of aging and dementia at the Charles F. and Joanne Knight Alzheimer Disease Research Center (Knight ADRC) at Washington University School of Medicine in St. Louis. Their burst design study required 1 week of surveys with a 6-month and 1-year follow up. Participants completed four sessions of testing per day, for 7 days. The authors reported high feasibility (86.50 % of those approached chose to enroll), with 80.42% adherence and 4.82% drop out. In summary, compliance and adherence have been demonstrated to be exceptional (>80%) across studies of varying designs and measurement schedules, in both healthy and mildly impaired adults. Of note, participants across studies received varying degrees of compensation for study involvement; Brewster et al., did not specify, Thompson et al. provided \$20.00 for returning the device, and Nicosia et al. and Sliwinski et al. provided remuneration that appeared to be related to the number of surveys completed.

Reliability BP reliability refers to the proportion of variance in cognitive scores due to differences between individuals (as opposed to fluctuations in scores within-individuals due to systematic factors and randomness), while WP reliability is the variation in individual test performance that is due to systematic, time-varying factors (such as, for example, stress, sleep,

etc., as opposed to randomness). All studies examined BP reliabilities. In Brewster et al.'s (2021) study, DM showed BP reliabilities at or above the range deemed acceptable (.80) across all bursts except one, where it was .79. SS BP reliabilities were at or above .94 across bursts. Sliwinski et al. (2018) found that when using the mean score across measurement occasions (five administrations per day, for 14-days), BP reliabilities were enhanced to between 97% and 98%. Reliabilities were comparable ($>.90$) in Nicosia et al.'s (2022) study for the weekly aggregate, and in Thompson et al. (2022) (.96 for SS). These findings demonstrate strong BP reliability when aggregating across various time scales, including days, weeks, and longitudinal bursts over the year.

Sliwinski et al. (2018) additionally examined WP reliability of DM and SS, and found .50 and .53 of the WP variances, respectively, reflected systematic variance across assessment occasions (i.e., meaningful fluctuation in scores across each testing occasion, as opposed to random error derived from trial-to-trial variance within one administration of a test). However, time-varying factors that may be related to this variance (e.g., stress, activity, sleep, etc.) were not explored. WP reliability of SS was relatively higher in Thompson et al.'s (2022) study, at .97. Differences in WP reliability may be related to population differences (older adults versus adults), or the way in which reliability was calculated, as Thompson et al. does not delineate the analytic approach. Though WP reliabilities were not examined in Brewster et al. (2021), time-varying factors were included in their models as predictors (affect, distractibility, pain, fatigue, subjective effort), though none yielded significant results suggesting that no coupled association with cognitive performance on testing was observed for these variables.

Practice Effects Practice effects were examined to an extent across studies (See Table 2). In Brewster et al.'s (2021) study (five bursts, over the course of 1 year) there was a significant

linear slope effect across burst for DM, indicating improved performance over the year. However, there was no linear or quadratic slope effect within bursts (i.e., no practice effects across the 5 days within a burst). For Symbol Search, both linear and quadratic slopes were significant across bursts, suggesting progressively faster performance over the year, with a gradual slowing of gains over time. Additionally, there was a significant linear within-burst slope, indicating progressively faster performance across days within each burst on this speeded task. Neither task showed a cross-level interaction, meaning that individual slopes within a burst were not changing over time. In keeping with these findings, Thomsson et al. (2022) report that reaction times on SS improved over the first 4 to 5 days of their study, before leveling off. However, Nicosia et al. (2022) did not find improvements on their analogous Grid task across timepoint (6-month and 1-year follow up), but in keeping with Brewster et al.'s finding, SS improved from visit 1 to 2, but not between 2 and 3, suggesting that practice effects may have diminished after completion of the first testing cycle. Day to day slopes were not examined. Taken together, these findings may suggest that speeded tasks, rather than accuracy tasks, are more sensitive to the effects of practice across days, while both types of tests may be susceptible to practice effects across longer periods, though these effects may be pronounced in reaction time tasks.

In regard to the impact of practice effects, rather than looking at slopes, Sliwinski et al. (2018) examined whether reliability remained constant across the 2-week ambulatory assessment period. The BP reliabilities of the average scores for each day remained stable across the 14 days; the standard deviations of daily reliabilities were .01 for SS and .02 for DM. The authors concluded that increased practice with the tasks did not alter the reliability of aggregated scores. Null practice effects from day-to-day in this study, compared to practice effected observed in

Brewster et al.'s and Thompson et al.'s study (for SS) may be due to differences in the way practice effects were examined, but more likely, due to procedural differences in that the participants in Sliwinski et al.'s study had 2 days of practice with the EMA protocol before being formally enrolled in the study (to determine compliance). It is possible that practice effects were present across these days, and gains diminished at the onset of the study.

Table 1.2 Summary of Practice Effect Findings Across Studies Employing DM, SS, and Grid

Authors	Across Bursts	Across Days
Brewster et al., 2021	DM: sig. linear slope across bursts SS: sig. linear and quadratic slope across bursts	DM: NS across days within bursts SS: sig. linear slope within burst
Nicosia et al., 2022	Grid: NS across 3 bursts SS: sig. across burst 1 and 2, but not 2 and 3	NA
Thompson et al., 2022	NA	SS: "improved successively over the first 4 to 5 days before leveling off"

Notes. DM = Dot Memory, SS = Symbol Search, sig. = significant, and NS = nonsignificant.

Validity Construct validity was established through correlations with in-lab tasks across studies. Brewster et al. (2021) correlated the burst one aggregate score with the National Alzheimer's Coordinating Center (NACC) Uniform Data Set (UDS) neuropsychological test battery and found that DM (a visual short-term memory task) correlated moderately with measures of complex attention/working memory ($r = -.51$ to $-.54$), visual memory ($r = -.59$), and semantic memory ($r = .64$), whereas SS (a task primarily assessing attention and information processing speed) correlated moderately with measures of simple attention ($r = -.51$) and controlled sequential processing ($r = .66$). In Sliwinski et al.'s (2018) study, tasks correlated with in lab tests as hypothesized; SS correlated with speeded tasks ($r = .61$ to $.74$), and DM correlated

with in lab working memory tasks ($r = -.39$ to $-.45$). Sliwinski et al. further fit a confirmatory factor analysis (CFA). When allowing for covariance between tests that shared common method variance, their CFA model was deemed acceptable and standardized regression weights were all significant for the working memory factor ($>|.56|$) and speeded factor ($>|.66|$). Nicosia et al. (2022) showed that the mHealth cognitive tasks were correlated with conventional measures of the same domain (r 's = $-.22$ to $.57$), and further demonstrated that the composite scores from these measures were significantly associated with all AD biomarkers to a degree comparable to the conventional composite scores. Biomarkers included cerebral spinal fluid measures and neuroimaging findings.

The Daily experiences of Affect, Stress, and Health Study (DASH)

Considering the limited research on any one mHealth cognitive task or battery, the present study aims to replicate and expand on the aforementioned studies in order to corroborate the use of the DM and SS task described and substantiate findings. This research used two (DM and SS) of the three cognitive tasks previously examined, and include an additional Stroop (ST) task. This study aims to provide additional evidence as to whether, and to what extent scores on repeated mHealth assessments capture systematic BP and WP variability in cognitive performance through reliability analyses, and further quantify re-test effects using multilevel models to capture short-term trends (Sliwinski, Hoffman, & Hofer, 2010). Specifically, multi-level modeling (MLM) will be used to partition the variance in cognitive scores at the BP and WP level. In this way, BP and WP reliability indices may be calculated using the same formula as Sliwinski et al. WP reliabilities will be examined across sessions and across days. Finally, both linear and quadratic slopes will be examined across sessions and days within the study for each cognitive task, so that practice-effects may be elucidated and compared for varying time metrics.

As observed in past studies, it is hypothesized that the reliabilities of the aggregate variables will be high ($>.80$; Brewster et al., 2021), and greater than those observed on only one occasion. Moderate WP reliabilities similar to those reported by Sliwinski et al. (2018) are expected from day to day, while the examination of WP reliabilities from session to session is novel to this study and exploratory. It is expected that practice effects will be observed early in the study since participants did not receive prior practice with the tests like in Sliwinski et al.'s study, and that quadratic slopes, where the effects of practice level-off, will better characterize the data compared to a linear trajectory. These effects are expected to be pronounced on the two reaction time tasks (SS and ST), compared to the accuracy task (DM) in light of past findings.

Method

Procedure

The DASH study uses an EMA framework, with participants completing brief daily surveys on a laboratory provided Android device. Autonomously, they completed four brief surveys per day, for the duration of 14 days. They were instructed to complete morning and evening surveys within an hour of waking and an hour of going to sleep, respectively. Twice throughout the day, the mobile device notified participants to complete additional momentary surveys at quasi-random times. They had an hour to do so, and they were instructed to skip the survey past this time limit. Each survey included questions and rating scales about contextual, lifestyle, and psychosocial factors. All response options were presented categorically (checkbox) or on a sliding scale. The surveys concluded with “brain games” that assessed several domains of cognition, including the visual working memory task (DM), processing speed task (SS), and an executive functioning task of inhibition (ST).

Prior to beginning the EMA portion of the study, participants attended an intake session at the Institute on Aging and Lifelong Health at the University of Victoria. Consent was reviewed in full and obtained by participants. Participants then received an orientation about the study and training on how to complete the daily surveys. Near the end of the intake session, participants completed a short, computerized intake questionnaire through LimeSurvey (enquiring about demographic factors and psychosocial factors such as stress, physical activity, and mood). Participants were given a training guide for the study devices to take home with them. After two weeks of EMA, participants returned to the IALH for a debrief session, where they provided feedback about their experience in the study and returned all devices. Participants were provided with a \$75 honorarium (\$25 at intake and \$50 at debrief) for their participation. The project was approved by the University of Victoria Human Research Ethics Board in the spring of 2019 under ethics protocol number 18-1069.

Participants

Healthy, community residing older adults aged 65-75 years old were recruited for the study. Participants were sought through presentations, events, and dissemination of posters to local organizations in person and through email blasts sent out on local list-serves. Interested individuals contacted the research team and went through a scripted screening process and brief questionnaire with a lab member over the telephone. Individuals met eligibility if they could read and write in English, had not participated in similar studies at UVic, and were relatively healthy for their age. This meant excluding individuals if they had been diagnosed with a neurodegenerative cognitive disorder or psychiatric illness, and/or had a neurological history (i.e., epilepsy, stroke, and/or moderate to severe head injury), or any untreated medical condition that may affect cognition.

Measures

Demographic Demographic information was collected at the intake session and included age, sex, gender, and education. Age was measured in years and centered at the grand mean. Sex was assessed by a categorical item with the following response options: female, male, intersex, and other. Gender was assessed using an open-ended response option. Since all participants selected female or male, sex was used in analyses, with female as the reference group since most of the sample was female. Participants were also asked to indicate their highest level of education obtained from the following categorical response options in which higher values indicate more education: 1=did not complete high school; 2=high school; 3=some university/college; 4=trade school; 5=undergraduate degree; 6=graduate, law or medical school. Education was grouped and treated as a continuous scale ranging from 0 to 3 (amalgamated as follows: 0 = response options 1, 2, and 3; 1 = response options 4 and 5; 3 = response option 6).

Cognitive Participants completed cognitive measures at the end of each survey. Across surveys, DM, SS, and ST were completed (See Figure 1). The cognitive data is nested, in that each test consists of trials within a session, and sessions within a day, which can be further nested within each participant. As such, trial level data can be considered the raw data, with two scores created for each test; A session score (i.e., the mean score across trials for that session), and an end of day (EOD) score (the mean of the session scores from that day, i.e., the 4 survey occasions).

All occasions with incomplete trials were excluded, as this indicated that participants did not complete the task. This resulted in excluding 8 out of 42391 trials for SS, 94 out of 7098 trials for DM, 124 out of 11194 trials for Colour-Shape, and 246 out of 7098 trials for the Stroop task. The distribution of reaction times was also examined, and trials that took unreasonable

short or long were eliminated (characterized as half a second or less, and more than 3.5 – 5.5 seconds for ST and SS, respectively, as these were tasks in which participants were instructed to respond as quickly as possible, and 25 seconds for DM, an un-speeded task). This resulted in exclusion of 3%, 2.4%, and 2.11% of the trials for SS, ST, and DM, respectively.

Symbol Search SS is a measure of processing speed. Participants were presented with two rows of two cards. All cards contained a pair of symbols. Participants were instructed to select the card on the bottom of the screen that was identical to the card on the top of the screen, as quickly as possible. On non-lure trials, the inaccurate choice did not contain any symbol presented in the top row. On lure trials, one of the symbols on the wrong response-option card was also present up top. These trials were meant to be more difficult, as additional processing is required to rule out amongst the selection. The stimuli remained on screen until a response was made.

The dependent variable was mean response time in milliseconds for lure trials, unless otherwise specified (where it would be the score across all trial types). SS consisted of 12 trials per session and was completed across all surveys within a day (four sessions per day).

Dot-Memory DM is a measure of working memory. Each trial consisted of three phases. First, participants were presented with a 5x5 grid containing three dots and asked to remember their location. After a 3-second study period, participants were presented with a distractor task for 8 seconds, consisting of an array of Fs and Es. They were instructed to touch all the Fs. Next, an empty 5x5 grid appeared, and participants were to recall the location of the previously encoded three dots, and press “Done” once complete. They completed two trials of this task per session, each presenting the dots in different locations.

The dependent variable was an error score, with partial credit given based on the deviation from the correct position. If all dots were recalled correctly, they received a score of 0. In the case of errors, Euclidean distance of the location of the incorrect dot or dots to the correct grid location was calculated, with higher scores indicating less accurate placement. DM consisted of two trials per session and was completed across all surveys within a day (four sessions per day).

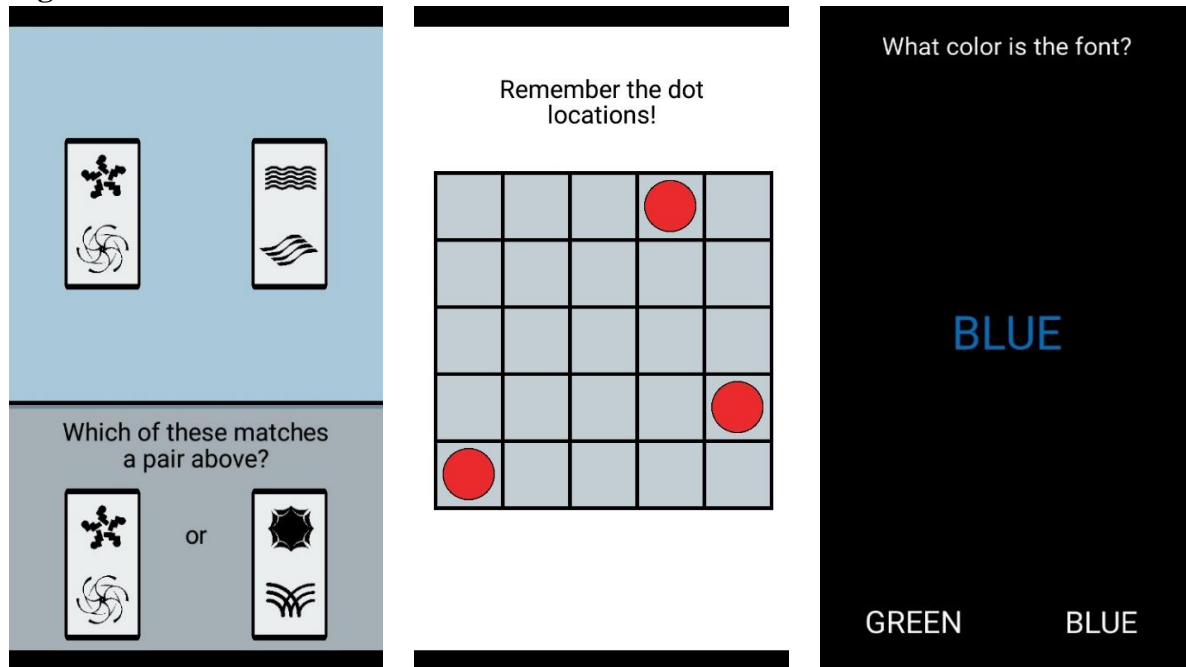
Stroop Task ST is a measure of response inhibition, an element of executive functioning. On each trial, participants were presented with a printed word, written in either an incongruent or congruent ink colour. They were asked to indicate *the ink colour* by touching one of two possible answers at the bottom of the screen. On congruent trials, the stimulus was matched, such that the ink colour and the word were harmonious. On incongruent conditions, the stimulus was mismatched (e.g., the word purple, in red font), and one of the response options was the colour of the ink (red, the correct response) and the other was the printed word (purple, the incorrect response). These trials were meant to be more difficult, as additional processing is required to inhibit the more automatic response (to respond with the word that was read). The stimuli remained on screen until a response was made.

The dependent variable was mean response time in milliseconds of incongruent trials, unless otherwise specified (where it would be the score across all trial types). ST consisted of 12 trials per session, and was completed only in the morning and evening survey (two sessions per day)

Effort Participants indicated the degree of effort (0 to 100, with 100 being maximum effort) they put into the brain games at the end of each survey. Effort is expected to account for

variability in cognitive outcomes and was controlled for. Effort was centered at the mean, so that 0 represents a score equivalent to the study's mean level of reported effort.

Figure 1 Brain Games used in DASH



Note. SS (non-lure trial), DM, and ST (congruent trial), respectively.

Analytical Approach

All data analyses were conducted in RStudio 2022.02.1+461. A two-level mixed multi-level model (MLM) was estimated to determine *intraclass correlations (ICC) and BP reliability*. Model 0 displays an unconditional model, where β_{0i} and e_{ij} reflect the intercept for person i and the variability about their own scores, respectively. The Level-1 model is nested within a second level, such that the distributions of these person specific parameters are modeled around population averages. Therefore, the population parameters (*fixed-effects*) inform population-level trends in, for example, the intercept (γ_{00}), while idiosyncratic deviations from these parameters (U_{0i}) indicate whether there is significant individual variation from population averages (*random effect*).

$$\begin{array}{ll}
 \text{Level 1:} & \text{Cog}_{ti} = \beta_{0i} + e_{ij} \\
 \text{Level 2:} & \beta_{0i} = \gamma_{00} + U_{0i}
 \end{array}
 \qquad \text{Model 0}$$

To calculate BP reliability based on Model 0, the following formula used by Sliwinski et al. (2018) was employed:

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

The intraclass correlation (ICC) is the expected correlation between two randomly sampled measurements from the same person and reflects the stability of measurements. To determine the ICC based on a single measurement, n would be equal to 1. To determine the reliability of aggregate scores based on the average of multiple measurements, n would be equal to the number of observations on which the average is based. Therefore, n is equal to the mean number of sessions completed for each cognitive test across the study.

A three-level mixed MLM was estimated to determine ***WP reliabilities across sessions and days***. WP reliability reflects systematic WP variation in cognitive performance across assessments (i.e., fluctuations that can be accounted for systematically, rather than error). This was done in two ways; To examine session-to-session reliability, trial (level 1) was nested within *session (level 2)*, which was then nested within person (level 3). To examine day-to-day reliability, trial (level 1) was nested within *day (level 2)*, which was then nested within person (level 3). Here, $U_{0si/di}$ represents trial-specific deviations from the persons predicted outcome, and γ_{000} is the grand mean, with V_{00i} being person-specific deviations from it.

$$\begin{array}{ll}
 \text{Level 1:} & \text{Cog}_{tsi/tdi} = \beta_{0si/di} + e_{tsi/tdi} \\
 \text{Level 2:} & \beta_{0si} = \delta_{00i} + U_{0si/di}
 \end{array}$$

Level 3: $\delta_{00i} = \gamma_{000} + V_{00i}$

Model 1

Similarly, to calculate WP reliability, the following formula used by Sliwinski et al. was employed:

$$WP\ Reliability = \frac{Var(WP\ occasion)}{\left(Var(WP\ occasion) + \frac{Var(WP\ error)}{i}\right)}$$

For **session-to-session reliability**, $Var(WP\ occasion)$ refers to session to session variability, which is considered systematic. Trial to trial variability within sessions is considered error. i denotes the number of trials the session aggregate was based on. Similarly, for **day-to-day reliability**, $Var(WP\ occasion)$ refers to systematic variability across days, and $Var(WP\ error)$ is trial level variability within a day. i denotes the number of trials the daily aggregate was based on. For this analysis, total scores (rather than just SS lure and ST incongruent scores) were used as the outcome, as the number of lure and incongruent trials varied within a session for SS and ST, respectively, so a mean i would not accurately represent each session or day if only these trials were used.

To examine **practice effects** two-level MLM's were fit to the data. Generally, in MLM, a unique trajectory for each participant in the study is determined at the first level of analysis (See Model Time). Therefore, time (and time-varying predictors) are specified at the first level. Building on Model 0, β_{1i} in the conditional model represents the rate of change for person i , γ_{10} represents population-level trends in the slope, and U_{1i} represents individual variation from it. The level-2 slope variance (U_{1i}) can be constrained so that all participants are assumed to be represented by the fixed effects, or specified as freely estimated, so that a unique trajectory for each individual is computed.

$$\text{Level 1: } \text{Cog}_{ti} = \beta_{0i} + \beta_{1i}(\text{Time}_{ti}) + e_{ti}$$

$$\text{Level 2: } \beta_{0i} (\text{intercept}) = \gamma_{00} + U_{0i}$$

$$\beta_{1i} (\text{slope}) = \gamma_{10} + U_{1i}$$

Time Model

Models with time defined as sessions across the study and days across the study were fit to elucidate practice effects for these various timescales, with linear and quadratic predictors for both. Demographic factors were added as predictors of the level 2 intercept to account for the affects of age, education, and sex on person intercepts. Effort for each survey will be controlled for at the first level of analyses (as it varied from session to session, and day to day)

$$\text{Level 1: } \text{Cog}_{ti} = \beta_{0i} + \beta_{1i}(\text{Time}_{ti}) + \beta_{2i}(\text{Time}^2_{ti}) + \beta_{3i}(\text{effort}_{ti}) + e_{ti}$$

$$\text{Level 2: } \beta_{0i} = \gamma_{00} + \gamma_{01}(\text{AgeBLi}) + \gamma_{02}(\text{SexMi}) + \gamma_{03}(\text{Edui}) + U_{0i}$$

$$\beta_{1i} = \gamma_{10} + U_{1i}$$

$$\beta_{2i} = \gamma_{20}$$

$$\beta_{3i} = \gamma_{30}$$

Predictor Model

Results

Demographic and Study Characteristics

Twenty-six participants who were contacted for the telephone-screening interview were not eligible to participate. Most exclusions ($N=21$) were due to serious medical concerns (e.g., having undergone recent surgery or chemotherapy treatment, recent stroke, substantial hearing loss, etc.) and five participants were already participating in a study using the same battery of cognitive measures. Two individuals dropped out due to COVID-19 concerns, and four additional participants dropped out for other reasons. In all, 66 community-residing older adults ($M_{\text{age}}=70.75$ years) from Victoria, B.C. and the surrounding communities participated. However,

the COVID-19 pandemic halted research that required in-person sessions, and we were forced to cancel intake sessions with the remaining participants.

The sample had proportionately more females (71%). Eighty-five percent of participants were Caucasian, 6% were Asian, and 9% left this question blank. The sample was highly educated with 40% of participants obtaining a graduate, law, or medical degree, and 30% obtaining an undergraduate degree or finishing trade school. The remainder (with the exception of three participants who did not specify) had completed some college/university, high school, or neither. These characteristics are largely representative of the recruitment population. Since participants completed 14 days of surveys (four/day), 56 surveys would be completed if compliance was perfect. On average, participants completed 49.79 sessions ($SD = 12.56$, 89% compliance). Some individuals completed sessions beyond the 14 days in the window between study completion and debriefing (range 1 – 64 surveys).

Table 1.3 Baseline Characteristics of Participants (N=66)

Variable	M (SD)	N (%)	Range
Age	70.75 (3.45)		64 – 78
Education			
Did not complete high school		1 (1.6)	
High school		4 (6)	
Trade school		4 (6)	
Some college/university		12 (18.2)	
Undergraduate degree		16 (24.2)	
Graduate, law or medical school		25 (37.9)	
Sex			
Female		47 (71.2)	
Male		16 (24.4)	
Ethnicity			
Caucasian		56 (84.9)	
Asian		4 (6)	
NA		6 (9)	
Cognitive functioning			
SS RT (Lure)	2.38 (.68)		.85 – 5.00
DM accuracy	8.51 (7.06)		.00 – 37.44
ST RT (Incongruent)	1.60 (.36)		.88 – 3.20
Effort	86.38 (18.22)		8 – 100

Note. M = Sample Mean; SD = Standard deviation; RT=reaction time converted to seconds. Study means of cognitive tasks are derived from the mean of person average scores across all sessions.

ICC and BP Reliabilities

The ICC is the correlation between two observations within the same individual. ICC values and BP reliabilities are reported in Table 1.2. ICC's ranged from 34% for DM, 42% for SS, and 59% for ST. In other words, 66% of the variance in DM trial scores, 58% of the variance in SS trial scores, and 41% of the variance in ST trial scores is WP. When aggregate scores are

used (i.e., the average score across study sessions for each person), BP reliability is enhanced to 96% for DM, and 97% for ST and SS. That is, the aggregate scores show minimal variability WP, and high variability between-people.

Table 1.4 ICC's and Between-Person Reliabilities for all Cognitive Measures

	DM	ST	SS
BP variance	16.92	75798	196860
Residual	33.03	53693	270631
ICC	.34	.59	.42
BP Reliability	.96	.97	.97

Note. i is the number of sessions completed on average for each cognitive test across the study. $i=50$ for DM and SS; $i=26$ for ST.

WP Reliabilities

WP reliability is the proportion of systematic WP variance that transpired across sessions (hours apart in this study) relative to WP variability that transpired within sessions (trial-level variability, which is considered random error) (Sliwinski et al., 2018). In other words, it refers to variance in performance likely attributable to systematic, time-varying factors, as opposed to randomness. WP reliabilities for each cognitive task are reported in Table 1.3. From session-to-session, WP-reliability ranged from .10 for DM, .65 for ST, and .56 for SS. From day-to-day, values ranged from .24, .75, and .77 for DM, ST, and SS, respectively. These findings show that DM was less able to reliably capture systematic WP fluctuations in performance compared to the other two tasks. The two reaction time tasks (ST and SS) showed relatively higher systematic variation (.56-.65 at the session level, and .75-.77 at the daily level) in scores, meaning that these measures may better capture systematic WP fluctuations. Figure 1.1 and 1.2 depicts individual variation in scores for each cognitive test across session and across day, respectively, for a subset of randomly selected participants.

Table 1.5 Within-Person Reliabilities Across Sessions and Days for all Cognitive Measures

	DM		ST		SS	
	Session	Day	Session	Day	Session	Day
WP Reliability	.10	.24	.65	.75	.56	.77

Note. Outcome for all trials of ST and SS, not just of the incongruent and lure trials. i = the number of observations for each time scale (number of trials within a session, and number of trials within a day). $i = 2$ trials/session and 8 trials/day for DM; $i = 12$ trials/session and 24 trials/day for Stroop; $i = 12$ trials/session and 48 trials/day for Stroop.

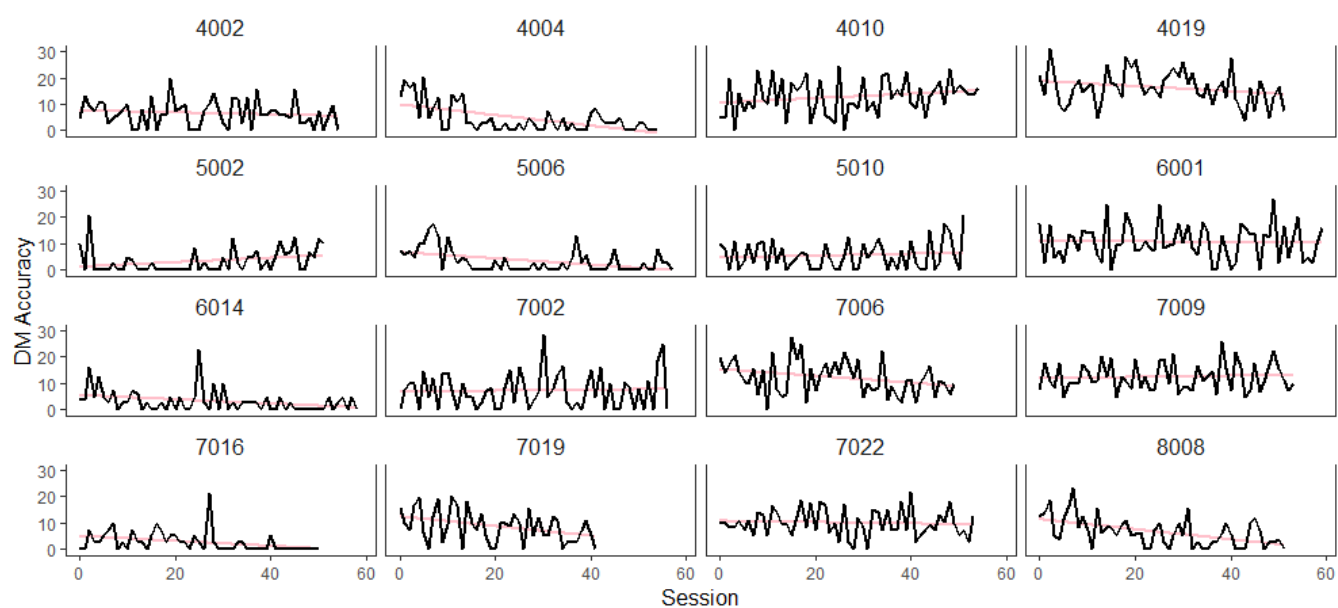
Figure 1.1a Dot Memory Accuracy Score Trajectories for a Random Subset of Participant Across Sessions

Figure 1.1b Stroop Reaction Time Trajectories for a Random Subset of Participant Across Sessions

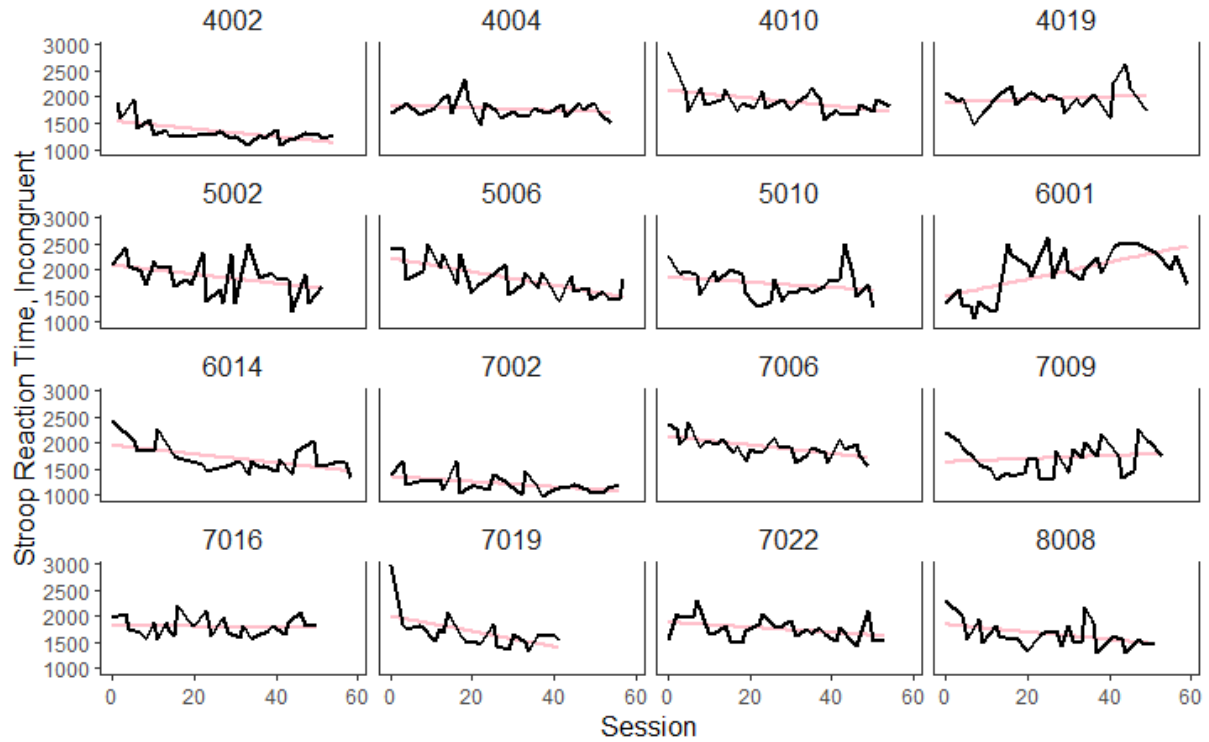


Figure 1.1c Symbol Search Reaction Time Trajectories for a Random Subset of Participant Across Sessions

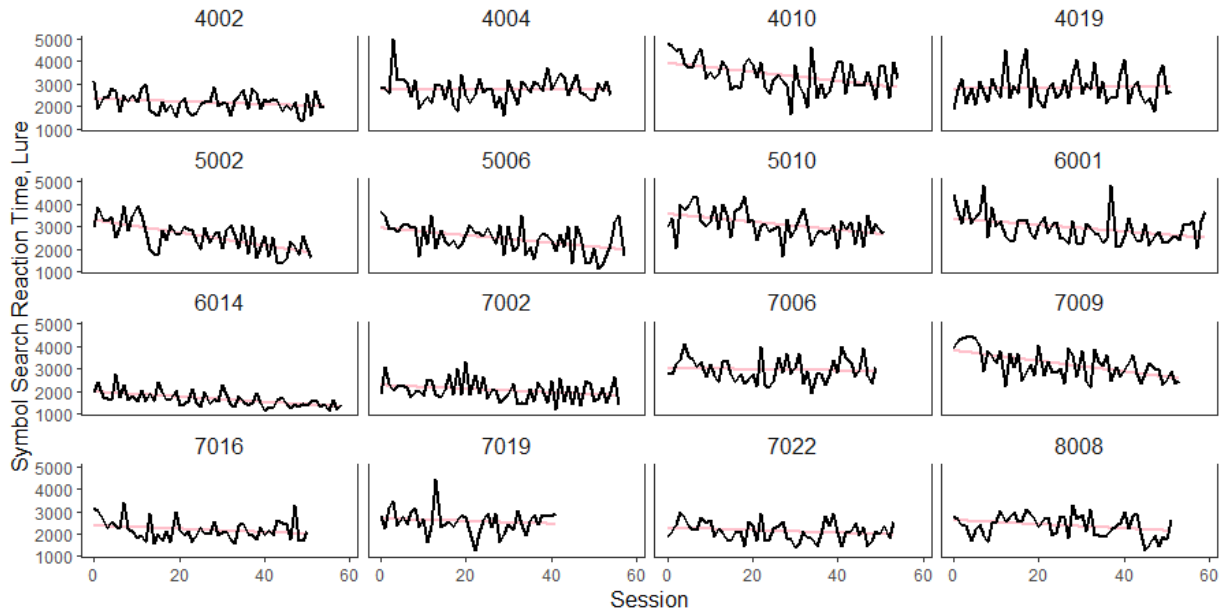


Figure 1.2a Dot Memory Accuracy Score Trajectories for a Random Subset of Participant Across Days

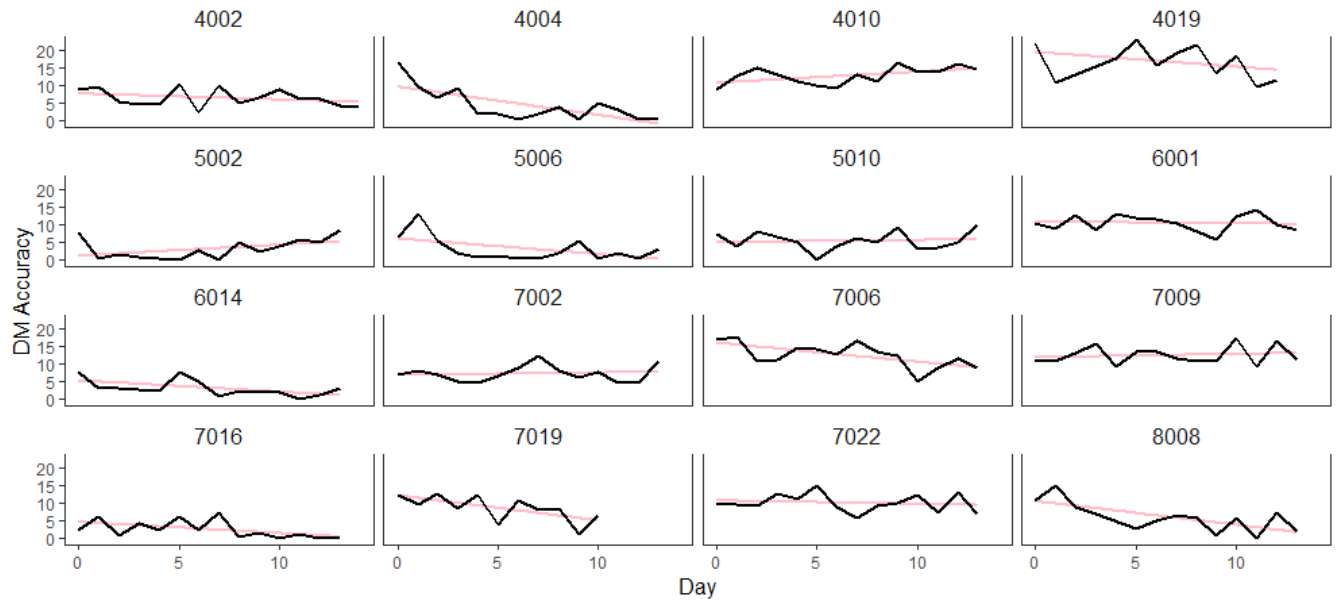


Figure 1.2b Stroop Reaction Time Trajectories for a Random Subset of Participant Across Days

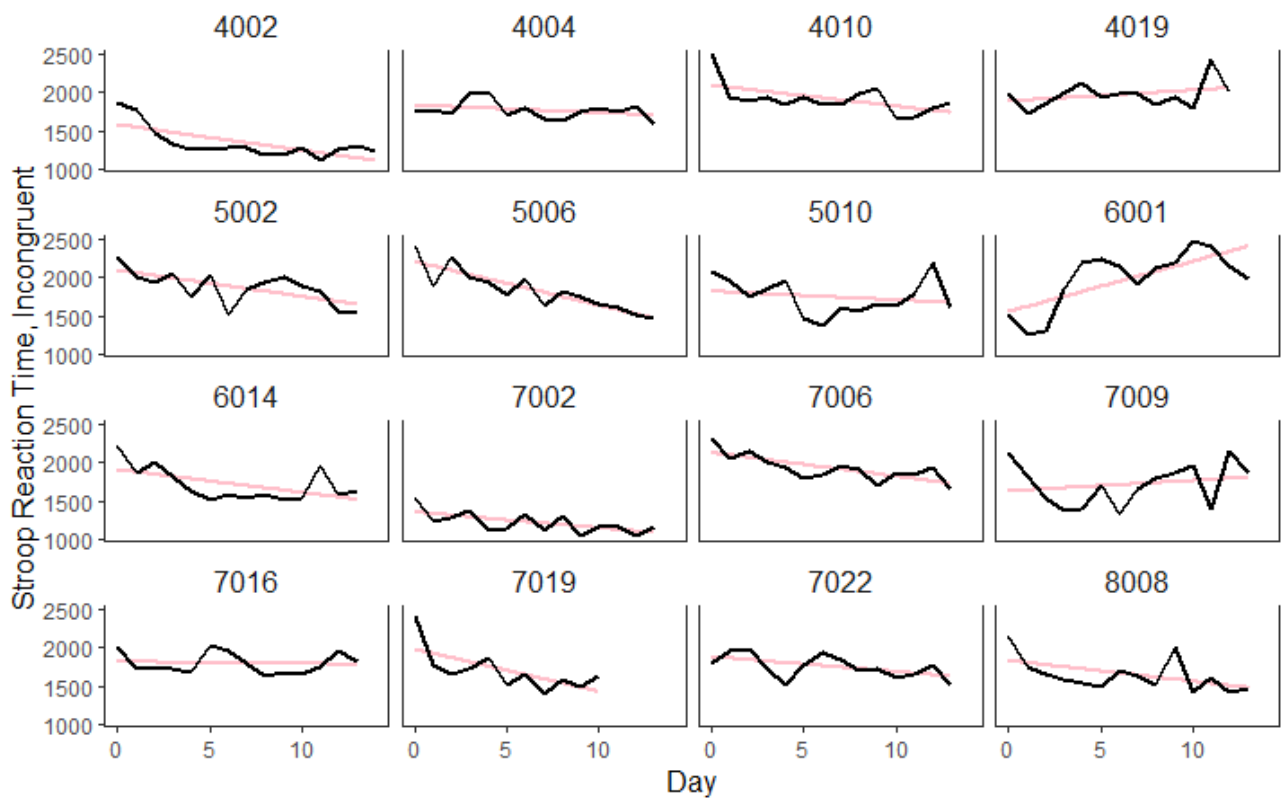
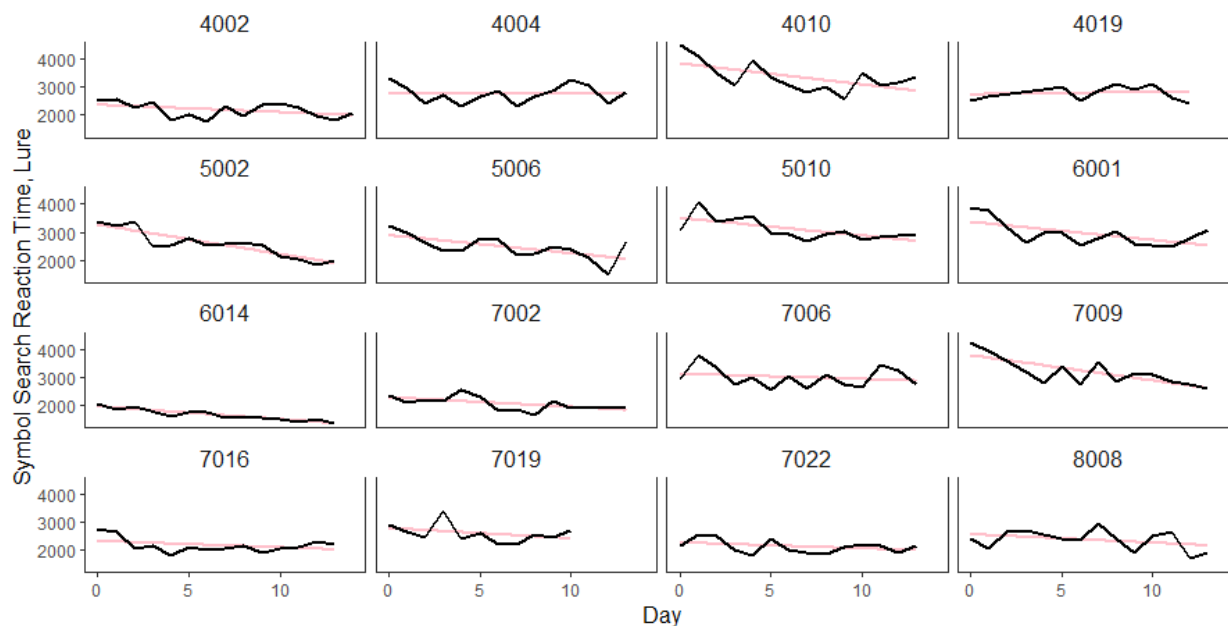


Figure 1.2c Symbol Search Reaction Time Trajectories for a Random Subset of Participant Across Days



Practice Effects

Across Sessions Models with and without a unique slope (trajectory) for each individual were compared for each cognitive outcome. An ANOVA comparison indicated that specifying a random slope for each individual improved the model fit across measures, with lower AIC and BIC values. Thus, the random slope model was used. Including a quadratic term for time (session) also improved fit across models. Table 1.4 presents the results of the final MLM model for each cognitive task. Both linear and quadratic time were significantly related to cognitive performance on all tasks, across sessions. Figure 1.3 depicts cognitive scores across sessions of the study for each individual, with a line of best fit.

No other predictors (effort, age, sex, and education) were found to be significant, except for the effects of education on DM performance, with higher education improving accuracy (i.e., lowering the score, indicating more accurate placement). Effort and age approached significance for DM, with more effort improving accuracy, and older individuals performing less accurately.

Table 1.6 Multilevel Estimates of the Effects of Session, Effort, and Demographic Factors on Cognitive Tasks

Variable	Dot Memory Estimate (SE)	Stroop Estimate (SE)	Symbol Search Estimate (SE)
Fixed Effects			
WP variables			
Intercept (γ_{00})	12.85 (.87)***	1765.40 (64.24)***	2796.55 (106.20)***
Linear time (γ_{10})	-2.56 (.51)***	-222.37 (28.31) ***	-405.85 (46.16)***
Quadratic time (γ_{20})	0.42 (.18)*	51.06 (9.41) ***	88.87 (15.87)***
Effort (γ_{30})	-.02 (.01)	1.14 (.75)	.09 (1.19)
BP Variables			
Age (γ_{01})	0.24 (.14)	-0.89 (10.34)	6.03 (16.17)
Sex (γ_{02})	-1.83 (1.12)	-13.52 (83.80)	51.39 (131.81)
Edu (γ_{03})	-1.37 (.60)*	8.44 (45.07)	-94.98 (70.50)
	Variance (SD)	Variance (SD)	Variance (SD)
Random effects			
WP (σ_e^2)	30.64 (5.54)	44174 (210.18)	242987 (492.9)
Between-person			
Intercept (σ_0^2)	13.84 (3.72)	81752 (285.92)	267684 (517.4)
Linear time (σ_1^2)	1.44 (1.20)	9030 (95.03)	16635 (129.0)

Note. Results are based on 3124, 1659, and 3153 assessments for DM, ST, and SS respectively. Lower values indicate better performance on all cognitive tasks. * $p < .05$; ** $p < .01$; *** $p < .001$

Figure 1.3a Dot Memory Accuracy Scores Across Sessions Within the Study for Individual Participants with a Line of Best Fit.

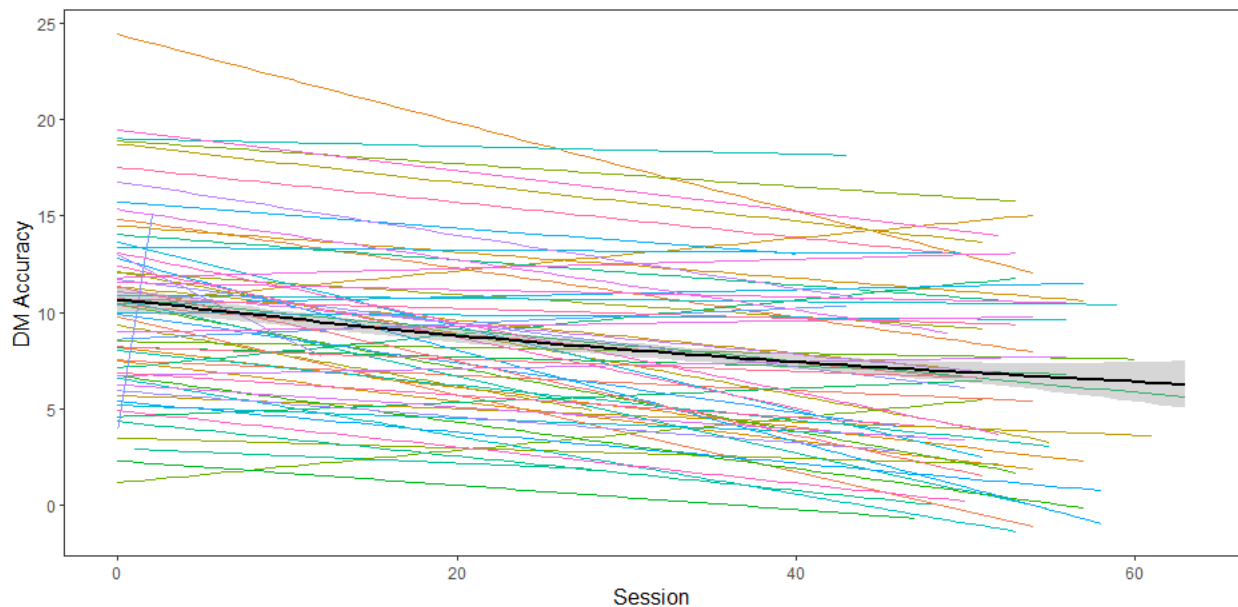


Figure 1.3b Stroop Reaction Time Scores Across Sessions Within the Study for Individual Participants with a Line of Best Fit.

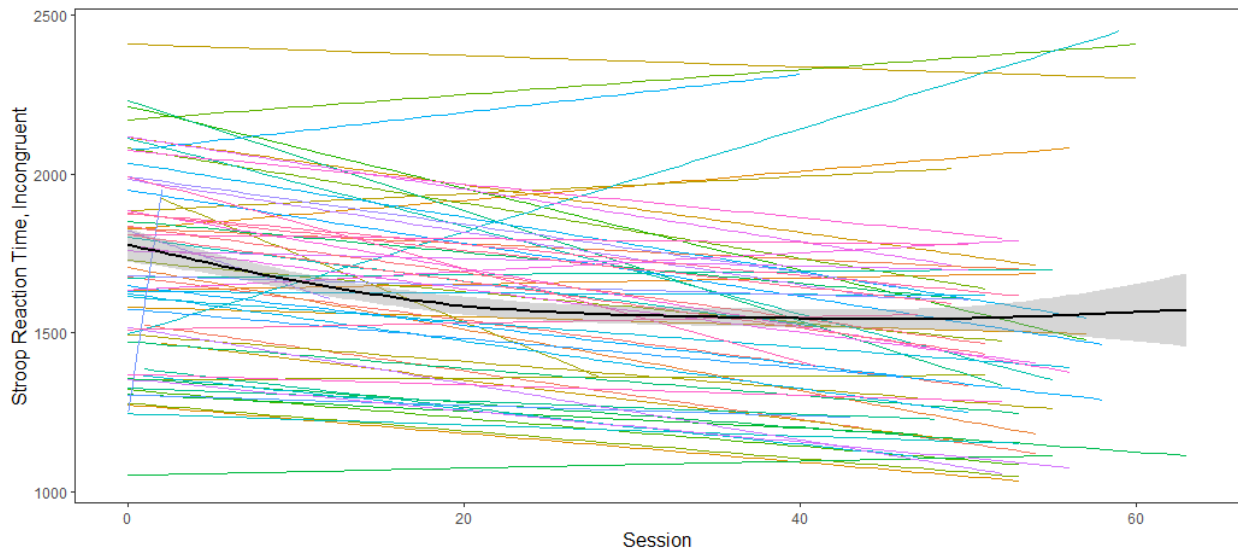
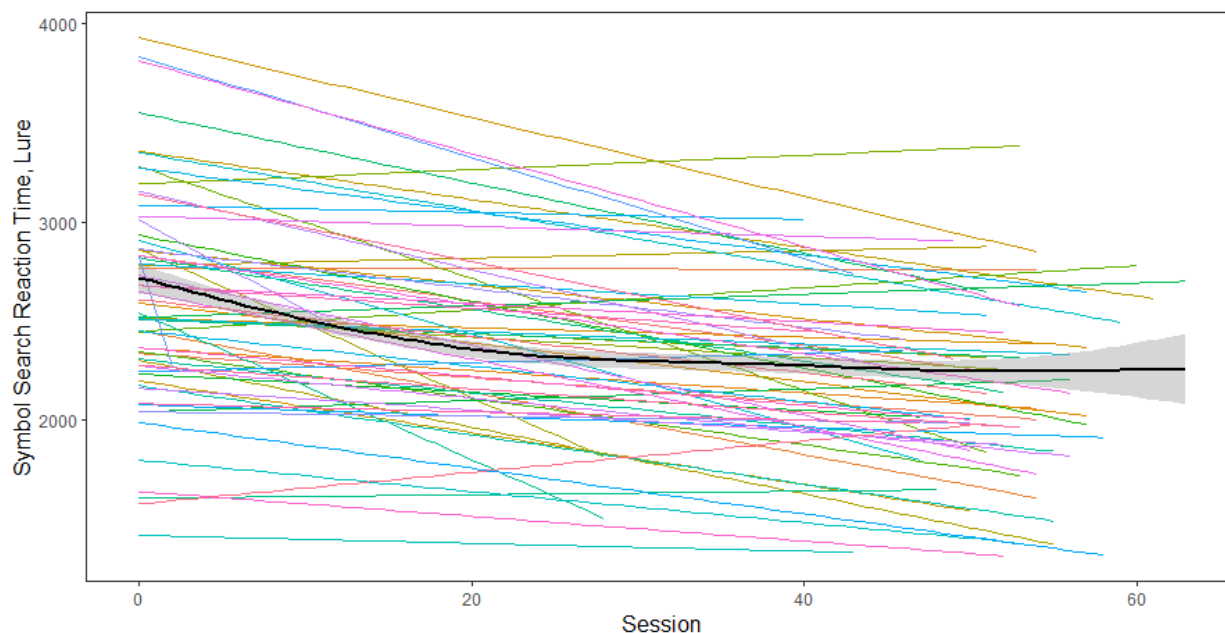


Figure 1.3c Symbol Search Reaction Time Scores Across Sessions Within the Study for Individual Participants with a Line of Best Fit.



Across Days Models with and without a unique slope (trajectory) for each individual were compared for each cognitive outcome. An ANOVA comparison indicated that specifying a random slope for each individual improved the model fit across measures, with lower AIC and BIC values. Including a quadratic term for time (day) also improved fit across models. Table 1.4 presents the results of the final MLM model for each cognitive task. Both linear and quadratic time were significantly related to cognitive performance on all tasks, across days. Figure 1.4 depicts cognitive scores across days of the study for each individual, with a line of best fit.

The effects of effort on ST performance were significant, with more effort resulting in a slower score. Age approached significance for DM accuracy, suggesting younger individuals performing more accurately. Lastly, higher educational attainment resulted in more accurate scores on DM. No other predictors (effort, age, sex, and education) were found to be significant.

Table 1.7 Multilevel Estimates of the Effects of Day, Effort, and Demographic Factors on Cognitive Tasks

	Dot Memory	Stroop	Symbol Search
Variable	Estimate (SE)	Estimate (SE)	Estimate (SE)
Fixed Effects			
WP variables			
Intercept (γ_{00})	12.74 (.88)***	1774.83 (65.12)***	2780.01 (105.48)***
Linear time (γ_{10})	-4.57 (.98)***	-440.28 (55.75) ***	-755.05 (98.65)***
Quadratic time (γ_{20})	1.40 (0.69)*	205.38 (37.46) ***	332.53 (69.98)***
Effort (γ_{30})	-.01 (0.02)	2.59 (1.18)*	1.74 (2.04)
BP Variables			
Age (γ_{01})	0.24 (.14)	-0.81 (10.49)	7.33 (16.02)
Sex (γ_{02})	-1.89 (1.13)	-37.03 (85.04)	45.68 (130.40)
Edu (γ_{03})	-1.43 (.61)*	-0.18 (45.81)	-101.64 (70.35)
	Variance (SD)	Variance (SD)	Variance (SD)
Random effects			
WP (σ_e^2)	8.45 (2.91)	24777 (157.4)	86777 (294.6)
Between-person			
Intercept (σ_0^2)	13.545 (3.68)	81359 (285.2)	246522 (496.5)
Linear time (σ_1^2)	4.53 (2.13)	33701 (183.6)	48140 (219.4)

Note. based on 810, 812, and 815 observations for DM, ST, and SS respectively. Lower values indicate better performance on all cognitive tasks. * $p < .05$; ** $p < .01$; *** $p < .001$

Figure 1.4a Dot Memory Accuracy Scores Across Days Within the Study for Individual Participants with a Line of Best Fit.

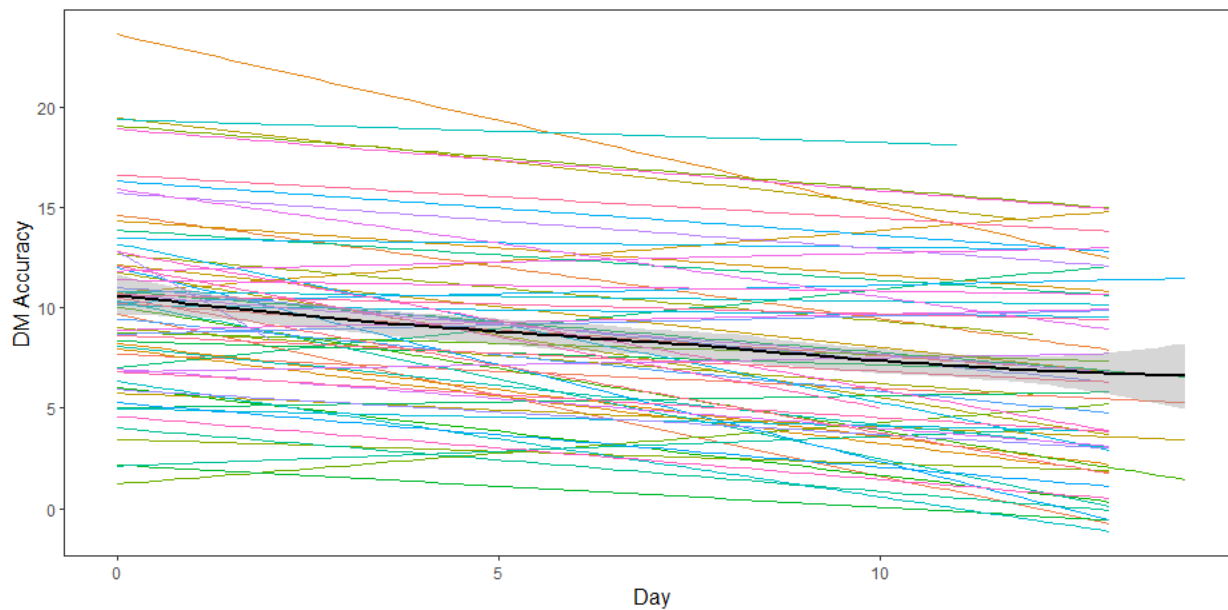


Figure 1.4b Stroop Reaction Time Scores Across Days Within the Study for Individual Participants with a Line of Best Fit.

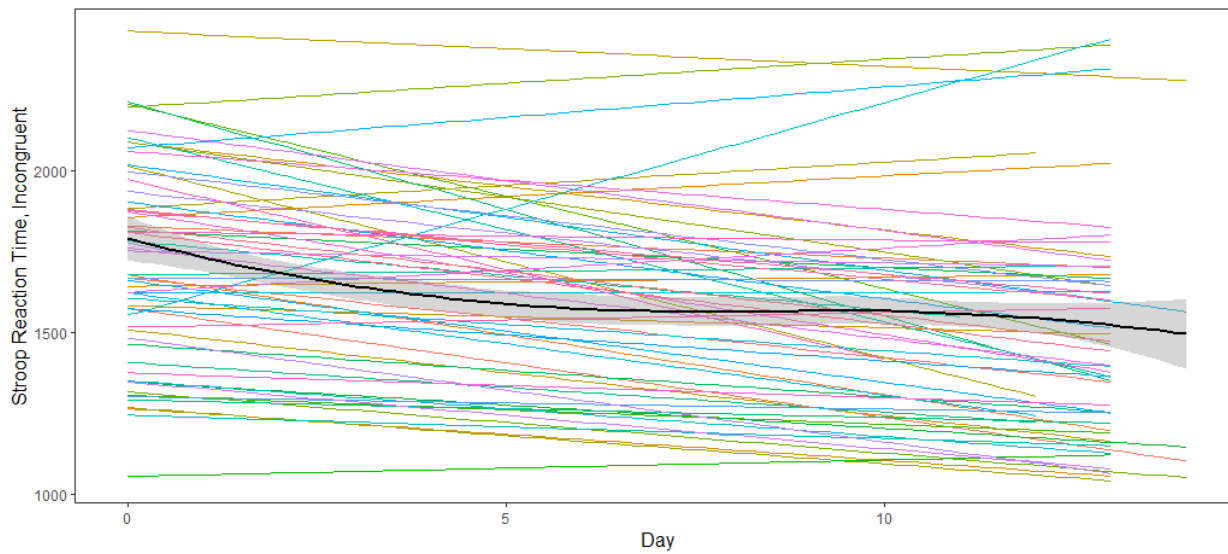
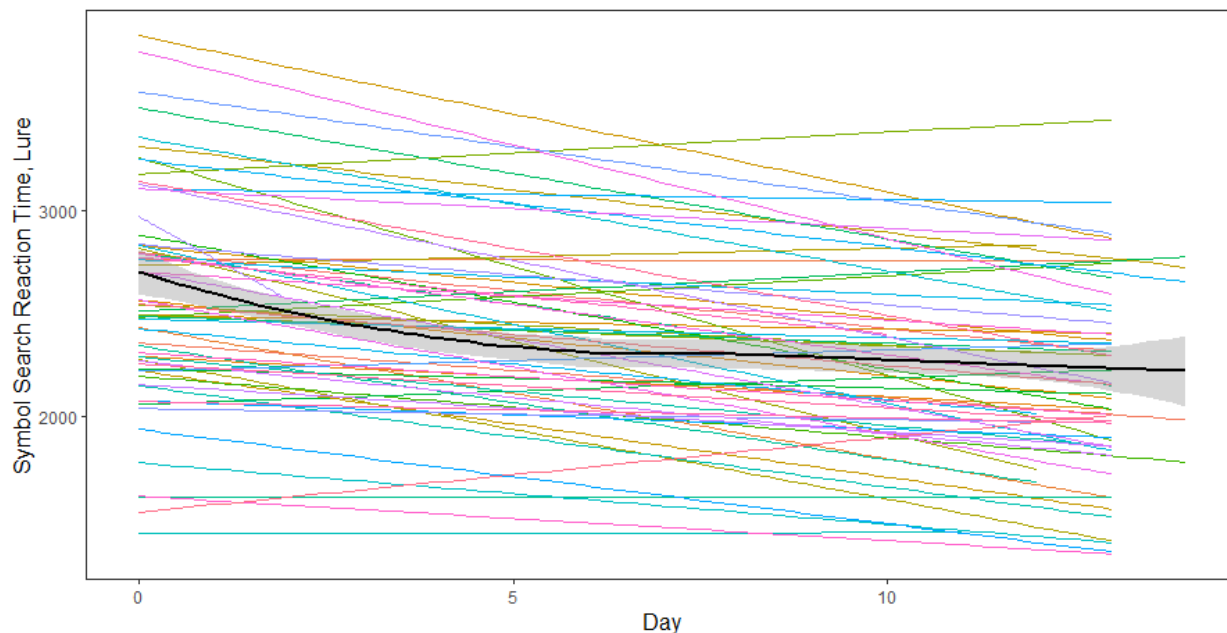


Figure 1.4c Symbol Search Reaction Time Scores Across Days Within the Study for Individual Participants with a Line of Best Fit.



Discussion

This project replicated and expanded on a growing body of research using a battery of mobile cognitive tests in a healthy older adult population. Specifically, previous analyses of BP and WP reliability were replicated, with a novel examination of WP reliability and practice effects across multiple time scales. Two cognitive tasks in this study (DM and SS) have been previously researched, while the present study also included ST. Measures of visual working memory, processing speed, and executive functioning, respectively, were included.

BP reliabilities using the aggregate score were found to be solid in this study, with reliabilities at or above .96 across tasks. These findings are comparable to previous results, where BP reliabilities are consistently above .90 across tasks, with the exception of DM in Brewster et al.'s (2021) study, with BP *burst* reliabilities between .79 to .80. Differences here are likely due to differences in testing protocol, where two assessments per day, for 5 days were completed in Brewster et al.'s longitudinal burst study, compared to 3-5 assessments completed

per day, from anywhere between 8 to 14 days in the other studies that did not utilize longitudinal burst follow-up (i.e., resulting in more assessments aggregated in the latter).

ICC's are a type of BP reliability reflecting the stability of measurements (i.e., the expected correlation between two random scores from the same individual). ICC's from the present study indicate that DM (an accuracy-based task) is comprised of lower BP variance relative to WP (ICC for DM was .34, so 34% of the variance in performance across this task was BP). ICC's for the reaction time tasks were higher, suggesting relatively more variation in performance at the BP level compared to DM, with an almost equal split in variance across the BP and WP level (ICC's of .59 and .42 for ST and SS, respectively). These findings are comparable to Brewster et al. (2021), Sliwinski et al. (2018), and Thompson et al. (2022) who report ICCs of .39 and .46 for DM, and .54 and .71 for SS. This pattern of lower DM ICCs relative to ICCs on speeded tasks suggests that people seem to be performing similarly (and accurately) on the accuracy-based DM task, while performance on the two reaction time tasks may better discriminate between individuals. However, it is important to keep in mind that the ICC is a relative index, meaning that it indicates the proportion of BP variability relative to WP variability; as such, it does not reflect the magnitude of variability, just where the variability exists. Nicosia et al. (2022) found higher ICC values for Grid (a task similar to DM) and SS, above .85 across bursts. Relatively higher ICC's for the Grid task compared to DM task may be due to a more intensive measurement schedule (i.e., relatively greater assessment occasions within a burst compared to Brewster et al.), or differences in the stimuli used on the Grid task compared to DM task.

Similarly, WP reliabilities for DM were low across both time scales (.10 and .24 for session and day, respectively), suggesting that DM captures relatively less systematic WP

fluctuations (as there seems to be relatively minimal variability in session scores). On the other hand, WP reliability for the two reaction time tasks (.56 to .65 across sessions, and .75 to .77 across days) were modest across sessions and higher across days, meaning that these scores more reliably captured systematic fluctuations (see Appendix 1.1 for a histograms of test score count). The session level WP reliabilities in Sliwinski et al. (2018) were .53 for SS (comparable to this study), and .50 for DM (much higher than in this study). However, the findings from this analysis are consistent with past research, showing WP reliabilities in the range from .50-.70 for response time tasks, and .15 to .36 for accuracy-based measures (as reported in Sliwinski et al., 2018). Taken together, this suggest that an accuracy-based outcome measure, such as DM, may show less BP and WP systematic differences in performance, since individuals are mostly scoring similarly, while speeded measures might be more useful in capturing performance variation, across both BP and WP levels. Alternatively, it may be the case that the lower number of trials in DM (two trials per session, since it takes relatively longer) resulted in a less robust aggregate score compared to the reaction time tests comprised of more trials per session (12). As Brewster et al. (2021) astutely points out, higher user satisfaction with the study protocols to date may permit longer assessment intervals with more DM trials in future studies.

Across tasks, WP reliability was higher when day was used as the time-metric as opposed to session, meaning that daily aggregate scores may be more sensitive to non-random WP fluctuations. However, this finding is not necessarily generalizable, as the expected temporal association likely differs depending on which time-varying factors are at play (Brewster et al.; Sliwinski et al.). That is, a different timescale may be more useful for examining one time-varying factor compared to another, such as capturing the effects of alcohol use on cognition (i.e., which likely has effects within the hour), versus the effects of sleep on cognition (i.e.,

which likely has pronounced effects at the daily level). As Sliwinski et al. (2018) describes, reliabilities in this context are not necessarily a property of the tests per se. Rather, they depend on the specific assessment procedure. This is noteworthy for two reasons; Firstly, the reliabilities reported here are specific to testing two to four times per day, for 14 days, and secondly, reliability should always be examined when different measurement cycles are used in future studies exploring time-varying factors of varying cadences. To date, there is limited research examining the relationship between lifestyle/time-varying factors and cognitive performance on mobile testing. Brewster et al. (2021) did not find any association between time-varying predictors, such as affect, distractibility, and pain and fluctuations in cognitive scores, though other studies have demonstrated within person coupling of daily time-varying factors and cognitive performance (e.g., Allard et al., 2014; Riediger et al., 2014; Thomson, Nimmo, Tiplady, & Glen, 2009; Tiplady, Oshinowo, Thomson, & Drummond, 2009). Further research on the sensitivity of mHealth tests to time-varying factors will be crucial.

Regarding practice effects, the two past longitudinal studies (Brewster et al. 2021 and Nicosia et al. 2022) both found improvement over the year for SS, while only Brewster et al. found linear improvement for DM over the year. The present study examined practice effects using both the study day and session as time metrics, with an analysis across days also conducted by Brewster et al. and Thompson et al. (2022), who both reported a significant improvement for SS. Brewster et al., included DM in their analysis, and did not find any effects across days within bursts. On the contrary, the present study found linear and quadratic terms for day to be significant across tasks, suggesting that individuals improved across days early in the study, with eventual leveling off in performance. This suggests diminished practice effects over time. The examination of practice effects from session to session was unique to this study, showing the

same trend. Early practice effects are largely supported across studies, underscoring the importance of future studies controlling for these effect. Sliwinski et al. (2018) did not observe practice effects (though these authors analyses of practice effects differed compared to the other studies), and interestingly, these authors did not include the initial data from a trial-run before study enrollment. Though this was meant to demonstrate compliance before entry into the study, it may also have acted as a control period to eliminate practice effects so that by the time participants entered the study, practice gains were minimal.

Demographic variables (age, education, and sex) were largely found to be nonsignificant in the models. The only exception to this was that higher education resulted in more accurate DM scores aggregated at the session level. This fits with previous research showing a relationship between short-term memory and education level among older adults (Brewster et al.; Pliatsikas et al., 2019). Effort predicted ST scores at the daily level, with higher effort resulting in slower ST scores. This finding is interesting and may mean that those who put in more effort were less impulsive in their responding, favouring accuracy over speed. Surprisingly, effort was not significantly related to any other cognitive task at the session or day level.

In addition to the limitations and caveats mentioned herein, it is worth noting that the sample of older adults completing assessments in the present study were highly educated, limiting generalizability of findings to the broader population of older adults. Moreover, selection bias may be a factor, as participants who responded to the advertisement for a mobile-based study may be more tech-savvy in comparison to aged matched peers in the general population. It is also worth noting that participants in this study were given compensation for their participation. This could have influenced adherence positively. Additionally, the study was conducted with a sample of healthy older adults. Thus, findings cannot speak to the utility of

mobile testing with clinical populations. Future studies with more diverse populations, as well as clinical populations will be critical.

References

- Adlimoghaddam, A., Roy, B., & Albeni, B. C. (2018). Future trends and the economic burden of dementia in Manitoba: comparison with the rest of Canada and the world. *Neuroepidemiology*, *51*, 71-81.
- Allard, M., Husky, M., Catheline, G., Pelletier, A., Dilharreguy, B., Amieva, H., ... & Swendsen, J. (2014). Mobile technologies in the early detection of cognitive decline. *PLoS One*, *9*(12), e112197.
- Brewster, P. W., Rush, J., Ozen, L., Vendittelli, R., & Hofer, S. M. (2021). Feasibility and psychometric integrity of mobile phone-based intensive measurement of cognition in older adults. *Experimental aging research*, *47*(4), 303-321.
- Brouillette, R. M., Foil, H., Fontenot, S., Correro, A., Allen, R., Martin, C. K., ... & Keller, J. N. (2013). Feasibility, reliability, and validity of a smartphone based application for the assessment of cognitive function in the elderly. *PloS one*, *8*(6), e65925.
- Cain, A. E., Depp, C. A., & Jeste, D. V. (2009). Ecological momentary assessment in aging research: a critical review. *Journal of psychiatric research*, *43*(11), 987-996.
- Claveau, J.S., Presse, N., Kergoat, M.J., Villalpando, J.M. (2018). The lost years: delay between the onset of cognitive symptoms and clinical assessment at a memory clinic. *Can Geriatr J*, *21*, 152–6.
- Government of Canada, Action for senior's report. (2014)

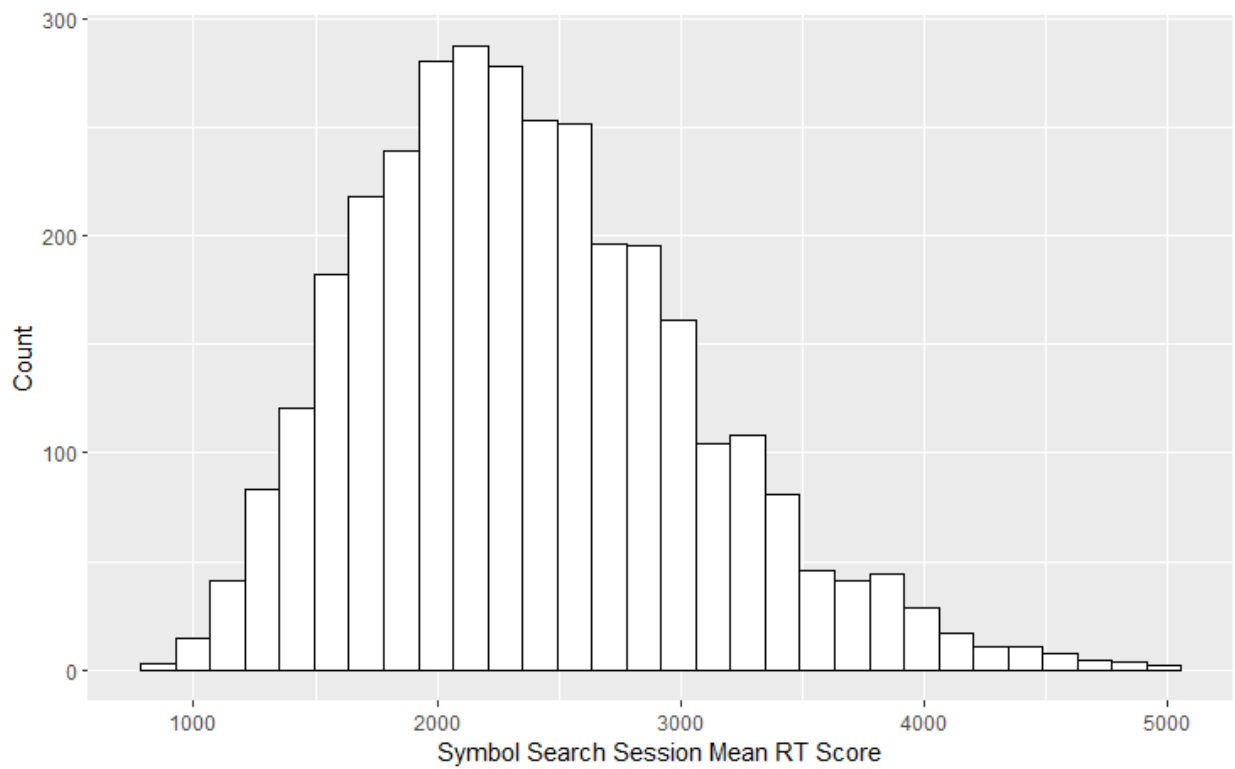
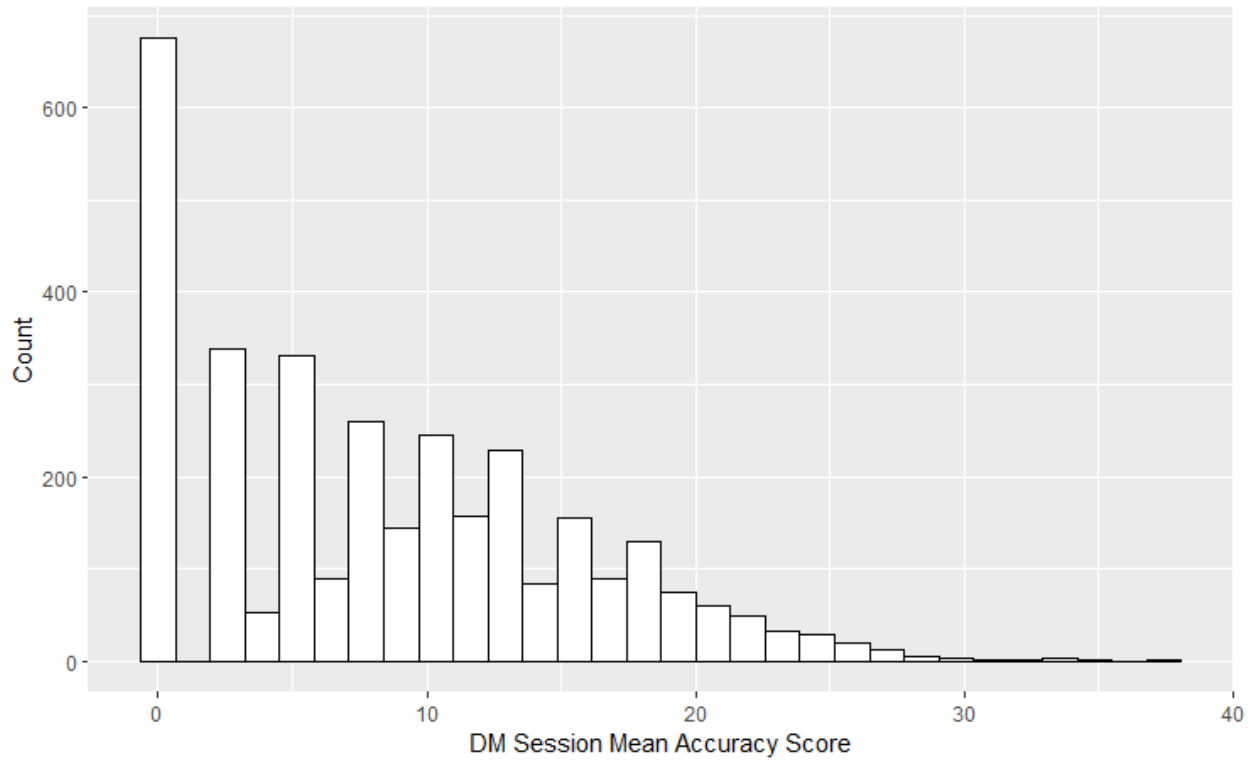
- Hassenstab, J., Aschenbrenner, A. J., Balota, D. A., McDade, E., Lim, Y. Y., Fagan, A. M., ... & Dominantly Inherited Alzheimer Network. (2020). Remote cognitive assessment approaches in the Dominantly Inherited Alzheimer Network (DIAN) Using digital technology to drive clinical innovation in brain-behavior relationships: a new era in neuropsychology. *Alzheimer's & Dementia*, *16*, e038144.
- Hassenstab, J., Chasse, R., Grabow, P., Benzinger, T. L., Fagan, A. M., Xiong, C., ... & Morris, J. C. (2016). Certified normal: Alzheimer's disease biomarkers and normative estimates of cognitive functioning. *Neurobiology of aging*, *43*, 23-33.
- Jongstra, S., Wijsman, L. W., Cachucho, R., Hoevenaer-Blom, M. P., Mooijaart, S. P., & Richard, E. (2017). Cognitive testing in people at increased risk of dementia using a smartphone app: the iVitality proof-of-principle study. *JMIR mHealth and uHealth*, *5*(5), e68.
- Karr, J. E., Graham, R. B., Hofer, S. M., & Muniz-Terrera, G. (2018). When does cognitive decline begin? A systematic review of preclinical change point studies on accelerated decline in cognitive and neurological outcomes preceding mild cognitive impairment, dementia, and death. *Psychology and Aging*, *33*(2), 195-218.
- Lancaster, C., Koychev, I., Blane, J., Chinner, A., Chatham, C., Taylor, K., & Hinds, C. (2020). Gallery Game: Smartphone-based assessment of long-term memory in adults at risk of Alzheimer's disease. *Journal of clinical and experimental neuropsychology*, *42*(4), 329-343.

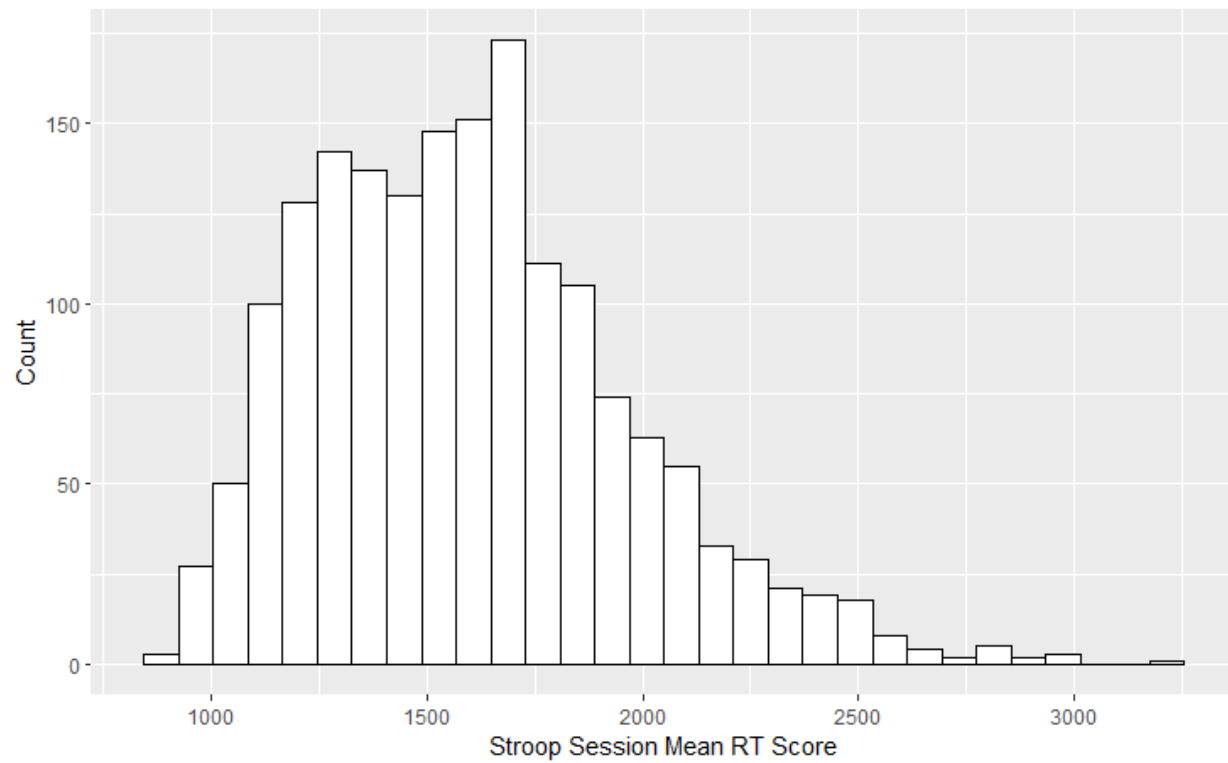
- Mackin, R. S., Insel, P. S., Truran, D., Finley, S., Flenniken, D., Nosheny, R., ... & Weiner, M. W. (2018). Unsupervised online neuropsychological test performance for individuals with mild cognitive impairment and dementia: results from the Brain Health Registry. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *10*, 573-582.
- Matar, E., Shine, J. M., Halliday, G. M., & Lewis, S. J. (2020). Cognitive fluctuations in Lewy body dementia: towards a pathophysiological framework. *Brain*, *143*(1), 31-46.
- 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.
- Nicosia, J., Aschenbrenner, A., Balota, D., Sliwinski, M., Tahan, M., Adams, S., ... & Hassenstab, J. (2022). Unsupervised high-frequency smartphone-based cognitive assessments are reliable, valid, and feasible in older adults at risk for Alzheimer disease [Unpublished Manuscript].
- Öhman, F., Hassenstab, J., Berron, D., Schöll, M., & Papp, K. V. (2021). Current advances in digital cognitive assessment for preclinical Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *13*(1), e12217
- Papp, K. V., Samaroo, A., Chou, H. C., Buckley, R., Schneider, O. R., Hsieh, S., ... & Amariglio, R. E. (2021). Unsupervised mobile cognitive testing for use in preclinical Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *13*(1), e12243.

- Pliatsikas, C., Veríssimo, J., Babcock, L., Pullman, M. Y., Gleib, D. A., Weinstein, M., ... & Ullman, M. T. (2019). Working memory in older adults declines with age, but is modulated by sex and education. *Quarterly Journal of Experimental Psychology*, *72*(6), 1308-1327.
- Rabin, L. A., Smart, C. M., & Amariglio, R. E. (2017). Subjective cognitive decline in preclinical Alzheimer's disease. *Annual review of clinical psychology*, *13*, 369-396.
- Riediger, M., Wrzus, C., Klipker, K., Müller, V., Schmiedek, F., & Wagner, G. G. (2014). Outside of the laboratory: Associations of working-memory performance with psychological and physiological arousal vary with age. *Psychology & Aging*, *29*(1), 103– 114.
- Salthouse, T. A., & Berish, D. E. (2005). Correlates of within-person (across-occasion) variability in reaction time. *Neuropsychology*, *19*(1), 77.
- Schweitzer, P., Husky, M., Allard, M., Amieva, H., Pérès, K., Foubert-Samier, A., ... & Swendsen, J. (2017). Feasibility and validity of mobile cognitive testing in the investigation of age-related cognitive decline. *International journal of methods in psychiatric research*, *26*(3), e1521.
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annu. Rev. Clin. Psychol.*, *4*, 1-32.
- Sliwinski, M., & Buschke, H. (2004). Modeling intraindividual cognitive change in aging adults: Results from the Einstein aging studies. *Aging Neuropsychology and Cognition*, *11*(2-3), 196-211.
- Sliwinski, M., Hoffman, L., & Hofer, S. (2010). Modeling retest and aging effects in a measurement burst design.

- 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.
- Thompson, L. I., Harrington, K. D., Roque, N., Strenger, J., Correia, S., Jones, R. N., ... & Sliwinski, M. J. (2022). A highly feasible, reliable, and fully remote protocol for mobile app-based cognitive assessment in cognitively healthy older adults. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *14*(1), e12283.
- Thomson, A. J., Nimmo, A. F., Tiplady, B., & Glen, J. B. (2009). Evaluation of a new method of assessing depth of sedation using two-choice visual reaction time testing on a mobile phone. *Anaesthesia*, *64*(1), 32– 38.
- 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.
- Wilks, H.M., Aschenbrenner, A.J., Gordon, B.A., Balota, D.A., Fagan, A.M., Musiek, E., BallsBerry, J., L.S. Benzinger, T.L.S., Cruchaga C., Morris, J.C., Hassenstab, J. (2021). Sharper in the Morning: Cognitive Time of Day Effects Revealed with High-Frequency Smartphone Testing. *Journal of Clinical and Experimental Neuropsychology*.
<https://doi.org/10.1080/13803395.2021.2009447>
- Wimo A, et al: The worldwide economic impact of dementia 2010. *Alzheimers Dement* 2013;*9*:1–11.e13.
- Wister, A. V., & Speechley, M. (2015). Inherent Tensions Between Population Aging and Health Care Systems: What Might the Canadian Health Care System Look Like in Twenty Years?. *Journal of Population Ageing*, *8*(4), 227-243.

- Yoneda, T., Lewis, N. A., Knight, J. E., Rush, J., Vendittelli, R., Kleineidam, L., ... & Muniz-Terrera, G. (2021). The importance of engaging in physical activity in older adulthood for transitions between cognitive status categories and death: A coordinated analysis of 14 longitudinal studies. *The Journals of Gerontology: Series A*, 76(9), 1661-1667.
- Zygouris, S., & Tsolaki, M. (2015). Computerized cognitive testing for older adults: a review. *American Journal of Alzheimer's Disease & Other Dementias*®, 30(1), 13-28.

Appendix 1.1 Count of Scores for each Cognitive Test



Chapter 2

The Association between Physical Activity and Cognitive Abilities using Ecological Momentary Assessment

Abstract

Objective: It has been well established that physical activity (PA), and moderate to vigorous PA (MVPA) in particular, enhances cognitive abilities both short- and long-term. The purpose of this study is to determine if, and to what extent, MVPA contributes to both BP (i.e., interindividual differences) and WP (i.e., intraindividual variation) variability in cognitive performance as measured by smartphone-based mobile health (mHealth) assessments within an EMA protocol. In addition, to evaluate differences between objective and subjective measures of activity on these effects.

Method: Healthy community-residing older adults ($n=66$; $M_{\text{age}}=70.75$ years) from Victoria, B.C. completed surveys up to four times per day for 14 days and wore Fitbit Charge 2 accelerometers for the duration of the study. The surveys concluded with the cognitive measures. Multi-level models (MLMs) were used to estimate intraindividual and interindividual variation in cognitive performance in relation to BP differences and WP fluctuations in MVPA.

Results: Objective and subjective measures of MVPA were moderately positively correlated. For objectively measures MVPA, MVPA was not associated with cognitive performance at the BP and WP level.

Conclusions: PA has been shown to have robust effects on a variety of within-person dynamics (e.g., well-being, sleep) that are theoretically related to cognitive functioning. Future research may aim to clarify the optimal measurement schedule to better capture the transient effects on cognition. A more robust research design, including larger sample size and increased EMA sampling to enhance WP reliability of cognitive measures may enhance statistical power in capturing the effects of time-varying factors on cognition.

Background

There is a growing body of research supporting the role of certain health behaviours in improving cognition and preventing or prolonging dementias (Moon et al., 2021; Ponvel et al., 2021). Physical Activity (PA) has unequivocally gained the most support in this regard. There are many possible mechanisms by which PA influences cognitive health, starting at the cellular level to more broad influences on other biological systems and even other health behaviours, such as sleep. The American College of Sports Medicine recommend adults participate in a minimum of 30 minutes of moderate-intensity activity on 5 days of the week, or vigorous-intensity aerobic PA for a minimum of 20 minutes on 3 days of the week (Haskell et al., 2007). Notably, evidence for the beneficial effects of both resistance training and aerobic exercise have been reported (Erickson et al., 2019). PA has been demonstrated to benefit a variety of cognitive domains such as executive function, episodic memory, visuospatial function, word fluency, processing speed, and global cognitive function (Barha, Davis, Falck, Nagamatsu, & Lui-Ambrose, 2017; Chang, Labban, Gapin, & Etnier, 2012; Lambourne & Tomporowski, 2010), however, executive functioning processes such as planning, scheduling, working memory, multitasking and inhibition have accrued the most evidence (Colcombe & Kramer, 2003; Erickson et al., 2019; Karr, Areshenkoff, Rast, & Garcia-Barrera, 2014). These benefits have been observed across both cognitively impaired and healthy subjects (Karr et al.; Groot et al., 2016).

The research on PA and cognition is vast and includes studies across the lifespan (children to older adults), with varying definitions of PA (e.g., play, daily activities, structured exercise programs; see Caspersen, Powell, & Christenson, 1985). One important differentiation is between longitudinal studies on the effects of long-term PA, which includes research on the

effects of chronic PA on cognitive aging outcomes, and studies examining shorter term effects of PA, such as those occurring after acute bouts of exercise (Erickson et al., 2019). To date, however, there is a dearth of research examining the coupled association between PA and cognitive abilities within an ecological momentary assessment (EMA) framework. This research will be important to elucidate the effects of PA on cognition within individuals day-to-day life, across varying time scales (e.g., momentary effects, prolonged effects, and effects of chronic PA).

Mechanisms by which PA Influences Brain Health

At the cellular level, PA directly influences the expression of neurotransmitters and neurotrophic factors, which in turn influence synaptic plasticity and cell proliferation and survival (Erickson et al., 2019). PA exposes neurons to metabolic stress eliciting an adaptive and protective response. It has been demonstrated to enhance mRNA protein expression of vascular endothelial growth factor (VEGF), which plays a critical role in mitigating ischemic infarct damage, alleviating neurological deficits, and promoting survival of new neurons (Wang, Dong, Zhang, & Zhang, 2015). At a more macro-level, PA mitigates vascular risks which in turn prevents or prolongs the onset of cognitive impairment and dementia due to cerebrovascular factors (Aarsland, Sardahaee, Anderssen, Ballard, & the Alzheimer's Society Systematic Review group, 2010). There is strong evidence that exercise is a crucial primary prevention strategy for risk factors such as type 2 diabetes, cardiovascular disease and premature mortality (Colberg et al., 2010; Haskell et al., 2007). PA may also influence extraneous health behaviours, such as improving sleep, which in turn will have beneficial effects on cognition (Erickson et al., 2019).

Additionally, brain-derived neurotrophic factor (BDNF) is a protein found in the hippocampus and cerebral cortex (Lommatzsch et al., 2005) that is up-regulated during exercise

(Rothman & Mattson, 2013). At the acute level, exercise induces transient effects on BDNF, which in turn promotes transient improvements in cognitive skills (Chang, Labban, Gappin, & Etnier, 2012). Exercise induced “arousal” for a period of time after exercising has been the leading hypothesis in regard to the mechanisms behind the cognitive benefits incurred just after exercising (Labourn & Tomporowski, 2010). PA has been demonstrated to enhance cognitive skills during exercise, immediately after exercise, and for lengths of time following bouts of exercise (Chang, Labban, Gappin, & Etnier). In the long term, BDNF facilitates neuronal survival, differentiation, axonal and dendritic growth, long-term potentiation (LTP), and synaptic transmission (Knaepen, Goekint, Heyman, & Meeusen, 2010). This typically leads to improvements in learning and memory (Rothman & Mattson, 2013). Through these mechanisms and others, PA is thought to be one of the most critical lifestyle factors that contribute to healthy brain aging.

The Current Body of Research on PA and Cognition

There is ample longitudinal and prospective support for the role of consistent PA in preventing and prolonging the onset of cognitive decline, as well as for beneficial acute effects on cognitive abilities. In a recent scoping review of meta-analyses and systematic reviews on the affects of PA across age groups, neurodivergent and neurodegenerative populations, and a variety of health conditions, *long-term* and *acute bouts* of MVPA were shown to be associated with improved cognitive outcomes and cognitive performance, respectively (Erickson et al., 2019). Overall, moderate effects were concluded for adults over the age of 50 for both acute bouts and chronic MVPA, with exercise interventions showing improvements in brain structure and function, as well as cognitive outcomes.

Erickson et al.'s (2019) review defines chronic PA as exercise that is repeated and lasts longer than a single session or episode. As such, the effects of chronic PA reflect a change in an individual's baseline, that is not tightly coupled in time to the last bout of exercise. Studies on chronic exercise include PA behaviour over a span of weeks, months, or years. Much of the research on the long-term effects of PA is with older adults, with findings from prospective observational research consistently supporting the benefits of a physically active lifestyle on cognitive outcomes with aging (Erickson et al., 2019; Kramer & Erickson, 2007). Erickson et al.'s review found strong evidence that greater amounts of PA are associated with a reduced risk of cognitive decline. For example, two large-scale prospective studies (Beckett, Ardern, & Rotondi, 2015; Sofi et al., 2011) both reported an approximate 40% reduced risk of developing dementia in individuals who were relatively more active, and the maintenance of midlife exercise has been linked to better cognitive outcomes (Palta et al., 2019). Conversely, low levels of self-reported PA have been associated with the development of Alzheimer's disease (AD) (Buchman et al., 2012; Erickson, Weinstein, & Lopez, 2012; Podewils et al., 2005) and cognitive decline (Yaffe, Barnes, Nevitt, & Lui, 2001) in older adults. Briefer intervention studies with elderly participants have also demonstrated significant gains in cognitive performance (Colcombe et al., 2006; Kramer & Erickson, 2007).

Erickson et al.'s (2019) review also determined that there is strong evidence for the effects of acute bouts of moderate PA on transient improvements in cognition during the post recovery period. It was reported that larger effects (Hedge's g) were realized for older adults (.67 [.40, .93]), relative to adolescents and young adults for tasks of executive functioning, with similar age differences in effect sizes for other aspects of cognition. Eleven-to-twenty-minute bouts of PA were found to be most beneficial, with smaller effects on cognition for shorter or

longer stints of exercise (Chang, Labban, Gapin, & Etnier, 2012). Notably, this research has implications when examining the effects of PA as it occurs in participants day to day life, such that daily exercise will likely be coupled with enhanced cognitive abilities.

Ecological momentary assessment (EMA) methodology gathers samples of an individual's behaviours (and/or mood, environmental and contextual factors, etc.) through briefer and more frequent assessments within the person's naturally occurring environment, as close in time to these experiences as possible (Cain, Depp, & Jeste, 2009) in order to maximize ecological validity (Shiffman, Stone, & Hufford, 2008). EMA has been ubiquitously employed social science and health research to better understand complex relationships, such the association between contextual factors and chronic pain experiences (May, Junghaenel, Ono, Stone, & Schneider, 2018), and situational factors that precipitate eating disorder behaviours and cognitions (Engel et al., 2016) and alcohol consumption (Jones, Tiplady, Houben, Nederkoorn, & Field, 2018). It is advantageous in ageing research, as it reduces memory bias and captures intraindividual variability.

In their review of EMA in aging research, Cain, Depp, and Jeste (2009) found that the most common assessment content was for affect, activities of daily living, PA, and social exchanges. Though infrequent, the sensitivity of ecologically embedded cognitive testing to environmental and contextual factors has been demonstrated in past studies (Allard et al., 2014; Riediger et al., 2014; Thomson, Nimmo, Tiplady, & Glen, 2009; Tiplady, Oshinowo, Thomson, & Drummond, 2009). For example, Allard et al. found that momentary reports of engagement in intellectually stimulating activities were prospectively associated with increased semantic memory performance. However, there are few studies taking advantage of EMA to elucidate the effects of PA on cognition on a shorter timescale, as it naturally occurs, despite evidence to

suggest this relationship exists. In Zapata-Lamana, Lalanza, Losilla, Parrado, and Capdevila's (2020) review of smartphone-based EMA studies on PA, only three out of 76 studies were in older adults, and none looked at the association between PA and cognition. This is surprising, given the usefulness in understanding how an active lifestyle influences cognitive abilities within an individual's day to day experience, rather than in a contrived research setting.

Moreover, within the PA literature, consistently and precisely quantifying activity remains a limitation. Many studies employ self-report measures, despite influences from recall biases, health status, depression, and cognitive ability, especially for older adults (Zhu et al., 2017). For example, a meta-analytic review found correlations from $-.71$ to $.96$ between self-report and objective measures of PA in adult comparison studies (Prince et al., 2008), suggesting remarkable inconsistencies across studies. This is highly relevant for studies with older adults, as large discrepancies between self-reported and accelerometer-assessed PA have been linked to cognitive functioning (Herbolsheimer, Riepe, & Peter, 2018). Additionally, self-report measures typically focus on structured exercise only. However, PA has been defined as "any bodily movement produced by skeletal muscles that results in energy expenditure" (Caspersen, Powell, & Christenson, 1985, p. 126). As such, lifestyle physical activity, such as household or occupational, captures the full spectrum of activities which may be considered PA. This energy expenditure may be more accurately captured by objective measures of PA such as accelerometers, which have been demonstrated to be acceptable and accurate when used in older adult populations (Tedesco et al., 2019).

To date, few studies examining the effects of PA on cognition have employed objective measures of PA, likely due to cost and practical obstacles in large-scale research. Though some studies exist (Johnson et al., 2016; Kerr et al., 2013; Wu et al., 2020; Zhu et al., 2017), cognitive

skills have not been measured in relation to fluctuation in PA (i.e., typically, baseline measures of cognitive abilities are used instead). That is, integrating objective accelerometer measures of PA within an EMA protocol alongside measures of cognition has not been done to the authors knowledge. Through objectively quantifying interindividual and intraindividual variability in PA and the relationship between frequent and brief cognitive assessments on a relatively shorter time scale, a more clear and ecologically valid picture of the BP and WP effects of PA on cognition can be elucidated.

The Daily experiences of Affect, Stress, and Health Study (DASH).

As discussed, the extant literature suggests that PA may be responsible for both BP and WP differences in cognitive performance. That is, chronically active individuals, or those more active on average, incur cognitive benefits compared to generally inactive individuals, while stints of activity cause transient cognitive improvements WP. However, the effects of lifestyle PA (i.e., daily activities that result in energy expenditure, such as chores, leisurely exercise, etc. in addition to more formal activities) on cognitive abilities from day to day has not been explored. The purpose of this manuscript is to determine if, and to what extent lifestyle PA contributes to variability in cognitive performance measured using smartphone-based assessments within an EMA protocol. Specifically, it is hypothesized that MVPA will be associated with BP variability in cognitive performance (i.e., relatively more active individuals will perform better on cognitive tests), as well as WP variability in cognitive performance (i.e., individuals will perform better on cognitive testing on days when they are more active than usual). Moreover, self-report measures of MVPA will be compared to objectively measured MVPA. In light of past research, these measures are not expected to be strongly correlated.

However, objectively measured MVPA using Fitbit accelerometers is anticipated to be related to cognitive performance as described.

Methods

Procedure

The DASH study procedure has been described in Chapter 1, and the study participants and protocol are identical. Briefly, participants in DASH were older adults aged 65-75 years. They completed EMAs consisting of brief daily surveys 4 times per day on a smartphone, for 14 days. Each survey included questions and rating scales about contextual, lifestyle, and psychosocial factors, including a question about PA levels that day. Each survey concluded with cognitive assessments, including the visual working memory task (DM), processing speed task (SS), and an executive functioning task (ST).

Additionally, participants were required to wear Fitbit accelerometers to measure their level of activity. They were instructed to keep the Fitbit on at all times, and charge it when they were sedentary, such as when watching TV. Prior to beginning the EMA portion of the study, participants attended an intake session at the Institute on Aging and Lifelong Health at the University of Victoria. Here, a member of the research team reviewed appropriate use of the Fitbit. They were lent a “FitKit” consisting of an Android device for survey completion (with the survey installed), Fitbit, and the appropriate chargers.

Measures

Demographic factors were obtained on the intake survey, and cognitive outcomes and degree of self-reported effort on each cognitive task were measured in the EMA surveys, as described in Chapter 1.

Self-Reported PA. Unique to the end of day (EOD) survey only, participants completed a self-report PA question, indicating how long they engaged in MVPA that day (**EOD-PA-SR**, where SR stands for self-report). They selected from the response options “0”, “30”, “45”, “60”, “1 hour, 30 mins”, “2 hours”, “2 hours and 30 mins.”, or “more than 3 hours”. This variable was then made numeric and used as a continuous predictor.

Objectively Measured PA. PA was objectively recorded using the Fitbit Charge 2 wrist-worn accelerometer. Two scores were generated from the Fitbit data for each participant to capture daily activity and total activity. First total active minutes spent in moderate and vigorous activity that day were summed to get a daily total, and then averaged at the level of the participant to get a mean score for each person in the study (**AV-PA-O**, where O stands for objective measure). This person average score was centered upon the grand mean, such that 0 represents a participant’s average MVPA level across the study that is commensurate with the overall study mean MVPA. A positive score represents a participant’s study average that is higher than the overall study average, and a negative score represents a participant’s study average that is lower than the overall study average. In this way, the effects of relatively higher or lower activity levels relative to the sample can be examined at the BP level (i.e., a comparison of more active to less active participants). Second, for the analyses of WP variability in MVPA, each participant’s daily total MVPA score was centered upon the persons mean level of MVPA across the study. In this was, a daily score of 0 represents a day where a participant engaged in their average level of activity, a positive score represents more activity that day than their usual, and a negative score represents less activity that day than their usual. This WP person daily Fitbit score will be referred to as **Day-PA-O**.

Analytical Approach

All data analyses were conducted in RStudio 2022.02.1+461. Pearsons Product-Moment Correlation was used to examine the association between objectively measured and self-reports of MVPA. A two-level coupled change mixed multi-level model (MLM) was estimated according to Model 1 below. Time varying predictors (day, daily effort, and WP changes in MVPA) were specified at the first level. β_{1i} represents the rate of change for person i , γ_{10} represents population-level trends in the slope, and U_{1i} represents individual variation from it, such that a unique trajectory for each individual is computed. Similarly, β_{4i} represents day to day fluctuations in MVPA for each person i , and γ_{40} represents population level trends in the effects of relatively more or less MVPA than a person's usual performance on average.

$$\text{Level 1: } \text{Cog}_{ti} = \beta_{0i} + \beta_{1i}(\text{Day}_{ti}) + \beta_{2i}(\text{Day}^2_{ti}) + \beta_{3i}(\text{effort}_{ti}) + \beta_{4i}(\text{Day-PA-O}_{ti}) + e_{ti}$$

$$\text{Level 2: } \beta_{0i} = \gamma_{00} + \gamma_{01}(\text{Age}_i) + \gamma_{02}(\text{Sex}_i) + \gamma_{03}(\text{Edu}_i) + \gamma_{04}(\text{AV-PA-O}_i) + U_{0i}$$

$$\beta_{1i} = \gamma_{10} + U_{1i}$$

$$\beta_{2i} = \gamma_{20}$$

$$\beta_{3i} = \gamma_{30}$$

$$\beta_{4i} = \gamma_{40}$$

Model 1

Results

Correlation Between Objective and Subjective MVPA

On average, participants engaged in 49.07 minutes of MVPA per day (SD = 52.79, range 0 to 329) as objectively measured by Fitbit, and 55.97 minutes of MVPA per day (SD = 48.08, range 0 to 180) according to self-reported accounts. The correlation between Day-PA-O and EOD-PA-SR was found to be moderately positively correlation, $r(847) = .44$, $p < .001$. Given this association, only objectively measures MVPA was included in the final model.

MLM Results for WP and BP Effects of PA on Cognition

Findings show both linear and quadratic time to be significantly associated with cognitive scores, such that individuals tend to become more proficient at the tasks over time. Mainly, individuals become faster at the ST and SS task, and more accurate at the DM task. Both BP and WP MVPA levels did not influence cognitive scores. The amount of effort put into the task was only associated with ST performance, where higher effort resulted in significantly slower scores. Education was associated with DM performance, such that higher education resulted in better performance. Age and Sex were also marginally significant for DM; Older individuals performed somewhat worse, and males tended to score slightly more accurately.

Table 2.1 Multilevel Estimates of the Effects of Session, Effort, and Demographic Factors, and Physical Activity on Cognitive Tasks

Variable	Dot Memory	Stroop	Symbol Search
	Estimate (SE)	Estimate (SE)	Estimate (SE)
Fixed Effects			
WP variables			
Intercept (γ_{00})	12.87 (.90) ***	1773.84 (66.41) ***	2764.56 (107.85) ***
Linear time (γ_{10})	-4.76 (0.98) ***	-444.10 (56.39) ***	-763.17 (99.98) ***
Quadratic time (γ_{20})	1.50 (.70) *	210.05 (37.96) ***	339.51 (70.92) ***
Day-PA-O (γ_{40})	.00 (.00)	.017 (.14)	.21 (.26)
Effort (γ_{30})	-.01 (.02)	2.54 (1.19) *	1.94 (2.05)
BP Variables			
Age (γ_{01})	0.24 (.14) .	-0.63 (10.49)	5.79 (16.01)
Sex (γ_{02})	-2.14 (1.17) .	-32.12 (87.34)	50.97 (134.22)
Edu (γ_{03})	-1.47 (.63) *	3.09 (46.76)	-90.82 (71.57)

AV-PA-O (γ_{04})	0.01 (.02)	.53 (1.18)	-1.58 (1.801)
Random effects			
	Variance (SD)	Variance (SD)	Variance (SD)
WP (σ_e^2)	8.37 (2.89)	24669 (157.1)	87067 (295.1)
Between-person			
Intercept (σ_0^2)	13.77 (3.71)	80807 (284.3)	250363 (500.4)
Linear time (σ_1^2)	4.47 (2.11)	33270 (182.4)	49772 (223.1)

Note. based on 797, 799, and 802 observations for DM, ST, and SS respectively. Lower values indicate better performance on all cognitive tasks. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Discussion

This project examined the relationship between day-to-day lifestyle MVPA, as objectively measured using Charge 2 Fitbit accelerometers, and cognitive outcomes on brief and frequent EMA of cognitive functioning, as well as looked at the association between objectively measured and subjectively reported activity levels. The association between objectively measured MVPA and daily subjective reports of MVPA was moderate and positive at .44. This is in keeping with current research, showing a correlation of $r = .36$, $p < 0.0001$ in a large-scale study with individuals aged 8 to 79-years ($n = 2,372$) comparing self-reported MVPA with accelerometer measured MVPA (Colley, Butler, Garriguet, Prince, &, Roberts, 2018).

Objectively measured MVPA was not related to cognitive outcomes at the WP level across days. In light of past research finding a link between acute exercise and transient cognitive improvements, it was hypothesized that on days when individuals exercised more than their usual, they would incur cognitive benefits. However, this was not the case on any of the three measures. Notably, as reported in chapter 1, the WP reliabilities for DM, ST, and SS at the daily level were .24, .75, and .77, respectively. Though the reaction time tasks had relatively good WP

reliabilities, these values may be suboptimal across measures in capturing coupled associations between cognitive performance and time-varying state factors, such as exercise (Brewster et al., 2021). However, low reliabilities alone are usually not problematic within an EMA protocol in light of the number of observations potentially offsetting these effects (Sliwinski et al., 2018). In fact, previous studies have reported significant effects of time-varying variables on outcomes with low to modest WP reliabilities, such as Charles, Piazza, Mogle, Sliwinski, & Almeida's (2013, as cited in Sliwinski et al., 2018) study on time-varying influences on negative affect.

Additionally, future research is needed in order to determine the optimal measurement schedule for capturing any possible effects. It is highly possible that the temporal relationship between when PA occurred and when the survey was completed will make a difference. The present study looks at activity and scores on cognitive tests *across the day*. A more fine-grained approach, examining the effects of bouts of activity on cognitive tests performed just after may hold more promise in showing WP associations. As mentioned, acute bouts of PA have been linked to cognitive improvements for a range of time following exercise (Chang, Labban, Gapin, & Etnier, 2012), however, this interval is not well-defined within the literature. Moreover, little is known about the optimal interval in relation to the initial intensity and duration of PA. Though combining moderate and vigorous exercise to create a PA variable in the present study is in keeping with past research, these distinct intensities may affect cognition differently.

Incongruent to previous research, we did not find an association between PA and cognitive outcomes at the BP level. That is, inconsistent with past findings, individuals who were relatively more active did not have higher levels of cognitive performance. This is likely explained by sample selectivity. The DASH sample in the present study engaged in MVPA for approximately 1-hour per day, which is well above population characteristics, as Statistics

Canada (2013) reported that only 1 in 5 older adult Canadians achieve the recommended 150 minutes of MVPA per week. Another potential consideration related to this is the initial fitness level of participants (Chang, Labban, Gapin, & Etnier, 2012). Participants in the present study may have been more fit compared to the general population. Though BP variability in MVPA existed, examination of the histogram in Appendix A highlights that individual deviations from the sample's mean MVPA level was relatively minimal. This point underscores the importance of obtaining a sample with diverse characteristics, a limitation of the present study (i.e., all highly educated, healthy, community dwelling older adults in the Victoria area). Given that past research has established PA as critical for cognitive health and cognitive aging outcomes, future research examining daily PA and the long-term benefits of maintaining daily activity are required.

References

- Aarsland, D., Sardahaee, F. S., Anderssen, S., Ballard, C., & the Alzheimer's Society Systematic Review group. (2010). Is physical activity a potential preventive factor for vascular dementia? A systematic review. *Aging & mental health, 14*(4), 386-395.
- Allard, M., Husky, M., Catheline, G., Pelletier, A., Dilharreguy, B., Amieva, H., ... & Swendsen, J. (2014). Mobile technologies in the early detection of cognitive decline. *PLoS One, 9*(12), e112197.
- Barha, C. K., Davis, J. C., Falck, R. S., Nagamatsu, L. S., & Liu-Ambrose, T. (2017). Sex differences in exercise efficacy to improve cognition: a systematic review and meta-analysis of randomized controlled trials in older humans. *Frontiers in neuroendocrinology, 46*, 71-85.
- Beckett, M. W., Ardern, C. I., & Rotondi, M. A. (2015). A meta-analysis of prospective studies on the role of physical activity and the prevention of Alzheimer's disease in older adults. *BMC geriatrics, 15*(1), 1-7.
- Buchman, A. S., Boyle, P. A., Yu, L., Shah, R. C., Wilson, R. S., & Bennett, D. A. (2012). Total daily physical activity and the risk of AD and cognitive decline in older adults. *Neurology, 78*(17), 1323-1329.
- Cain, A. E., Depp, C. A., & Jeste, D. V. (2009). Ecological momentary assessment in aging research: a critical review. *Journal of psychiatric research, 43*(11), 987-996.
- Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public health reports, 100*(2), 126.

- Chang, Y. K., Labban, J. D., Gapin, J. I., & Etnier, J. L. (2012). The effects of acute exercise on cognitive performance: a meta-analysis. *Brain research, 1453*, 87-101.
- Colberg, S. R., Sigal, R. J., Fernhall, B., Regensteiner, J. G., Blissmer, B. J., Rubin, R. R., ... & Braun, B. (2010). Exercise and type 2 diabetes: the American College of Sports Medicine and the American Diabetes Association: joint position statement. *Diabetes care, 33*(12), e147-e167.
- Colcombe, S. J., Erickson, K. I., Scalf, P. E., Kim, J. S., Prakash, R., McAuley, E., ... & Kramer, A. F. (2006). Aerobic exercise training increases brain volume in aging humans. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 61*(11), 1166-1170.
- Colcombe, S., & Kramer, A. F. (2003). Fitness effects on the cognitive function of older adults: a meta-analytic study. *Psychological science, 14*(2), 125-130.
- Colley, R. C., Butler, G., Garriguet, D., Prince, S. A., & Roberts, K. C. (2018). Comparison of self-reported and accelerometer-measured physical activity in Canadian adults. *Health Rep, 29*(12), 3-15.
- Engel, S. G., Crosby, R. D., Thomas, G., Bond, D., Lavender, J. M., Mason, T., ... & Wonderlich, S. A. (2016). Ecological momentary assessment in eating disorder and obesity research: a review of the recent literature. *Current psychiatry reports, 18*(4), 37.
- Erickson, K. I., Hillman, C., Stillman, C. M., Ballard, R. M., Bloodgood, B., Conroy, D. E., ... & Powell, K. E. (2019). Physical activity, cognition, and brain outcomes: a review of the 2018 physical activity guidelines. *Medicine and science in sports and exercise, 51*(6), 1242.

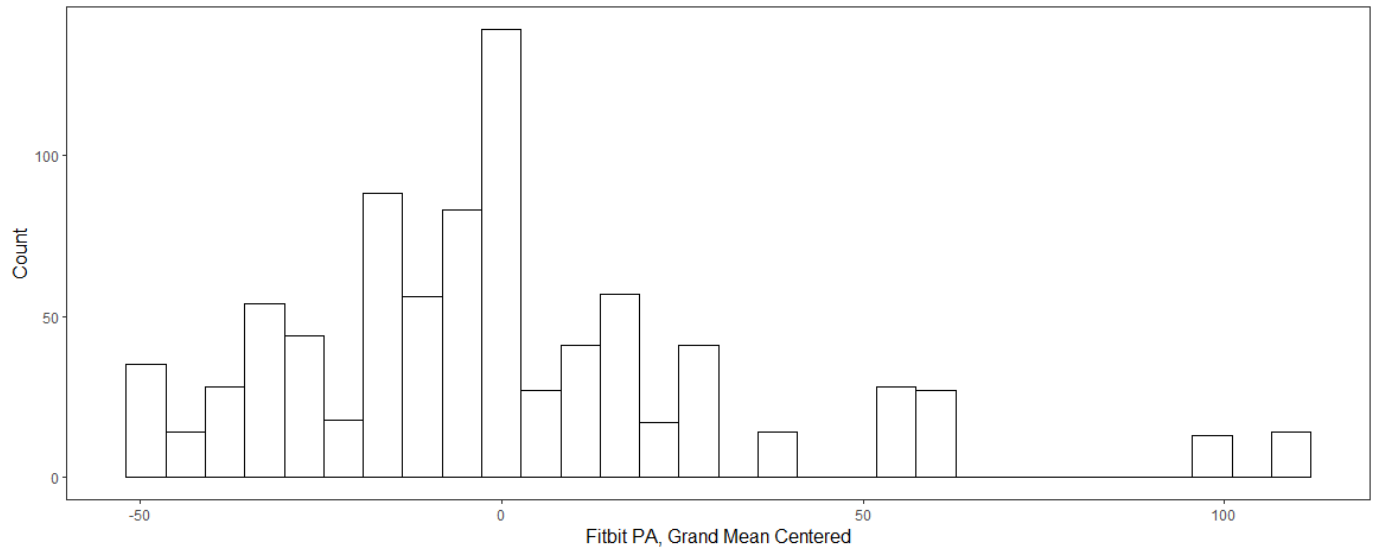
- Erickson, K. I., Weinstein, A. M., & Lopez, O. L. (2012). Physical activity, brain plasticity, and Alzheimer's disease. *Archives of medical research*, *43*(8), 615-621.
- Gheysen, F., Poppe, L., DeSmet, A., Swinnen, S., Cardon, G., De Bourdeaudhuij, I., ... & Fias, W. (2018). Physical activity to improve cognition in older adults: can physical activity programs enriched with cognitive challenges enhance the effects? A systematic review and meta-analysis. *International Journal of Behavioral Nutrition and Physical Activity*, *15*(1), 1-13.
- Groot, C., Hooghiemstra, A. M., Raijmakers, P. G. H. M., Van Berckel, B. N. M., Scheltens, P., Scherder, E. J. A., ... & Ossenkoppele, R. (2016). The effect of physical activity on cognitive function in patients with dementia: a meta-analysis of randomized control trials. *Ageing research reviews*, *25*, 13-23.
- Herbolsheimer, F., Riepe, M. W., & Peter, R. (2018). Cognitive function and the agreement between self-reported and accelerometer-accessed physical activity. *BMC geriatrics*, *18*(1), 1-9.
- Johnson, L. G., Butson, M. L., Polman, R. C., Raj, I. S., Borkoles, E., Scott, D., ... & Jones, G. (2016). Light physical activity is positively associated with cognitive performance in older community dwelling adults. *Journal of science and medicine in sport*, *19*(11), 877-882.
- Jones, A., Tiplady, B., Houben, K., Nederkoorn, C., & Field, M. (2018). Do daily fluctuations in inhibitory control predict alcohol consumption? An ecological momentary assessment study. *Psychopharmacology*, *235*(5), 1487-1496.

- Kerr, J., Marshall, S. J., Patterson, R. E., Marinac, C. R., Natarajan, L., Rosenberg, D., ... & Crist, K. (2013). Objectively measured physical activity is related to cognitive function in older adults. *Journal of the American Geriatrics Society*, *61*(11), 1927-1931.
- Lambourne, K., & Tomporowski, P. (2010). The effect of exercise-induced arousal on cognitive task performance: a meta-regression analysis. *Brain research*, *1341*, 12-24.
- Haskell, W. L., Lee, I. M., Pate, R. R., Powell, K. E., Blair, S. N., Franklin, B. A., ... & Bauman, A. (2007). Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Circulation*, *116*(9), 1081.
- Karr, J. E., Areshenkoff, C. N., Rast, P., & Garcia-Barrera, M. A. (2014). An empirical comparison of the therapeutic benefits of physical exercise and cognitive training on the executive functions of older adults: A meta-analysis of controlled trials. *Neuropsychology*, *28*(6), 829.
- Knaepen, K., Goekint, M., Heyman, E. M., & Meeusen, R. (2010). Neuroplasticity—exercise-induced response of peripheral brain-derived neurotrophic factor. *Sports medicine*, *40*(9), 765-801.
- Kramer, A. F., & Erickson, K. I. (2007). Capitalizing on cortical plasticity: influence of physical activity on cognition and brain function. *Trends in cognitive sciences*, *11*(8), 342-348.
- Liu, S. H., Eaton, C. B., Driban, J. B., McAlindon, T. E., & Lapane, K. L. (2016). Comparison of self-report and objective measures of physical activity in US adults with osteoarthritis. *Rheumatology international*, *36*(10), 1355-1364.
- Lambourne, K., & Tomporowski, P. (2010). The effect of exercise-induced arousal on cognitive task performance: a meta-regression analysis. *Brain research*, *1341*, 12-24.

- Lommatzsch, M., Zingler, D., Schuhbaeck, K., Schloetcke, K., Zingler, C., Schuff-Werner, P., & Virchow, J. C. (2005). The impact of age, weight and gender on BDNF levels in human platelets and plasma. *Neurobiology of aging*, *26*(1), 115-123.
- May, M., Junghaenel, D. U., Ono, M., Stone, A. A., & Schneider, S. (2018). Ecological momentary assessment methodology in chronic pain research: a systematic review. *The Journal of Pain*, *19*(7), 699-716.
- Moon, S. Y., Hong, C. H., Jeong, J. H., Park, Y. K., Na, H. R., Song, H. S., ... & Choi, S. H. (2021). Facility-based and home-based multidomain interventions including cognitive training, exercise, diet, vascular risk management, and motivation for older adults: a randomized controlled feasibility trial. *Aging (Albany NY)*, *13*(12), 15898.
- Palta, P., Sharrett, A. R., Deal, J. A., Evenson, K. R., Gabriel, K. P., Folsom, A. R., ... & Heiss, G. (2019). Leisure-time physical activity sustained since midlife and preservation of cognitive function: The Atherosclerosis Risk in Communities Study. *Alzheimer's & Dementia*, *15*(2), 273-281.
- Podewils, L. J., Guallar, E., Kuller, L. H., Fried, L. P., Lopez, O. L., Carlson, M., & Lyketsos, C. G. (2005). Physical activity, APOE genotype, and dementia risk: findings from the Cardiovascular Health Cognition Study. *American journal of epidemiology*, *161*(7), 639-651.
- Ponvel, P., Shahar, S., Singh, D. K. A., Ludin, A. F. M., Rajikan, R., Rajab, N. F., ... & Kivipelto, M. (2021). Multidomain intervention for reversal of cognitive frailty, towards a personalized approach (AGELESS Trial): study design. *Journal of Alzheimer's Disease*, *82*(2), 673-687.

- Prince, S. A., Adamo, K. B., Hamel, M. E., Hardt, J., Gorber, S. C., & Tremblay, M. (2008). A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *International journal of behavioral nutrition and physical activity*, 5(1), 56.
- Riediger, M., Wrzus, C., Klipker, K., Müller, V., Schmiedek, F., & Wagner, G. G. (2014). Outside of the laboratory: Associations of working-memory performance with psychological and physiological arousal vary with age. *Psychology and aging*, 29(1), 103.
- Rothman, S. M., & Mattson, M. P. (2013). Activity-dependent, stress-responsive BDNF signaling and the quest for optimal brain health and resilience throughout the lifespan. *Neuroscience*, 239, 228-240.
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annu. Rev. Clin. Psychol.*, 4, 1-32.
- Sofi, F., Valecchi, D., Bacci, D., Abbate, R., Gensini, G. F., Casini, A., & Macchi, C. (2011). Physical activity and risk of cognitive decline: a meta-analysis of prospective studies. *Journal of internal medicine*, 269(1), 107-117.
- Tedesco, S., Sica, M., Ancillao, A., Timmons, S., Barton, J., & O'Flynn, B. (2019). Validity Evaluation of the Fitbit Charge2 and the Garmin vivosmart HR+ in Free-Living Environments in an Older Adult Cohort. *JMIR mHealth and uHealth*, 7(6), e13084.
- Thomson, A. J., Nimmo, A. F., Tiplady, B., & Glen, J. B. (2009). Evaluation of a new method of assessing depth of sedation using two-choice visual reaction time testing on a mobile phone. *Anaesthesia*, 64(1), 32-38.

- 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.
- Wang, Y., Li, M., Dong, F., Zhang, J., & Zhang, F. (2015). Physical exercise-induced protection on ischemic cardiovascular and cerebrovascular diseases. *International journal of clinical and experimental medicine*, 8(11), 19859.
- Wu, Z. J., Wang, Z. Y., Hu, B. Q., Zhang, X. H., Zhang, F., Wang, H. L., & Li, F. H. (2020). Relationships of accelerometer-based measured objective physical activity and sedentary behaviour with cognitive function: a comparative cross-sectional study of China's elderly population. *BMC geriatrics*, 20(1), 1-8.
- Yaffe, K., Barnes, D., Nevitt, M., Lui, L. Y., & Covinsky, K. (2001). A prospective study of physical activity and cognitive decline in elderly women: women who walk. *Archives of internal medicine*, 161(14), 1703-1708.
- Zapata-Lamana, R., Lalanza, J. F., Losilla, J. M., Parrado, E., & Capdevila, L. (2020). mHealth technology for ecological momentary assessment in physical activity research: a systematic review. *PeerJ*, 8, e8848.
- Zhu, W., Wadley, V. G., Howard, V. J., Hutto, B., Blair, S. N., & Hooker, S. P. (2017). Objectively measured physical activity and cognitive function in older adults. *Medicine and science in sports and exercise*, 49(1), 47.

Appendix 2.1 Count Distribution of Participant's Study Mean Fitbit Activity Level

Note. 0 represents the study mean activity level.

Chapter 3

The Association Between Stress and Cognitive Abilities using Ecological Momentary Assessment

Abstract

Objective: Though it has been well established that chronic stress is negatively associated with cognition, less is known about the acute effects of stress on cognitive abilities within daily life. The purpose of this study is to determine if, and to what extent perceived stress contributes to both BP (i.e., interindividual differences) and WP (i.e., intraindividual differences) variability in cognitive performance as measured by smartphone-based mobile health (mHealth) assessments within an EMA protocol, as well as explore the effects of perceived and objective stress on these relationships.

Method: Healthy community-residing older adults ($n=66$; $M_{\text{age}}=70.75$ years) from Victoria, B.C. completed surveys up to four times per day for 14 days. These surveys inquired about subjective feelings of stress and concluded with the cognitive measures. Coupled multi-level models (MLMs) were used to estimate intraindividual and interindividual variation in cognitive performance in relation to BP differences and WP fluctuations in stress.

Results: BP differences were found for perceived stress and performance on a processing speed task. However, coupling of stress and cognition were not found across sessions or days.

Conclusions: EMA holds promise for advancing the field of research on stress and cognition through elucidating the impact of naturally occurring stress within the individual daily life. However, this field is incipient and gold-standard measurement protocols have yet to be established. More research is required in order to determine how stress influences cognitive abilities in day-to-day life. Future considerations include determining appropriate measurement schedules, identifying individual moderators, and the development of more cohesive and encompassing measures of stress which capture various dimensions of the stress experience (e.g.,

perception, cognitive processes such as rumination, and a variety of physiological factors associated with stress).

Background

Stress is typically operationally defined as situations where the demand is perceived to be taxing or exceeding one's own capacity or resources to manage it, which threatens one's well-being (Calvo & Guterrez-Garcia, 2016). In response to acute stress, the hypothalamic-pituitary-adrenal axis (HPA) is activated. This leads to a cascade of events that signal the release of glucocorticoids and catecholamines (i.e., stress hormones). Subsequently, metabolic and immune systems alter functioning to mobilize energy so that the organism can respond to the demands of the situation (Tsatsoulis & Fountoulakis, 2006). This adaptive response to stress is termed "allostasis", or stability through change (McEwen, 1998). If this response is chronic and continuous, the system becomes inefficient, resulting in "allostatic load". Thereafter, cardiovascular, metabolic, and inflammatory pathology ensue (Zsoldos et al., 2018), leaving one susceptible to a myriad of tertiary diseases (McEwen & Stellar, 1993). Stress hormones can cross the blood-brain barrier, with the highest occurrence of glucocorticoid receptors located in the hippocampus and frontal lobes (Lupien, Maheu, Fiocco, & Schramek, 2007). The insidious effect of chronic stress on the hippocampus has been extensively studied (Lupien et al., 1996; Seeman, McEwen, Singer, Albert, & Roew, 1997). Specifically, the glucocorticoid-cascade hypotheses (Sapolsky, Krey, & McEwen, 1986) suggests that exposure to glucocorticoids over extended periods of time degrades hippocampal receptors leading to neuronal death, disruption in the HPA regulatory loop, and vulnerability to acute stressors. Additionally, glucocorticoids secreted in response to stress hinder the production of BDNF and exacerbate neuropathological

changes related to dementia (Nation et al., 2011). In animal models, chronic and unpredictable stress has been demonstrated to decrease BDNF levels in the prefrontal cortex and hippocampus (Liu et al., 2014), resulting in neurodegeneration. The damaging effects of stress are speculated to be magnified in older adults, as less efficient biological systems increase vulnerability to stress (McEwen & Morrison, 2013).

Both long-term and acute effects of stress on cognitive abilities have been observed. However, the research on acute stress is mixed compared to the well-established findings on the adverse effects of chronic stress on cognition. Acutely, both positive and negative effects are observed. Some general findings have emerged but are often not consistent across studies. This is likely due to intraindividual differences in the stress response, cross-study differences in the nature of the stressors and sample characteristics, and the differential effects of stress on various domains of functioning (e.g., it may influence memory, executive functioning, and language differently) and processes within domains (e.g., it may differentially influence encoding, retention, retrieval in memory, and working memory, response inhibition, and flexibility in the executive domain) (Shields, 2019). Generally, findings from the acute stress literature suggest that stress intensity largely contributes to subsequent cognitive performance. Mild to moderate stress is often facilitative, particularly for simple tasks with lesser cognitive load, while exposure to higher stress is typically impairing. The latter seems to be especially evident in tasks mediated by the hippocampus and prefrontal cortex, such as explicit memory tasks and those requiring flexible reasoning, while implicitly learned and well-rehearsed task performance is enhanced by high levels of stress (Sandi, 2013).

Mechanisms have been explored for both opposing actions of acute stress. For instance, acute stress can be beneficial since the physiological response is mobilizing, such as the

upregulation of glutamatergic, excitatory brain systems (Chaouloff & Groc, 2011). Moreover, dopamine and norepinephrine in mild to moderate increases are considered to enhance functional connectivity with the prefrontal cortex, while excessive release in highly stressful situations impairs prefrontal functions (Arnsten, 2009; Porcelli et al., 2008). Acute stress is also suspected to hinder performance through decreased neural efficiency (Wu & Yan, 2017), the reallocation of cognitive resources (Hasher & Zacks, 1979; Kahneman, 1973), and cognitive interference (i.e., interference from stress-related thoughts; Klein & Boals, 2001; Stawski, Sliwinski, & Smyth, 2006).

Researching the effects of acute stress on cognition is challenging, and the current state of research presents many limitations. Many of the studies examining the stress-cognition relationship are often highly contrived and lack ecological validity, which may in part account for the variability in findings within the acute stress literature. For example, in-lab endogenous cortisol manipulation (Beckwith, Petros, Scaglione, & Nelson, 1986; Newcomer et al., 1999) and lab induced psychosocial stressors, such as mental arithmetic tasks (Starcke, Wiesen, Trotzke, & Brand, 2016; Qi et al., 2016) are most often used to study stress in-lab. Moreover, there are also ethical limitations to the level of stress that can be induced within the laboratory setting (Wu & Yan, 2017). Taken together, this means that the nature and intensity of stress experienced in daily life is likely much different than that in research settings, calling into question generalizability of the extant literature. When daily stressors are assessed (rather than laboratory-based stress), typically self-report checklists of stressful events are used, rather than considering the idiosyncratic perceptions of a potential stressor. However, it has been demonstrated that individuals' subjective appraisal, or reactivity to stressors, is more important for outcomes compared to the "objective" experience of a stressor per se (Piazza, Charles, Sliwinski, Mogle, &

Almeida, 2013; Leung, et al., 2016). That is, a particular event may impact people differently, depending on how they perceive and react to the situation (e.g., viewing it as unmanageable and surpassing their resources to cope, versus manageable).

To enhance ecological validity, measures of stress have been embedded in Ecological Momentary Assessment (EMA) studies, where the effects of naturally occurring stressors as they occur within an individual's daily environment are explored. For example, stress has been related to *self-reported* (i.e., subjective) memory lapses (Neupert, Almeida, Mroczek, & Spiro, 2006). However, objective cognitive testing has rarely been integrated into EMA protocols. When cognitive tests are employed (rather than self-reported/subjective cognitive performance), systematic within-person (WP) variance in cognitive performance of older adults has been observed at both weekly testing intervals (Hertzog, Dixon, & Hulstsch, 1992; Li, Aggen, Nesselroade, & Baltes, 2001) and across multiple daily tests within a week (Brewster et al., 2021; Salthouse & Berish, 2005; Sliwinski et al., 2018). However, very few studies have examined factors that may be responsible for these fluctuations in cognitive skills, such as stress. One study examined the link between stress and fluctuations in cognitive performance on *in-lab* cognitive tasks (Sliwinski, Smyth, Hofer, & Stawski, 2006), finding that on days where more stress is reported, individuals performed more slowly on a working memory (2-back) and divided attention (2-count) tasks. However, to date, there are no EMA studies examining the relationship between perceived stress and cognition as it occurs in a naturalistic setting.

The Daily experiences of Affect, Stress, and Health Study (DASH)

The present study addresses many of the aforementioned limitations. EMA (four brief surveys per day, for 14 days) was used to capture stress as it occurs within participants' daily lives, alongside brief embedded measures of cognition. The effects of fluctuations in momentary stress at each survey and daily stress, (i.e., WP, intraindividual variability), as well as average

level of participant stress (between-person; BP, interindividual variability) were examined in relation to cognition. Importantly, *perceived stress* was used as a predictor (i.e., subjective feeling of stress), rather than the objective experience of what could be a stressor. At the BP level, it is expected that individuals who report more subjective feelings of stress on average will perform worse on cognitive tests. At the WP level, it is expected that on days when a person reports more feelings of daily stress than their usual level, they will also perform worse on cognitive testing that day; The same effect is expected when an individual reports relatively more feelings of stress than their usual prior to completing a survey (i.e., momentary stress). However, the occurrence of a stressor in and of itself is not expected to be associated with cognitive performance at that testing occasion.

Method

Procedure

The DASH study procedure has been described in Chapter 1, and the participants and study protocol are identical. Briefly, participants in DASH were older adults aged 65-75 years. They completed EMAs consisting of brief daily surveys four times per day on a smartphone, for 14 days. Each survey included questions and rating scales about contextual, lifestyle, and psychosocial factors, including a question about stress during the day at the end of the day. Each survey concluded with “brain games” that assessed several domains of cognition, including the visual working memory task (DM), processing speed task (SS), and an executive functioning task (ST). Prior to beginning the EMA portion of the study, participants attended an intake session at the Institute on Aging and Lifelong Health at the University of Victoria. Here, a member of the research team reviewed the study protocol, and participants were lent an Android device for survey completion (with the survey installed).

Measures

Demographic factors were obtained on the intake survey, and cognitive outcomes and degree of self-reported effort on each cognitive task were measured in the EMA surveys, as described in Chapter 1. Stress was measured at the end of day, using items from the Perceived Stress Scale (PSS), and on each survey within the day (besides the morning survey), participants were asked about the occurrence of stressful events and indicated how stressed they felt in that moment. WP variability in momentary stress, and baseline PSS ratings have been associated with fluctuations in ambulatory measures of blood pressure, and BP differences in blood pressure, respectively (Bowen et al., 2014).

Perceived Stress Scale (PSS) Unique to the end of day (EOD) survey only, participants completed an abbreviated version of the PSS (Cohen, Kamarch, & Mermelstein, 1983). Specifically, the PSS evaluates the degree to which individuals believe their life has been unpredictable, uncontrollable, and overloaded during the previous month. The items are general, rather than pertaining to specific experiences or events. Participants were asked to reflect on their day (“over the course of the day, how often have you...”) and rate the following 4 items on a sliding scale from never to often (which corresponded to 0 to 100): “Been upset because something happened unexpectedly”; “Been angered because of things that were outside of your control”, “Felt confident about your ability to handle your personal problems”, and “felt like you could not cope with all the things you have to do?” For each participant, the responses to these items were summed for each day to create a PSS-Day score (with the third item reverse scored). This score was used to create a daily WP stress variable (i.e., WP fluctuations in stress across days) and a BP stress variable (i.e., average stress experienced by the participant across the 14 days) as follows:

PSS-BP To examine the BP effects of perceived stress, each participant's PSS-Day ratings were averaged to yield a participant mean score. Each participant's PSS mean score was centered upon the average of the sample's PSS ratings, so that a score of 0 reflects a participant with a mean PSS score equal to the sample's mean score, a positive score represents a participant with a higher average PSS score across the study, and a negative score represents a participant who is reporting less stress on average than the study average. In this way, the BP effects of perceived stress (**PSS-BP**) could be interpreted with ease.

PSS-Day-WP To examine the effects of daily perceived stress on daily cognitive performance, each person's PSS-Day score was person-mean centered (Hoffman, 2015). Specifically, the average of each participant's PSS-Day scores across the entire study was subtracted from their PSS rating each day, so that a **PSS-Day-WP** score of 0 represents a day where that individual reported their average level of stress on the EOD PSS measure, a positive score represents a day where an individual reported more feelings of stress than their usual, and a negative score represents a day where an individual reported less stress than their usual.

Momentary Stress (MS) The occurrence of stressful events and appraisal of the event was measured at each testing occasion (excluding the morning survey). Participants were asked if anything stressful happened to them *since the last survey* and then asked to choose the most salient option amongst a list. They chose from the following response options: "No, nothing stressful happened"; "negative social interaction"; "health issue (yours)"; "health issue (spouse/partner)"; "a problem someone else had"; "a mistake you made"; "financial problem"; "heard bad or upsetting news"; "time pressure/running late"; "other stressful experience". This measure of objective stressor will be referred to as **MS-Objective**; It is equal to 0 if no stressor was endorsed and 1 if a stressor was endorsed.

To account for appraisal of the stressor, participants answered two rating scale questions. First, they rated their experience on the dimension of unpleasantness at the time the stressor occurred (sliding scale from not at all unpleasant to extremely unpleasant, which ranged from 0 to 100, respectively). Then, they rated how their feelings about it changed at the time of the survey (sliding scale from feeling worse to feeling better, which ranged from 0 to 100, respectively). This latter variable was centered to range from -50 (worse) to 50 (better), with 0 representing no change in the participants feelings about the stressful event. To create a variable that considers the overall subjective impact of the stressor, the degree of change was subtracted from their unpleasantness rating. For example, if Person A appraised the stressor to be 50 on the unpleasantness scale and they were feeling better by 20 points on the change scale at the time of the survey, then their acute rating would be 30. Alternatively, if they felt worse about it as indicated by a ranking of -20 on the change sliding scale, then their acute rating would be 70 ($50 - (-20)$). If, for example, Person B ranked the initial unpleasantness as 100 (much higher than person A), and also reported feeling better by 20 on the change scale, their score of 80 ($100 - 20$) will still reflect a more impactful subject experience than Person A who ranked the initial unpleasantness as 50 (resulting in the score of 30). Similarly, if the individual indicated feeling worse (-20), then their score would be 120 ($100 - (-20)$). In this way, both the initial severity of perceived stress is accounted for, along with participants change in their feeling about the stressor to get an overall picture of the subjective impact of the stressor at the time of the survey. As such, a higher score reflects a more subjectively distressing experience at the time of the survey. This variable was further within person centered, so that a score of 0 represents the persons mean feelings of subjective stress, a positive score indicates that they are more stressed

than their usual at that survey, and a negative score indicates they are less stressed than usual at that survey. This variable will be referred to as **MS-Subjective-WP** hereafter.

Analytical Approach

All data analyses were conducted in RStudio 2022.02.1+461. The data was nested such that sessions were nested within days, which were further nested within person. Two, 2-level mixed multi-level model (MLM) were estimated (Model 1 and Model 2). Model 1 uses session as the time metric in order to determine the effects of MS-Subjective-WP (i.e., feelings of stress at that session) and the effects of MS-Objective (i.e., the endorsement of a potential stressor prior to that survey) on cognition. Model 2 uses day as the time metric in order to determine the WP (PSS-Day-WP) and PSS-BP effects of daily perceived stress (i.e., EOD stress and study average stress, respectively, nested within person). In Model 1A, MS-Subjective-WP is specified at the first level of analysis, as a time-varying factor which fluctuates from session to session. In Model 1B, the occurrence of an objective stressor was specified instead. In Model 2, daily fluctuations in PSS ratings (PSS-Day-WP) is modeled in this same way, with day specified as the time metric, and BP PSS included at level 2, as a time-invariant predictor on the intercept.

$$\text{Level 1: } \text{Cog}_{ii} = \beta_{0i} + \beta_{1i}(\text{Session}_{ti}) + \beta_{2i}(\text{Session}^2_{ti}) + \beta_{3i}(\text{effort}_{ti}) + \beta_{4i}(\text{MS-Subjective-WP}_{ti}) \text{ OR} \\ \beta_{3i}(\text{MS-Objective}_{ti}) + e_{ii}$$

$$\text{Level 2: } \beta_{0i} = \gamma_{00} + \gamma_{01}(\text{Age}_i) + \gamma_{02}(\text{Sex}_i) + \gamma_{03}(\text{Edu}_i) + U_{0i}$$

$$\beta_{1i} = \gamma_{10} + U_{1i}$$

$$\beta_{2i} = \gamma_{20}$$

$$\beta_{3i} = \gamma_{30}$$

Model 1 A and B

$$\text{Level 1: } \text{Cog}_{ti} = \beta_{0i} + \beta_{1i}(\text{Day}_{ti}) + \beta_{2i}(\text{Day}^2_{ti}) + \beta_{3i}(\text{effort}_{ti}) + \beta_{4i}(\text{PSS-Day-WP}_{ti}) + e_{ti}$$

$$\text{Level 2: } \beta_{0i} = \gamma_{00} + \gamma_{01}(\text{Age}_i) + \gamma_{02}(\text{Sex}_i) + \gamma_{03}(\text{Edu}_i) + \gamma_{04}(\text{PSS-BP}_i) + U_{0i}$$

$$\beta_{1i} = \gamma_{10} + U_{1i}$$

$$\beta_{2i} = \gamma_{20}$$

$$\beta_{3i} = \gamma_{30}$$

Model 2

Results

In regard to WP findings of time-varying predictors, effort at each session as well as levels of WP stress appraisal (MS-Subjective-WP) at each session were not associated with cognitive performance for any of the cognitive tasks in Model 1A (See Table 3.1). When MS-Objective (i.e., the occurrence of a potential stressor) was used as a time-varying predictor instead, the results remained largely the same, showing no effects on cognition as expected (results not shown here). In Model 2, daily effort was significantly associated with ST performance that day (but no other cognitive outcomes), such that those who reported putting in more effort across the day performed slower on the tasks that day. WP daily stress (as measured by the PSS; PSS-Day-WP) was not related to daily fluctuations in cognition for any of the cognitive outcomes, meaning that more perceived stress within a day than a participant's usual level, did not affect overall cognitive performance that day (See Table 3.2).

In regard to BP effects, in Model 1A, higher educational attainment was significantly associated with more accurate DM scores, but no other cognitive outcomes. Age and Sex were not related to cognitive performance. In Model 2 age was marginally significant for DM only, with older individuals scoring less accurately. Higher education was significantly associated with better DM performance. Lastly, higher levels of perceived stress on the EOD PSS scale on

average (BP effects) resulted in a slower score on SS, meaning that individuals who reported more stress on the PSS scale on average tended to perform worse on the SS task (See Table 3.2).

No association between PSS-BP were found on the other two cognitive tasks.

Table 3.1 Multilevel Estimates of the Effects of Session, Effort, and Demographic Factors, and Momentary Subject Stress Appraisal on Cognitive Tasks (Model 1A)

Variable	Dot Memory	Stroop	Symbol Search
	Estimate (SE)	Estimate (SE)	Estimate (SE)
Fixed Effects			
WP variables			
Intercept (γ_{00})	12.93(1.12) ***	1746.13(89.67) ***	2782.52(127.50) ***
Linear time (γ_{10})	-2.57(1.04) *	-290.50(70.02) ***	-491.23(92.38) ***
Quadratic time (γ_{20})	0.60(.37)	80.45(25.00) **	114.83(33.1) ***
Effort (γ_{30})	-0.03(.02)	1.07(1.89)	0.22(2.24)
MS-Subjective-WP(γ_{40})	.01(.01)	0.37(.56)	0.55(.668)
BP Variables			
Age (γ_{01})	0.14(.16)	10.88(13.03)	8(18.2)
Sex (γ_{02})	1.03(1.33)	-36.48(104.79)	164.75(148.09)
Edu (γ_{03})	-1.82(.71) *	47.01(57.65)	-78.12(80.21)
Random effects	Variance (SD)	Variance (SD)	Variance (SD)
WP (σ_e^2)	29.35(5.42)	40062(200.2)	252834(502.83)
Between-person			
Intercept (σ_0^2)	12.89(3.59)	109677(331.2)	266630(516.36)
Linear time (σ_1^2)	3.55(1.88)	6321(79.5)	9864(99.32)

Note. based on 712, 239, 718 assessments for DM, ST, and SS respectively. Lower values indicate better performance on all cognitive tasks. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 3.2 Multilevel Estimates of the Effects of Session, Effort, and Demographic Factors, and Daily and Participant Mean Stress on Cognitive Tasks (Model 2)

Variable	Dot Memory	Stroop	Symbol Search
	Estimate (SE)	Estimate (SE)	Estimate (SE)
Fixed Effects			
WP variables			
Intercept (γ_{00})	12.68(.9) ***	1755.95(65.87) ***	2743.005 105.347***
Linear time (γ_{10})	-4.57(.97) ***	-439.95(55.75) ***	-755.2(98.65) ***
Quadratic time (γ_{20})	1.41(.69) *	204.75(37.45) ***	332.47(69.98) ***
Effort (γ_{30})	-.01(.019)	2.73(1.19) *	2.46(2.04)
PSS-Day-WP (γ_{40})	0.01(.01)	0.42(.60)	0.58(1.11)
BP Variables			
Age (γ_{01})	.24(.14)	-0.05(10.33)	8.51(15.44)
Sex (γ_{02})	-1.88(1.13)	-31.23(83.79)	54.64(125.68)
Edu (γ_{03})	-1.37(.63) *	15.24(46.63)	-70.71(70)
PSS-BP (γ_{04})	.01(.04)	3.86(2.79)	8.58(4.20) *
Random effects			
WP (σ_e^2)	Variance (SD)	Variance (SD)	Variance (SD)
	8.44(2.91)	24566(156.7)	86856(294.7)
Between-person			
Intercept (σ_0^2)	13.62(3.69)	81687(285.8)	245341(495.3)
Linear time (σ_1^2)	4.46(2.11)	33500(183)	48116(219.4)

Note. based on 810, 812, 815 observations for DM, ST, and SS respectively. Lower values indicate better performance on all cognitive tasks. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Discussion

This project aimed to address current limitation in the stress-cognition literature by examining how perceived stress (as opposed to objective occurrences of a possible stressor) is

related to cognition in a naturalistic setting, rather than within a laboratory where the stress experience is often contrived. Particularly, the coupled relationship between fluctuations in stress (WP, intraindividual effects), along with the effects of differences in average stress levels on cognitive performance (BP, interindividual effects) was explored using EMA methodology. WP stress was defined on two-time scales; Session to session, or momentary feelings of stress, and day to day feelings of global stress. For both timescales, perceived stress was not found to be associated to cognitive functioning. That is, momentary feelings of perceived stress and daily feelings of perceived stress were not found to be related to performance on brief cognitive tasks, contrary to expectations. When the occurrence of an objective experience that could be considered stressful was used as the time-varying predictor (as opposed to subjective *appraisal* of stress at each session), no associations were found as expected. Interestingly, BP effects were found for global feelings of stress on the SS task, showing that higher levels of perceived stress across the study was related to slower performance on the reaction time task. This finding supports the hypothesis that stress hindered speeded performance on SS, though this relationship was not replicated for the other cognitive tasks.

Similar to the PA findings in Chapter 2, WP coupling at the session and day level with cognitive outcomes was not observed. Notably, session level WP reliability (as opposed to daily level reliabilities for PA) involves aggregating across fewer observations, so it is even more difficult to capture coupling at the session level. In the present study, session-level WP reliabilities were .10, .65, and .56 for DM, ST, and SS, respectively (where day-level WP reliabilities were .24, .75, and .77). Moreover, trial level variability (i.e., varying performance on the individual trials of the cognitive task that are aggregated to create a session level score) directly effects the WP reliability of the session-level aggregate scores. As Sliwinski et al. (2018)

points out, time varying factors such as stress may actually increase trial level variability (thus, decreasing reliability of the aggregate), especially in older adults, since reaction time variability increases with age (Rutter, Vahia, Passell, Forester, & Germine, 2021). In future studies, it would be interesting to examine the effects of stress on trial-to-trial level variability of the cognitive tasks.

Additionally, similar to challenges that arose when examining the acute effects of PA on cognition, timing of the assessment in relation to the event may be important here too. The consideration of measurement timing is especially important when studying older adults and stress, as existing theories imply that chronologicity of events is crucial in observing age differences in stress reactivity (Scott, Sliwinski, and Blanchard Fields, 2013). It is thought that older adults are better able to draw on adaptive skills to cope with stress (i.e., Socioemotional selectivity theories of aging; Carstensen, Fung, & Charles, 2003), though this advantage is now considered to be circumstantial and contextual (Strength and Vulnerability Integration theory; Charles, 2010). From these theories, Scott, Sliwinski, and Blanchard Fields (2013) hypothesize that stress may impact older adults similar to, or even more, than younger adults at the time of occurrence, but over-time, they may draw on adaptive skills such as using attentional strategies and reappraisals to experience to attenuate the impact of stress more so than younger individuals. This notion contrasts with previous beliefs that older adults are more susceptible to stress due to biological vulnerability. Future research should be guided by existing theories on aging and stress reactivity when considering the timeframe of events accordingly.

In the present study, however, an attempt was made to circumvent this limitation of the momentary stress variable by considering the participants past appraisal of unpleasantness as well as any change in their feelings at the time of the assessment in order to capture the

participants current experience of the stressor. However, this measure is also not without limitations. Firstly, subjective reporting is prone to response-biases, so it is possible that participants were not accurately reporting on their experience of stress. However, weak construct validity is likely more so a confounding factor on the stress-appraisal question, since there is no clear standard for how to best capture the dimension(s) of stress that drives cognitive effects. That is, since the mechanisms of how acute stress affects cognition are not clearly defined, how to best measure this likely relationship remains unclear. For example, ruminative thoughts long after the stressor could possibly affect performance, but if the mechanism driving the effects of stress on cognition is more physiological, effects will not be observed during post-stress rumination if physiological reactivity has been attenuated. Interestingly, momentary stress as measured by a simple question “How stressed do you feel right now” has been shown to be coupled with ambulatory blood pressure. (Bowman et al., 2014). However, the experience of stress is complex and dynamic, and the variety of stress impacts (e.g., cognitive, and physiological) do not necessarily occur together.

It is also worth noting that although both WP indicators of stress (MS-Subjective-WP, and PSS-Day) reflected perceived stress, these variables capture different facets of stress. Mainly, MS-Subjective-WP considered stress severity, or feelings of stress in the event that a stressor occurred, where PSS-Day captures overall, global feelings of stress, such as unpredictability and lack of control. As such, they are subject to different processes that may mediate their relationship to cognition. For example, stress-related rumination and interference may be more likely to occur in the event of a stressor, while overall feelings of global stress may be more in keeping with feelings of helplessness and subsequent depression over time. What is interesting is that feelings of chronic stress, such as ongoing thwarted control and autonomy,

have been shown to mediate the effects of stress reactivity for occurrences in daily life (Scott, Sliwinski, & Fields, 2013). It would be worthwhile to examine this interaction in future EMA research.

As mentioned, the state of knowledge on acute stressors largely extrapolates from the laboratory setting, where artificial physiological manipulation and in-lab stress tasks are mainly used in research. The present study used EMA to examine naturally occurring stress, outside of the research setting in participants day-to-day life. Though this approach holds promise for advancing the field of stress research, gold standard measurement tools do not exist (Smets, De Raedt, & Van Hoof, 2018). Within this body of research, stress is largely measured by way of self-report, though some EMA studies have recorded objective data to quantify stress, such as blood pressure. For example, Kamarck et al. (2002) found ambulatory blood pressure measured within an EMA study to be associated with time-varying psychosocial factors WP, including negative affect, arousal, task demand, decisional control, and social conflict (which are all tangential to, if not direct indicators of stress) in a sample of 340 older adults. EMA studies using ambulatory blood pressure are growing, though the development of wearable devices to record a variety of physiological data in naturalistic settings is required to more accurately capture the stress experience and advance this line of research (Yoshiuchi, Yamamoto, & Akabayashi, 2008). Projects with this aim are underway; for example, devices measuring skin conductance, respiration, heart rate, and electromyography have been developed (Choi, Ahmed, and Gutierrez-Osuna, 2011), but uptake into clinical research is lagging.

Overall, the current literature remains inconclusive about the effects of acute stress on cognition (relative to the well-established adverse effects of chronic stress); It is likely that stress influences cognition in a variety of ways, (i.e., negatively, and positively) through a myriad of

avenues, and that these associations are moderated by a number of individual differences. Though examining stress within the laboratory has brought some clarity to the nuanced circumstances influencing how stress may affect cognition differently, there is a dearth of research within the naturalistic setting examining the sequela of stress and cognition and exploring possible mediators of this potential relationship. In keeping with findings from the present study, the literature more consistently shows that pervasive feelings of global stress hinder cognitive performance and have deleterious effects on aging, if chronic.

References

- Arnsten, A. F. (2009). Stress signalling pathways that impair prefrontal cortex structure and function. *Nature reviews neuroscience*, *10*(6), 410-422.
- Beckwith, B. E., Petros, T. V., Scaglione, C., & Nelson, J. (1986). Dose-dependent effects of hydrocortisone on memory in human males. *Physiology & behavior*, *36*(2), 283-286.
- Bowen, K. S., Uchino, B. N., Birmingham, W., Carlisle, M., Smith, T. W., & Light, K. C. (2014). The stress-buffering effects of functional social support on ambulatory blood pressure. *Health Psychology*, *33*(11), 1440.
- Brewster, P. W., Rush, J., Ozen, L., Vendittelli, R., & Hofer, S. M. (2021). Feasibility and psychometric integrity of mobile phone-based intensive measurement of cognition in older adults. *Experimental aging research*, *47*(4), 303-321.
- Calvo, M. G., & Gutiérrez-García, A. (2016). Cognition and stress. In *Stress: Concepts, cognition, emotion, and behavior* (pp. 139-144). Academic Press.
- Carstensen, L. L., Fung, H. H., & Charles, S. T. (2003). Socioemotional selectivity theory and the regulation of emotion in the second half of life. *Motivation and emotion*, *27*(2), 103-123.
- Chaouloff, F., & Groc, L. (2011). Temporal modulation of hippocampal excitatory transmission by corticosteroids and stress. *Frontiers in neuroendocrinology*, *32*(1), 25-42.
- Charles, S. T. (2010). Strength and vulnerability integration: a model of emotional well-being across adulthood. *Psychological bulletin*, *136*(6), 1068.
- Choi, J., Ahmed, B., & Gutierrez-Osuna, R. (2011). Development and evaluation of an ambulatory stress monitor based on wearable sensors. *IEEE transactions on information technology in biomedicine*, *16*(2), 279-286.

- Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of health and social behavior*, 385-396.
- Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of experimental psychology: General*, 108(3), 356.
- Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1992). Intraindividual change in text recall of the elderly. *Brain and Language*, 42(3), 248-269.
- Hjortskov, N., Garde, A. H., Ørbæk, P., & Hansen, Å. M. (2004). Evaluation of salivary cortisol as a biomarker of self-reported mental stress in field studies. *Stress and Health: Journal of the International Society for the Investigation of Stress*, 20(2), 91-98.
- Hoffman, L. (2015). *Longitudinal analysis: Modeling within-person fluctuation and change*. Routledge: Taylor and Francis Group, New York and London.
- Joseph, N. T., Jiang, Y., & Zilioli, S. (2021). Momentary emotions and salivary cortisol: A systematic review and meta-analysis of ecological momentary assessment studies. *Neuroscience & Biobehavioral Reviews*, 125, 365-379.
- Kamarck, T. W., Janicki, D. L., Shiffman, S., Polk, D. E., Muldoon, M. F., Liebenauer, L. L., & Schwartz, J. E. (2002). Psychosocial demands and ambulatory blood pressure: a field assessment approach. *Physiology & Behavior*, 77(4-5), 699-704.
- Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Englewood Cliffs, NJ: Prentice-Hall.
- Klein, K., & Boals, A. (2001). The relationship of life event stress and working memory capacity. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 15(5), 565-579.

- Leung, A. W., Barrett, L. M., Butterworth, D., Werther, K., Dawson, D. R., & Brintnell, E. S. (2016). Neural plastic effects of working memory training influenced by self-perceived stress in stroke: a case illustration. *Frontiers in psychology, 7*, 1266.
- Li, S. C., Aggen, S. H., Nesselroade, J. R., & Baltes, P. B. (2001). Short-term fluctuations in elderly people's sensorimotor functioning predict text and spatial memory performance: The MacArthur Successful Aging Studies. *Gerontology, 47*(2), 100-116.
- Liu, D., Zhang, Q., Gu, J., Wang, X., Xie, K., Xian, X., ... & Wang, Z. (2014). Resveratrol prevents impaired cognition induced by chronic unpredictable mild stress in rats. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 49*, 21-29.
- Lupien, S., Lecours, A., Schwartz, G., Sharma, S., Hauger, R. L., Meaney, M. J., & Nair, N. P. V. (1996). Longitudinal study of basal cortisol levels in healthy elderly subjects: evidence for subgroups. *Neurobiology of Aging, 17*(1), 95-105.
- Lupien, S. J., Maheu, F., Tu, M., Fiocco, A., & Schramek, T. E. (2007). The effects of stress and stress hormones on human cognition: Implications for the field of brain and cognition. *Brain and cognition, 65*(3), 209-237.
- McEwen, B. S. (1998). Stress, adaptation, and disease: Allostasis and allostatic load. *Annals of the New York Academy of Sciences, 840*(1), 33-44.
- McEwen, B. S. (2002). Sex, stress and the hippocampus: allostasis, allostatic load and the aging process. *Neurobiology of aging, 23*(5), 921-939.
- McEwen, B. S., & Morrison, J. H. (2013). The brain on stress: vulnerability and plasticity of the prefrontal cortex over the life course. *Neuron, 79*(1), 16-29.
- McEwen, B. S., & Stellar, E. (1993). Stress and the individual: mechanisms leading to disease. *Archives of internal medicine, 153*(18), 2093-2101.

- Nation, D. A., Hong, S., Jak, A. J., Delano-Wood, L., Mills, P. J., Bondi, M. W., & Dimsdale, J. E. (2011). Stress, exercise, and Alzheimer's disease: A neurovascular pathway. *Medical hypotheses, 76*(6), 847-854.
- Neupert, S. D., Almeida, D. M., Mroczek, D. K., & Spiro III, A. (2006). Daily stressors and memory failures in a naturalistic setting: Findings from the va normative aging study. *Psychology and aging, 21*(2), 424.
- Newcomer, J. W., Selke, G., Melson, A. K., Hershey, T., Craft, S., Richards, K., & Alderson, A. L. (1999). Decreased memory performance in healthy humans induced by stress-level cortisol treatment. *Archives of general psychiatry, 56*(6), 527-533.
- Piazza, J. R., Charles, S. T., Sliwinski, M. J., Mogle, J., & Almeida, D. M. (2013). Affective reactivity to daily stressors and long-term risk of reporting a chronic physical health condition. *Annals of Behavioral Medicine, 45*(1), 110-120.
- Porcelli, A. J., Cruz, D., Wenberg, K., Patterson, M. D., Biswal, B. B., & Rypma, B. (2008). The effects of acute stress on human prefrontal working memory systems. *Physiology & behavior, 95*(3), 282-289.
- Qi, M., Gao, H., Guan, L., Liu, G., & Yang, J. (2016). Subjective stress, salivary cortisol, and electrophysiological responses to psychological stress. *Frontiers in psychology, 7*, 229.
- Rutter, L. A., Vahia, I. V., Passell, E., Forester, B. P., & Germine, L. (2021). The role of intraindividual cognitive variability in posttraumatic stress syndromes and cognitive aging: A literature search and proposed research agenda. *International Psychogeriatrics, 33*(7), 677-687.
- Salthouse, T. A., & Berish, D. E. (2005). Correlates of within-person (across-occasion) variability in reaction time. *Neuropsychology, 19*(1), 77.

- Sandi, C. (2013). Stress and cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(3), 245-261.
- Sapolsky, R. M., Krey, L. C., & McEwan, B. S. (1986). The neuroendocrinology of stress and aging: the glucocorticoid cascade hypothesis. *Endocrine reviews*, 7(3), 284-301.
- Scott, S. B., Sliwinski, M. J., & Blanchard-Fields, F. (2013). Age differences in emotional responses to daily stress: the role of timing, severity, and global perceived stress. *Psychology and aging*, 28(4), 1076.
- Seeman, T. E., McEwen, B. S., Singer, B. H., Albert, M. S., & Rowe, J. W. (1997). Increase in urinary cortisol excretion and memory declines: MacArthur Studies of successful aging 1. *The Journal of Clinical Endocrinology & Metabolism*, 82(8), 2458-2465.
- Shields, G. S. (2019). Stress and Cognition: A User's Guide to Designing and Interpreting Studies. *Psychoneuroendocrinology*, 104475.
- 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.
- 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.
- Smets, E., De Raedt, W., & Van Hoof, C. (2018). Into the wild: the challenges of physiological stress detection in laboratory and ambulatory settings. *IEEE journal of biomedical and health informatics*, 23(2), 463-473.
- Starcke, K., Wiesen, C., Trotzke, P., & Brand, M. (2016). Effects of acute laboratory stress on executive functions. *Frontiers in Psychology*, 7, 461.
- Stawski, R. S., Sliwinski, M. J., & Smyth, J. M. (2006). Stress-related cognitive interference predicts cognitive function in old age. *Psychology and aging*, 21(3), 535.

- Traustadóttir, T., Bosch, P. R., & Matt, K. S. (2005). The HPA axis response to stress in women: effects of aging and fitness. *Psychoneuroendocrinology*, *30*(4), 392-402.
- Tsatsoulis, A., & Fountoulakis, S. (2006). The protective role of exercise on stress system dysregulation and comorbidities. *Annals of the New York Academy of Sciences*, *1083*(1), 196-213.
- Wu, J., & Yan, J. (2017). Stress and cognition. *Frontiers in psychology*, *8*, 970.
- Yoshiuchi, K., Yamamoto, Y., & Akabayashi, A. (2008). Application of ecological momentary assessment in stress-related diseases. *BioPsychoSocial Medicine*, *2*(1), 1-6.
- Zsoldos, E., Filippini, N., Mahmood, A., Mackay, C. E., Singh-Manoux, A., Kivimäki, M., ... & Ebmeier, K. P. (2018). Allostatic load as a predictor of grey matter volume and white matter integrity in old age: The Whitehall II MRI study. *Scientific reports*, *8*(1), 1-11.

Chapter 4
Future Direction of mHealth Cognitive Assessments

Abstract

This dissertation contributes to a growing body of research on mobile health (mHealth) cognitive assessments, replicating and expanding on previous research using a consistent battery. The reliability of three brief cognitive measures (Dot memory, Stroop task, and Symbol Search, referred to as DM, ST, and SS respectively) was established in a sample of 66 healthy older adults using ecological momentary assessment (EMA) of cognitive abilities four times per day for 14 days. Specifically, both between person (BP) reliability (i.e., sensitivity in detecting interindividual differences) and within-person (WP) reliability (i.e., sensitivity in detecting intraindividual variability) of this battery was examined. Practice effects were modeled across sessions and days of the study, showing initial gains with an eventual plateau in performance for both timescales. Moreover, psychosocial factors (stress and physical activity) were modeled as predictors of BP cognitive differences, with one cognitive test (SS) showing significantly slower performance for higher levels of perceived daily stress. Novel to this EMA study, psychosocial factors were also included as predictors of WP variability in cognitive performance across sessions and days. I hypothesized that both stress and physical activity (PA) would account for short-term fluctuations in cognitive scores, however, no significant WP findings emerged and limitations in detecting these effects were discussed in chapters two and three. The future of cognitive mHealth assessments is promising for clinical research and applications. This chapter will review the current state of EMA as it pertains to mHealth cognitive assessments, and even more specifically to older adult populations at risk for dementia. Future directions, including expansion into large-scale research, clinical uptake, and examining WP psychosocial/contextual coupling will be explored.

Introduction

Considering the aging Canadian population, early detection of dementia has become paramount (Shah et al., 2016). This necessitates innovative tools for population surveillance and patient monitoring, which has historically been achieved through gold-standard in-person cognitive assessment methods that are not without limitations. Traditionally, neuropsychological assessment only occurs once the individual has demonstrated a degree of cognitive change. However, it is now generally understood that early, preclinical cognitive indicators are present years prior to objective changes (Karr, Graham, Hofer, & Muniz-Terrera, 2018). This issue is in part, related to long waitlists and cost of in-person private neuropsychological assessment. However, even if individuals are seen prior to appreciable decline, the changes that occur early in dementia are often subtle, and traditional neuropsychological testing is not sensitive enough to capture them, especially without a baseline assessment. Considering these challenges, technology-driven cognitive assessments has become of great interest given its potential to identify individuals in the early stages of dementia, who are often missed in the healthcare system (Sabbagh et al., 2020).

MHealth assessment allows for more frequent measurement occasions compared to traditional approaches, enhancing sensitivity to early changes. The accessibility and convenience of mHealth assessments may further improve care by empowering patients and extending clinical reach to more diverse and/or underserved population. Additional clinical benefits include reduced or potentially clinically useful practice effects, as these can be accounted for in longitudinal modeling (Sliwinski, Hoffman, & Hofer, 2010). Through more precise and sensitive assessment, at-risk individuals may be streamlined for earlier intervention. In addition to vast clinical potential, mHealth assessments are also remarkably fit for large-scale research

(Brewster, Rush, Ozen, Vendittelli, & Hofer, 2021), holding exceptional promise for advancements in the field of neuropsychology through their potential to elucidate the temporal sequela of cognitive processes in day-to-day life, mediating/moderating prognostic factors, and intra-individual differences for tailored intervention. Research in this regard, however, is limited to date.

Though rare, EMA studies on cognitive abilities are emerging in part due to advances in technology which have allowed for the development of mobile-based assessment tools that are brief and can be administered frequently (Sliwinski et al., 2018). Put briefly, EMA “involves repeated sampling of subjects’ current behaviours and experiences in real time, in subjects’ natural environments. EMA aims to minimize recall bias, maximize ecological validity, and allow study of microprocesses...” (Shiffman, Stone, & Hufford, 2008, p.1). Flexibility in EMA measurement schedules permit intensive longitudinal research that also minimizes participant burden and allows for researchers to examine processes across various timeframes, from minutes to years. In this way, EMA research has been instrumental in health and psychological research, which aims to answer questions about processes and behaviours that unfold in the “real world.” For example, EMA has been used to better understand the relationship between psychosocial factors and behaviours such as mental health and social contact during the COVID-19 pandemic (Huckins et al., 2020), affect and binge eating behaviours (Schaefer et al., 2020), and motivators for substance use in daily life (Votaw & Witkiewitz, 2021). However, few studies have incorporated mHealth *cognitive* assessments into EMA, despite the potential for these types of at-home assessments to revolutionize the way in which we monitor and assess individuals at risk for dementia, and overcome the barriers present in traditional in-person screening approaches (Sabbagh et al., 2020).

In their review of self-administered and repeated mHealth cognitive assessments, Moore, Swendsen, and Depp (2016) identified only 12 studies that fit their inclusion criteria, with individuals aged 14 to 83 years. More frequently, studies examine mobile/computerized measures of cognition at only one time-point, and/or under the supervision of a trained clinician/researcher rather than autonomously. That is, though computerized testing is not necessarily novel, incorporating it into EMA protocols is. In Koo and Vizer (2019) review of studies on *computerized* (i.e., not just mobile phones) cognitive assessments in older adult populations, 29 studies were identified with only three using repeated measurement design assessing specific domains of cognitive functioning (Allard et al., 2014; Jongstra et al., 2017) or proxies for a cognitive ability (Lange & Süß, 2014).

Additional studies on repeated measures mHealth cognitive assessments with adult and older adult populations have been published since both reviews (Brewster, Rush, Ozen, Vendittelli, & Hofer, 2021; Lancaster et al., 2019; Lancaster et al., 2020; Nicosia et al., 2022; Schweitzer et al., 2017; Sliwinski et al., 2018; Thompson et al., 2022). Overall, the use of mHealth cognitive assessments has largely been proven feasible with older adult populations across studies, with solid compliance, adherence, and low attrition, even in impaired populations (Nicosia et al., 2022). Moreover, the extant studies demonstrate validity against standardized neuropsychological measures (Jongstra et al. 2017; Sliwinski et al., 2018; Lancaster et al., 2019), as well as BP reliability of study aggregate scores (Brewster et al., 2021; Nicosia et al., 2022; Sliwinski et al; Thompson et al., 2022). These foundational studies have been instrumental in supporting the use of mHealth cognitive assessments through demonstrating feasibility and psychometric integrity. However, this field of research is in its inception. The remainder of this

chapter will outline future directions for this line of research, including large-scale research and clinical uptake, and a focus on WP reliability and variability.

Large-Scale Research and Clinical Potential

MHealth assessments are particularly amenable to measurement burst design, which merges traditional EMA with prospective longitudinal methods through including repeated “bursts” of intensive assessments over longer periods of time, such as months to years (Nesselroade, 1991; Nestler, 2021). Compared to traditional longitudinal studies where a single assessment occurs periodically over the duration of the study, measurement burst designs replace this single “snapshot” measurement with EMA assessments instead, which typically span days or weeks. This EMA burst is then repeated over the duration of the longitudinal study at predetermined times. In this way, measurement burst design “permits statistical decomposition of short-term variation and learning effects that overlay normative aging and provide stronger bases for detecting accelerated change due to pathological processes” (Rast, MacDonald, & Hofer, 2012, p.1). In other words, measurement burst design provides the opportunity to understand cognitive variability across different timeframes, such as minutes or hours (e.g., in relation to lifestyle factors as discussed in chapter two and three, or practice effects as discussed in chapter one), weeks or months (e.g., due to changes in health status or life circumstances), and years (e.g., long term changes associated with neurodegenerative processes or normative aging).

Traditional longitudinal approaches largely neglect intraindividual variability (Nesselroade, 1991) by assuming that a single measurement can adequately summarize the individual (Sliwinski et al., 2008). This issue is problematic when examining cognition, as research highlights the dynamic nature of cognitive processes across various time scales. For example, considerable WP variability in cognitive performance has been observed from trial-to-

trial on speeded executive control tasks (Nesselroade & Salthouse, 2004; West, Murphy, Armilio, Craik, & Stuss, 2022), and though fewer studies have looked at session-to-session variability, findings show similar yet attenuated results (Hultsch, Strauss, Hunter, & MacDonald, 2011). Short-term individual fluctuations have been demonstrated to be as high as 22 times greater than annual differences expected BP (Salthouse & Nesselroade, 2010). This WP variability calls into question the utility of single time point assessments, as they represent the individuals' abilities at a particular moment in time. However, through aggregating individual scores within a burst and examining changes in the average over time, reliability is enhanced (Brewster et al., 2021; Sliwinski, 2008; Sliwinski et al., 2018) through a closer approximation of an individual's "true score". In this way, more accurate modeling of individual trajectories may be accomplished.

Interestingly, WP variability in and of itself may also be a telling prognostic indicator. That is, changes in intraindividual variance may be an antecedent to developmental change (Hultsch, et al., 2011), with studies demonstrating a relationship between greater WP variability and central nervous system dysfunction such as in neurological conditions like Parkinsons Disease (Burton, Strauss, Hultsch, Moll, & Hunter, 2006) and dementia (Haynes, Bauermeister, & Bunce, 2017). For example, a recent meta-analysis of 13 longitudinal studies demonstrated WP variability both within (trial to trial) and across cognitive tasks of the same domain to be related to subsequent cognitive decline and/or conversion to MCI/dementia (Mumme et al., 2021). Notably, the studies in this review used standardized, in-person measures of cognitive abilities, rather than mobile assessments. These findings support the notion that WP variability may be related to processing efficiency and provide useful information about cognitive functioning beyond what a single or averaged score is able to (Hultsch et al., 2011), and provide

rational for using mHealth cognitive assessments in longitudinal measurement bursts design, as this approach is most amenable to modeling and monitoring *changes in variability*, which may be an early indicator of central nervous system dysfunction.

Considering the clinical relevance of intraindividual variability, mHealth assessments hold the potential to advance the field of neuropsychology if integrated into primary and secondary care through more frequent and precise measurements of change, that not only minimizes the confounding influence of WP variability through aggregation, but also accounts for it as a prognostic factor. Notably, the goal of ambulatory testing is not to replace current neuropsychological evaluation; rather, it may enable large-scale cognitive screening and the optimization of cognitively oriented care that is more accurate and comprehensive (Sabbagh et al., 2020). There are numerous frameworks which can be imagined for how these tools can be integrated into primary care in the future. Sabbagh et al., discuss an “ideal” approach, where score reports with baseline comparisons are sent directly to primary care providers to promote discussion about cognitive health during routine care. However, this field of research is in its early stages and extensive research is essential for eventual clinical uptake.

As Sabbagh et al. (2020) discuss in their expert consensus perspective on the future of at-home mHealth cognitive assessments, barriers to clinical uptake remain, and include the lack of robust validation studies and data needed to generate confidence in these tools (which in turn, hinders approval by regulatory bodies), privacy concerns and data security, and variability in testing conditions related to both the environment and the devices (e.g., type of smart device, operating system and versions, and software updates). Additionally, some crucial questions stand-out, and include determining the optimal measurement schedule for specific purposes (e.g., dementia monitoring), best practices for measuring and controlling for practice-effects and

establishing a normative database, as well as cross-cultural applicability/norms. For these tests to be useful clinically, they must first be deemed acceptable and valid for use in populations most at risk, such as ethnic minorities who may present with elevated rates of dementia and often nuances in the progression of the disease course (Brewster et al., 2019). Cross-cultural validation and applicability of westernized cognitive testing is a limitation within the broader field of neuropsychology (for example, applying western cognitive constructs cross-culturally), and only exacerbated within the field of mHealth assessments due to disparities in socioeconomic status and technology use.

WP Variability and Psychosocial/Contextual Factors

As discussed above, intraindividual variability in cognitive performance can be useful as a prognostic indicator, and its confounding effects on accurately capturing an individual's "true" abilities can be minimized through aggregation. However, time-varying factors that systematically account for variability in performance can also be parsed out and understood. For example, Allard et al. (2014) found that momentary reports of engagement in intellectually stimulating activities were prospectively associated with increased semantic memory performance, and Riediger et al. (2014) showed self-reported arousal to predict working-memory performance. Though WP variability is measured and defined in existing EMA studies of cognitive abilities (Brewster, Rush, Ozen, Vendittelli, & Hofer; Sliwinski et al., 2018; Valdes et al. 2016), there is little published research on factors that may account for this intraindividual variability. The examination of coupled effects between stress and exercise with cognition in this study was largely novel, though significant findings did not emerge. One consideration not yet addressed for detecting possible effects in MLM is sample power. Determining power in MLM is less well examined compared to single-level designs, and requires several additional

considerations, such as determining sample size needed at each level (Scherbaum & Ferreter, 2009). However, emerging research has demonstrated that high-frequency sampling reduces sample size requirements (Dodge, Zhu, Mattek, Austin, Kornfeld, & Kaye, 2009), and simulation studies have been published providing useful guidelines for two-level models (Arend & Schäfer, 2019). Findings from Arend and Schäfer's (2019) simulation analyses suggest that this study was appropriately powered. The dearth of published research on WP associations in EMA studies between psychosocial/contextual factors and concrete measures of cognitive skills may in part be attributable to publication biases, in that the significant findings required to build a body of research are not emerging. This is surprising, considering well-established evidence of associations in cross sectional research (e.g., those discussed in chapter two and three regarding exercise and stress and their relationship to cognition), and considering that associations among psychosocial variables which may serve as proxies for cognitive skills have been repeatedly established in EMA studies, such as the relationship between mood and sleep (Coysins et al. 2011; Patapoff et al., 2022), with both showing extensive evidence for their effect on thinking cross sectionally (Rock, Roiser, Riedel, & Blackwell, 2014; Wickens, Hutchins, Lauz, & Sebok, 2015). Rather than concluding that associations do not exist WP, it is more likely that the novelty of this field of research necessitates refinement in study design for effects to emerge.

Given the nascent nature of the mHealth field, clear, gold-standard operational definitions for time-varying psychosocial factors within this context do not yet exist. For instance, as discussed in chapter two, stress is operationally defined within the literature using subjective reports (of objective stressors, and/or subjective feelings of stress), as well as through physiological markers. However, the best way to operationalize and capture the definition of "stress" which is associated with enhanced or reduced cognitive abilities in day-to-day life is

unclear. Complicating matters further, studies examining the association between biological indicators and subjective reports of stress do not always show convergence. For example, in a review on the association between the salivary cortisol response and self-reported measures of stress, results varied widely from a positive association, negative association, or no association (Hjortskov, Garde, Ørbæk, & Hansen, 2004). On the contrary, a more recent meta-analysis found a significant positive association between momentary negative emotions and cortisol ($r = .06, p < .001$) and a significant negative association between momentary positive emotions and cortisol (Joseph, Jiang, & Zilioli, 2021). As discussed in Chapter three, the use of ambulatory blood pressure in EMA stress research is growing, and ambulatory devices aimed at recording a variety of physiological stress data are being explored (Choi, Ahmed, & Gutierrez-Osuna, 2011; Yoshiuchi, Yamamoto, & Akabayashi, 2008). It will be interesting to see this line of research advance, and potentially enhance the sensitivity of mHealth cognitive assessments to momentary experiences of stress using these novel physiological definitions.

In the same vein, PA is often self-reported or objectively measured using accelerometer data. In the broader PA literature, there remains ambiguity in the field regarding the “dose” of PA required to produce optimal effects (Erickson et al., 2019); That is, the volume, duration, frequency, and intensity of activity that is required for specific populations to experience cognitive improvements. Research suggests that a variety of activities, including leisure activities (Park, Choi, Choi, Kang, & Lee, 2019) motor training (e.g., balance and coordination), and aerobic exercise programs may benefit cognition (Ludyga, Gerber, Pühse, Looser, & Kamijo, 2020; Netz, 2019; van Uffelen, Paw, Hopman-Rock, & van Mechelen, 2008). However, even less is known about the “type” of PA associated with transient improvements captured on a shorter-time scale through EMA. Determining how to best operationally define time-varying

factors in a way that is suitable to EMA of cognition extends beyond just stress and PA and will be an important consideration in future research examining WP associations with mHealth cognitive outcomes.

Additionally, there may be many robust ways of operationalizing a variable; however, a unique consideration within the field of mHealth cognitive assessment is that different operationalization may be related to cognitive outcomes on different time scales. As Sliwinski et al. (2018) points out, the cadence of the underlying time-varying factor should be considered when determining how often and how frequently assessments are to be conducted (Ram & Gerstorf, 2009; Sliwinski, Almeida, Smyth, & Stawski, 2009). Put simply, one must consider the timing between the occurrence of the time-varying predictor and the cognitive assessment and/or the type of statistical modeling, such as lagged effects models which examine how events at previous time points effect the variable being measured at a later time; ideally, this consideration should be backed by supporting research. Within the PA literature, for example, post-exercise transient effects have been observed as discussed in chapter two, so assessing cognition just after bouts of exercise may be most appropriate. In chapter two, average activity across the day was modeled as a predictor of daily cognition. It is possible that a more fine-grained approach, modeling bouts of PA just prior to cognitive testing may have emerged as significant. Compared to the PA literature, even less is known about the effects of stress on cognition, and the literature is marked with inconsistencies partly attributable to individual level idiosyncrasies. These individual level moderating/mediating factors will be important considerations moving forward.

For example, there is some research suggesting that physical fitness, rather than average levels of PA alone may better predict cognitive outcomes, and that fitness level may mediate the relationship between acute bouts of PA and transient cognitive improvements (Bherer et al.,

2021). There is also emerging research demonstrating that PA in conjunction with environmental enrichment, such as intellectually stimulating activities, results in greater cognitive benefit as PA facilitates neuroplasticity, priming the brain for new connections (Gheysen et al., 2018). The results of Gheysen et al. (2018)'s meta-analyses of forty-one studies showed that relative to control groups, combined PA and cognitive intervention showed significantly larger gains in cognition. Of these, the studies that compared combined intervention with PA only showed small but significantly greater cognitive improvement in favor of combined interventions.

Interestingly, no differences in effects were found between older adults with and without mild cognitive impairments. In the stress research, factors such as exposure to chronic stress, and stress appraisal have been supported as moderators of the stress outcomes (Kiecolt-Glaser, Renna, Shrout, & Madison, 2020). Additionally, Bowen, Uchino, and Birmingham (2014) found that momentary stress as measured by a simple question "How stress do you feel right now" was significantly related to ambulatory blood pressure, with social support emerging as a buffer mitigating these effects. Future EMA of WP associations has the potential to elucidate many of these mediating and moderating factors in regard to how day-to-day processes interact to influence cognition.

Summary and Conclusion

This dissertation examined mHealth cognitive assessments to contribute to a novel field of research which as largely established solid feasibility with older adult populations as well as psychometric integrity, with a focus on BP reliability. This dissertation expanded on the current body of research through the inclusion of an extensive examination of WP considerations, including reliability, and factors which may account for intraindividual variance of three mHealth cognitive tasks. This last chapter delineates the vast clinical and research potential of

mHealth assessments for future consideration. Clinical application includes the possibility of more frequent, brief, less burdensome, and more accessible assessment for preventative care, with the added benefit of accounting for WP variance, both as a confound of cognitive estimates, and an indicator of neurological change. Future research using mHealth assessments may serve to elucidate shorter-term cognitive processes which overlay longitudinal change associated with normative and non-normative aging, as well how psychosocial and lifestyle factors independently and interactively moderate/mediate cognitive abilities on both of these time scales. Though in its inception, the field of mHealth assessments hold great promise for the future of cognitive oriented healthcare.

References

- Arend, M. G., & Schäfer, T. (2019). Statistical power in two-level models: A tutorial based on Monte Carlo simulation. *Psychological methods*, 24(1), 1.
- Allard, M., Husky, M., Catheline, G., Pelletier, A., Dilharreguy, B., Amieva, H., ... & Swendsen, J. (2014). Mobile technologies in the early detection of cognitive decline. *PLoS One*, 9(12), e112197.
- Brewster, P., Barnes, L., Haan, M., Johnson, J. K., Manly, J. J., Nápoles, A. M., ... & Mungas, D. (2019). Progress and future challenges in aging and diversity research in the United States. *Alzheimer's & Dementia*, 15(7), 995-1003.
- Bherer, L., Langeard, A., Kaushal, N., Vrinceanu, T., Desjardins-Crépeau, L., Langlois, F., & Kramer, A. F. (2021). Physical exercise training effect and mediation through cardiorespiratory fitness on dual-task performances differ in younger–old and older–old adults. *The Journals of Gerontology: Series B*, 76(2), 219-228.
- Bowen, K. S., Uchino, B. N., Birmingham, W., Carlisle, M., Smith, T. W., & Light, K. C. (2014). The stress-buffering effects of functional social support on ambulatory blood pressure. *Health Psychology*, 33(11), 1440.
- Brewster, P., Rush, J., Ozen, L., Jacobs, D. M., Kaye, J., Nygaard, H. B., ... & Canadian Thumbs Up Study Group. (2020). Intensive measurement of cognition to support early detection of cognitive change in individuals at risk of dementia: Advancing clinical trial design and methods in a digital age. *Alzheimer's & Dementia*, 16, e042599.

- Brewster, P. W., Rush, J., Ozen, L., Vendittelli, R., & Hofer, S. M. (2021). Feasibility and psychometric integrity of mobile phone-based intensive measurement of cognition in older adults. *Experimental aging research*, *47*(4), 303-321.
- Brose, A., Lövdén, M., & Schmiedek, F. (2014). Daily fluctuations in positive affect positively co-vary with working memory performance. *Emotion*, *14*(1), 1.
- 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.
- Choi, J., Ahmed, B., & Gutierrez-Osuna, R. (2011). Development and evaluation of an ambulatory stress monitor based on wearable sensors. *IEEE transactions on information technology in biomedicine*, *16*(2), 279-286.
- Cousins, J. C., Whalen, D. J., Dahl, R. E., Forbes, E. E., Olino, T. M., Ryan, N. D., & Silk, J. S. (2011). The bidirectional association between daytime affect and nighttime sleep in youth with anxiety and depression. *Journal of pediatric psychology*, *36*(9), 969-979.
- Dodge, H. H., Zhu, J., Mattek, N. C., Austin, D., Kornfeld, J., & Kaye, J. A. (2015). Use of high-frequency in-home monitoring data may reduce sample sizes needed in clinical trials. *PloS one*, *10*(9), e0138095.
- Erickson, K. I., Hillman, C., Stillman, C. M., Ballard, R. M., Bloodgood, B., Conroy, D. E., ... & Powell, K. E. (2019). Physical activity, cognition, and brain outcomes: a review of the 2018 physical activity guidelines. *Medicine and science in sports and exercise*, *51*(6), 1242.

- Gheysen, F., Poppe, L., DeSmet, A., Swinnen, S., Cardon, G., De Bourdeaudhuij, I., ... & Fias, W. (2018). Physical activity to improve cognition in older adults: can physical activity programs enriched with cognitive challenges enhance the effects? A systematic review and meta-analysis. *International Journal of Behavioral Nutrition and Physical Activity*, *15*(1), 1-13.
- Haynes, B. I., Bauermeister, S., & Bunce, D. (2017). A systematic review of longitudinal associations between reaction time intraindividual variability and age-related cognitive decline or impairment, dementia, and mortality. *Journal of the International Neuropsychological Society*, *23*(5), 431-445.
- Hjortskov, N., Garde, A. H., Ørbæk, P., & Hansen, Å. M. (2004). Evaluation of salivary cortisol as a biomarker of self-reported mental stress in field studies. *Stress and Health: Journal of the International Society for the Investigation of Stress*, *20*(2), 91-98.
- Huckins, J. F., Wang, W., Hedlund, E., Rogers, C., Nepal, S. K., Wu, J., ... & Campbell, A. T. (2020). Mental health and behavior of college students during the early phases of the COVID-19 pandemic: Longitudinal smartphone and ecological momentary assessment study. *Journal of medical Internet research*, *22*(6), e20185.
- Hultsch, D. F., Strauss, E., Hunter, M. A., & MacDonald, S. W. (2011). Intraindividual variability, cognition, and aging. In *The handbook of aging and cognition* (pp. 497-562). Psychology Press.
- Jongstra, S., Wijsman, L. W., Cachucho, R., Hoevenaer-Blom, M. P., Mooijaart, S. P., & Richard, E. (2017). Cognitive testing in people at increased risk of dementia using a smartphone app: the iVitality proof-of-principle study. *JMIR mHealth and uHealth*, *5*(5), e68.

- Joseph, N. T., Jiang, Y., & Zilioli, S. (2021). Momentary emotions and salivary cortisol: A systematic review and meta-analysis of ecological momentary assessment studies. *Neuroscience & Biobehavioral Reviews*, *125*, 365-379.
- Karr, J. E., Graham, R. B., Hofer, S. M., & Muniz-Terrera, G. (2018). When does cognitive decline begin? A systematic review of change point studies on accelerated decline in cognitive and neurological outcomes preceding mild cognitive impairment, dementia, and death. *Psychology and aging*, *33*(2), 195.
- Kiecolt-Glaser, J. K., Renna, M. E., ShROUT, M. R., & Madison, A. A. (2020). Stress reactivity: what pushes us higher, faster, and longer—and why it matters. *Current directions in psychological science*, *29*(5), 492-498.
- Koo, B. M., & Vizer, L. M. (2019). Mobile technology for cognitive assessment of older adults: a scoping review. *Innovation in aging*, *3*(1), igy038.
- Lancaster, C., Blane, J., Chinner, A., Wolters, L., Koychev, I., & Hinds, C. (2019). The Mezurio smartphone application: Evaluating the feasibility of frequent digital cognitive assessment in the PREVENT dementia study. *MedRxiv*, 19005124.
- Lancaster, C., Koychev, I., Blane, J., Chinner, A., Wolters, L., & Hinds, C. (2020). Evaluating the feasibility of frequent cognitive assessment using the Mezurio smartphone app: observational and interview study in adults with elevated dementia risk. *JMIR mHealth and uHealth*, *8*(4), e16142.
- Lange, S., & Süß, H. M. (2014). Measuring slips and lapses when they occur—Ambulatory assessment in application to cognitive failures. *Consciousness and cognition*, *24*, 1-11.

- Ludyga, S., Gerber, M., Pühse, U., Looser, V. N., & Kamijo, K. (2020). Systematic review and meta-analysis investigating moderators of long-term effects of exercise on cognition in healthy individuals. *Nature human behaviour*, 4(6), 603-612.
- 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.
- Mumme, R., Pushpanathan, M., Donaldson, S., Weinborn, M., Rainey-Smith, S. R., Maruff, P., & Bucks, R. S. (2021). Longitudinal association of intraindividual variability with cognitive decline and dementia: A meta-analysis. *Neuropsychology*.
- Nesselroade, J. R. (1991). Interindividual differences in intraindividual change.
- Nesselroade, J. R., & Salthouse, T. A. (2004). Methodological and theoretical implications of 10. INTRAINDIVIDUAL VARIABILITY, COGNITION, AND AGING 553 intraindividual variability in perceptual-motor performance. *Journal of Gerontology: Psychological Sciences*, 59B, P49–P55.
- Nestler, S. (2021). Modeling intraindividual variability in growth with measurement burst designs. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(1), 28-39.
- Netz, Y. (2019). Is there a preferred mode of exercise for cognition enhancement in older age?— a narrative review. *Frontiers in medicine*, 6, 57.
- Nicosia, J., Aschenbrenner, A., Balota, D., Sliwinski, M., Tahan, M., Adams, S., ... & Hassenstab, J. (2022). Unsupervised high-frequency smartphone-based cognitive assessments are reliable, valid, and feasible in older adults at risk for Alzheimer disease. Manuscript under review.

- Park, S., Choi, B., Choi, C., Kang, J. M., & Lee, J. Y. (2019). Relationship between education, leisure activities, and cognitive functions in older adults. *Aging & mental health, 23*(12), 1651-1660.
- Patapoff, M., Ramsey, M., Titone, M., Kaufmann, C. N., Malhotra, A., Ancoli-Israel, S., ... & Eyller, L. T. (2022). Temporal relationships of ecological momentary mood and actigraphy-based sleep measures in bipolar disorder. *Journal of Psychiatric Research, 150*, 257-263.
- 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.
- Ram, N., Rabbitt, P., Stollery, B., & Nesselroade, J. R. (2005). Cognitive performance inconsistency: intraindividual change and variability. *Psychology and Aging, 20*(4), 623.
- Rast, P., MacDonald, S. W., & Hofer, S. M. (2012). Intensive measurement designs for research on aging. *GeroPsych: The Journal of Gerontopsychology and Geriatric Psychiatry, 25*(2), 45.
- Riediger, M., Wrzus, C., Klipker, K., Müller, V., Schmiedek, F., & Wagner, G. G. (2014). Outside of the laboratory: Associations of working-memory performance with psychological and physiological arousal vary with age. *Psychology and aging, 29*(1), 103.
- Rock, P. L., Roiser, J. P., Riedel, W. J., & Blackwell, A. (2014). Cognitive impairment in depression: a systematic review and meta-analysis. *Psychological medicine, 44*(10), 2029-2040.

- Salthouse, T. A., & Nesselroade, J. R. (2010). Dealing with short-term fluctuation in longitudinal research. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 65(6), 698-705.
- Schaefer, L. M., Smith, K. E., Anderson, L. M., Cao, L., Crosby, R. D., Engel, S. G., ... & Wonderlich, S. A. (2020). The role of affect in the maintenance of binge-eating disorder: Evidence from an ecological momentary assessment study. *Journal of abnormal psychology*, 129(4), 387.
- Scherbaum, C. A., & Ferreter, J. M. (2009). Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organizational Research Methods*, 12(2), 347-367.
- Schweitzer, P., Husky, M., Allard, M., Amieva, H., Pérès, K., Foubert-Samier, A., ... & Swendsen, J. (2017). Feasibility and validity of mobile cognitive testing in the investigation of age-related cognitive decline. *International journal of methods in psychiatric research*, 26(3), e1521.
- Shah, H., Albanese, E., Duggan, C., Rudan, I., Langa, K. M., Carrillo, M. C., ... & Dua, T. (2016). Research priorities to reduce the global burden of dementia by 2025. *The Lancet Neurology*, 15(12), 1285-1294.
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annu. Rev. Clin. Psychol.*, 4, 1-32.
- Sliwinski, M. J. (2008). Measurement-burst designs for social health research. *Social and Personality Psychology Compass*, 2(1), 245-261.

- Sliwinski, M. J., Almeida, D. M., Smyth, J., & Stawski, R. S. (2009). Intraindividual change and variability in daily stress processes: findings from two measurement-burst diary studies. *Psychology and aging, 24*(4), 828.
- Sliwinski, M., Hoffman, L., & Hofer, S. (2010). Modeling retest and aging effects in a measurement burst design.
- 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.
- 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.
- Stone, A. A., & Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavioral medicine. *Annals of behavioral medicine.*
- Thompson, L. I., Harrington, K. D., Roque, N., Strenger, J., Correia, S., Jones, R. N., ... & Sliwinski, M. J. (2022). A highly feasible, reliable, and fully remote protocol for mobile app-based cognitive assessment in cognitively healthy older adults. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 14*(1), e12283.
- Valdes, E. G., Sadeq, N. A., Harrison Bush, A. L., Morgan, D., & Andel, R. (2016). Regular cognitive self-monitoring in community-dwelling older adults using an internet-based tool. *Journal of clinical and experimental neuropsychology, 38*(9), 1026-1037.
- van Uffelen, J. G., Paw, M. J. C. A., Hopman-Rock, M., & van Mechelen, W. (2008). The effects of exercise on cognition in older adults with and without cognitive decline: a systematic review. *Clinical Journal of Sport Medicine, 18*(6), 486-500.

- Votaw, V. R., & Witkiewitz, K. (2021). Motives for substance use in daily life: A systematic review of studies using ecological momentary assessment. *Clinical Psychological Science, 9*(4), 535-562.
- West, R., Murphy, K. J., Armilio, M. L., Craik, F. I. M., & Stuss, D. T. (2002). Lapses of intention and performance variability reveal age-related increases in fluctuations of executive control. *Brain and Cognition, 49*, 402–419.
- Wickens, C. D., Hutchins, S. D., Laux, L., & Sebok, A. (2015). The impact of sleep disruption on complex cognitive tasks: a meta-analysis. *Human factors, 57*(6), 930-946.
- Williams, M. T., Lewthwaite, H., Fraysse, F., Gajewska, A., Ignatavicius, J., & Ferrar, K. (2021). Compliance with mobile ecological momentary assessment of self-reported health-related behaviors and psychological constructs in adults: systematic review and meta-analysis. *Journal of medical Internet research, 23*(3), e17023.
- Yoshiuchi, K., Yamamoto, Y., & Akabayashi, A. (2008). Application of ecological momentary assessment in stress-related diseases. *BioPsychoSocial Medicine, 2*(1), 1-6.