6-8-2022

A DIGITAL APPLICATION FOR ASSESSMENT OF NEUROCOGNITIVE DISABILITIES

Thomas H. Auriemma
Rowan University

A DIGITAL APPLICATION FOR ASSESSMENT OF NEUROCOGNITIVE DISABILITIES

by

Thomas H. Auriemma

A Thesis

Submitted to the
Department of Computer Science
College of Science and Mathematics
In partial fulfillment of the requirement
For the degree of
Master of Science in Computer Science
at
Rowan University
May 16, 2022

Thesis Chair: Ganesh Baliga, Ph.D., Professor, Department of Computer Science

Committee Members:

David J. Libon, Ph.D., Professor, Department of Geriatrics and Gerontology and Department of Psychology

Andrea Lobo, Ph.D., Professor, Department of Computer Science

Acknowledgements

I would like to express my appreciation to Dr. Baliga and Dr. Libon for their guidance and expertise in developing this project. This would have been impossible without their help and I'm forever grateful for the opportunity to work on such impactful research. I would not be the professional that I am now without them.

I would also like to thank my parents for pushing me to complete my Master's degree and my amazing partner Leo for his love and unending support.

Lastly, I'd like to acknowledge the experienced programmers who helped along the way. Thank you to Rosty for building out the core of the iOS app, and thank you to Snazz for helping me learn Unity and work past roadblocks.

Abstract

Thomas H. Auriemma
A DIGITAL APPLICATION TO ASSESS FOR NEUROCOGNITIVE DISABILITIES
2021-2022
Ganesh Baliga, Ph.D.
Master of Science in Computer Science

Neuropsychological assessments designed to identify neurodegenerative diseases such as dementia rely on pen-and-paper tests administered by a trained medical practitioner such as a neuropsychologist. When digitally administered, these assessments can provide traditional test scoring metrics as well as many previously unobtainable time-based parameters, such as inter-response latency. Recent research has demonstrated that these digital assessments are able to identify cognitive impairments earlier that their traditional counterparts. The current research features data collected on fifty-one patients using an iPad app for the Philadelphia Pointing Span Test (PPST), a test designed to measure executive abilities. These patients were also assessed using the standard Montreal Cognitive Assessment (MoCA) test which provides neuropsychological indices measuring executive, language, and memory abilities. Our analysis shows correlations between PPST recall scores and MoCA's executive assessment. Additionally, PPST latencies correlate with patient cardiovascular risk.

In summary, the PPST appears to provide an effective assessment of executive abilities and could be a parsimonious and effective test to screen for neurocognitive impairment in the context of elevated cardiovascular risk in an ambulatory primary care environment.

Table of Contents

Table of Contents (Continued)

List of Tables

Table Page

Chapter 1

Introduction

1.1 Alzheimer's Disease

Dementia is an ever-growing world-wide public health problem. From 2000 to 2019 the rate of death from Alzheimer's disease (AD), probably the most well-known dementia subtype, has experienced a 145% increase [1]. In 2021, the United States alone spent approximately $355 billion in healthcare costs caring for AD and other dementias [1]. This number includes the money spent to care for patients and the many hours that family members set aside to care for their loved ones. By 2050, this number could rise to over a trillion US dollars [1]. As the number of dementia cases rise, so too will the stress on caregivers, medical professionals, and family members of those affected.

1.2 Mild Cognitive Impairment

Mild cognitive impairment (MCI) is characterized as a prodromal or transitional state where individuals present with reduced performance on selected neuropsychological tests and may go on to develop a dementia [2]. Early detection of MCI makes it possible to slow the onset of dementia with lifestyle changes such as regular exercise and improved nutritional choices. It is also well understood that when a disease modifying agent becomes available to treat dementia such as AD, early diagnosis and the identification of neurocognitive decline such as MCI will be critical for effective treatment. Much of the prior research on MCI has focused on predicting conversion to Alzheimer's disease (AD). A variety of biomarkers including reduced medial temporal lobe/ hippocampal volume [3-6] and the presence of Aβ40 and Aβ42 proteins in plasma

and cerebral spinal fluid [7-9] have proven useful in targeting individuals at increased risk for conversion to dementia. Some researchers associate MCI with an almost automatic conversion to AD [10]; however, research suggests that not all patients classified with MCI go on to develop AD. Some MCI patients demonstrate different outcomes including the development of non-AD dementia [11].

1.3 Mild Cognitive Impairment: Subtypes

Recent research now suggests the presence of several MCI subtypes [12]. Indeed, Petersen [13, 14] describes distinct MCI groups including single-domain amnesic MCI (aMCI) where impaired performance on tests of declarative memory dominates the clinical picture; single-domain non-memory MCI where patients present with isolated impairment in some non-memory domain of neuropsychological functioning such as executive control or language; and multi-domain MCI where several domains of neuropsychological functioning may be compromised. Using person-centered statistical techniques, Delano-Wood et al., [15] and Libon et al., [16] have provided evidence for the existence of multiple MCI groups. In their research, neuropsychological test performance obtained from groups of memory clinic MCI patients was subjected to cluster analysis. In both studies three distinct groups emerged – a relatively pure memory group; patients with impaired performance on tests of executive control and processing speed; and a mixed memory/ language group. In a world with an increasingly aging population, identifying distinct prodromal dementia states is critically important and may constitute one of the most effective tools currently available to identify, predict, and ultimately treat dementia.

1.4 Mild Cognitive Impairment and Vascular Risk

Additional evidence for the prevalence and prognosis of MCI subtypes comes from several large, community-based population studies. For example, Lopez and colleagues [17, 18] found a higher prevalence of multi-domain MCI compared with aMCI. Despite its lower prevalence rate, conversion studies suggest that individuals with aMCI were almost twice as likely to convert to dementia within 2.5 years compared with individuals diagnosed with other non-amnesic forms of MCI [19]. In another study examining dementia conversion, most amnesic and dysexecutive MCI patients were ultimately diagnosed with either AD or subcortical vascular dementia, respectively [20]. In this research the single-domain dysexecutive MCI group presented with more vascular co-morbidities and signs of vascular disease on T2-weighted MRI studies. In fact, increased stroke risk factors such as hypertension and diabetes in individuals with MCI were associated with greater white matter alterations and degraded white matter integrity as measured by diffusion tensor MRI [15]. Finally, Busse et al., [21] found that an initial diagnosis of aMCI was almost inevitably associated with a pathological diagnosis of AD at autopsy. By contrast, less pathological specificity was associated with individuals who were initially characterized as presenting with single domain, non-memory MCI or multi-domain MCI. In sum, the accumulated research suggests that MCI is a very complex and highly nuanced clinical phenomenon with emerging evidence for the co-morbid effect of cardiovascular risk factors [22].

1.5 Executive Control and Working Memory

Dysexecutive impairment is one of the key neuropsychological features associated with MCI [23]. Executive control is best understood as describing a wide

range of higher-order neurocognitive behaviors that allow individuals to regulate and control their behavior to accomplish a task. Executive control allows one to complete everyday tasks like work duties, what clothes to wear, what foods to eat, and choosing the best time to go to bed. With executive control comes working memory, characterized by the ability to mentally maintain and manipulate information for the purpose of completing a task. Remembering a phone number is a good example of working memory. Working memory is effectively our short-term memory that we use to complete tasks in our everyday life.

One test paradigm commonly used to measure executive abilities is the Digit Span Test – a traditional subtest in the Wechsler corpus of intelligence tests [24, 25]. The Wechsler version of the digit span test presents two trials of numbers to repeat with increasing span length. The test proceeds until the individual fails both trials within any single span length. Digit span forward provides a measure of auditory span or the number of bits of information that can be held in mind regardless of whether the completed response is correct. Digit span backward provides a measure of working memory or the capacity for mental manipulation [26]. Each trial is scored as either correct or incorrect.

Lamar and colleagues [27, 28] described data obtained using the Backwards Digital Span Test (BDST). This test is analogous to the commonly used Wechsler Digit Span Backwards subtest. However, there are some important differences. First, blocks of 7 trials of 3, 4, 5-span are administered. Second, there is no discontinuation as all test trials are administered. Third, rather than scoring each trial as correct or incorrect, Lamar and colleagues [27, 28] calculated ANY order and SERIAL order recall. ANY order

recall is scored by awarding a point for each correct response regardless of whether the order is correct (e.g., 48972 – "27894"= 5/5 or 100% correct). SERIAL recall is scored by awarding a point only when responses are recalled in their exact serial order position (e.g., 48972 – "27894"= 3/5 or 60% correct). In prior research Lamar and colleagues [27, 28] found that patients with dementia associated with MRI subcortical vascular disease presented with worse SERIAL versus ANY order recall. Worse SERIAL order recall has also been found in memory clinic patients when patients with single domain, dysexecutive MCI were compared to other MCI groups [16]. Lamar and colleagues [28] also observed worse SERIAL recall in relation to greater left- sided MRI white matter alterations (i.e., leukoaraiosis, LA). There was no correlation between ANY ORDER performance and LA [28] suggesting that LA does not interfere with immediate digit recall but does interfere with temporal re-ordering of digits.

Emrani and colleagues [29] analyzed BDST SERIAL order performance in memory clinic patients. Patients were classified into three groups – amnestic MCI (aMCI), combined mixed/dysexecutive MCI (mixed/dys-MCI), and cognitively normal (CN). The results showed that the mixed/dys-MCI group scored lower than non-MCI patients for serial positions 3, 4 and 5. The mixed/dys-MCI group also produced a greater number of transposition errors, omissions, and preservations than other groups. This data suggests that assessing mental performance through SERIAL order using the BDST can dissociate between MCI subtypes.

Recently, Emrani et al. [30] considered latencies produced within correct trails of the BDST by individuals with MCI (N = 22) and without MCI (N = 36). Their analyses revealed that non-MCI patients produced *slower* latencies in positions 2 and 4 in the 5-

span BDST. By contrast, MCI patients were slower in their position 3 response, which suggests that the allocation of time to bring this task to fruition was different for those patients. Thus, intra-component latencies within a response can provide valuable added information for MCI screening.

1.6 Significance of the Current Research

Current neuropsychological assessment relies on traditional pen-and-paper tests. However, recent research has demonstrated the potential for the early identification of emerging neurodegenerative disease through the use of digital neuropsychological assessment platforms [30-32]. Using digital assessment technology, we can now collect time- based data such as intra-component latencies (delays between responses). With digital technology, test results can now be objectively obtained and scored using the same criteria used in the pen-and-paper tests. Digital assessment technology can collect a patient's responses using voice recognition and/ or touch screen technology. All data collected using digital technology are time stamped allowing us to know when and where decisions were made within the context of the total time necessary to complete a task. By extending previous research described by Emrani and colleagues [29-32] using digital assessment tools, we can undertake research to collect data previously unobtainable.

In the current research, middle-aged participants were recruited from an ambulatory primary medical practice. Cardiovascular risks such as treatment for hypertension, diabetes, elevated cholesterol, and cardiac illness (e.g., MI, CAD, etc.) were coded. In the current research a digitally administered pointing span test, the Philadelphia Pointing Span Test (PPST), was deployed. The PPST is designed by Dr. Ganesh R. Baliga and Dr. David J. Libon and modeled after the Backward Digit Span

Test (BDST) [30]. The goals of the current research are 2-fold. First, the concurrent validity of the PPST as a measure of executive abilities will be investigated. Our first hypothesis is that PPST executive indices measuring both auditory span (ANY ORDER recall) and mental manipulation (SERIAL ORDER recall) will be related to other neuropsychological tests that assess executive abilities. Another goal of the current research is to assess how well the PPST can screen for neuropsychological difficulties in the context of cardiovascular risk. Our second hypothesis is that selected PPST latency measures will be associated with greater cardiovascular risk.

Chapter 2

Digital Assessment

2.1 Philadelphia Pointing Span Test

As noted above, digital neuropsychological assessment was conducted using the Philadelphia Pointing Span Test (PPST). The test was administered using an iOS app running on an iPad. The PPST is similar to another well-known neurocognitive test – The Corsi Block Tapping Test [33]. This test is designed to assess visuospatial working memory and is commonly used with children, adults, and those with neuropsychological disabilities [34].

The PPST was constructed based on the Corsi Block Tapping Test as described above as well as BDST. The test contained 5 trials of 3, 4, and 5- span test trials. As with the BDST, all PPST test trials are administered with no discontinuation. The digitally administered PPST contains two subtests – (1) a number only test condition, and (2) a combined numbers/letters test condition. In the numbers only test condition, all trials consisted of a series of non-repeating Arabic numbers. In this portion of the test, participants heard a series of numbers from the iPad and were instructed to touch the answer key ordering their response from the lowest number to the highest number (i.e., 3729 – "2379"). In the combined numbers/letters test condition, participants heard a series of non-repeating numbers and letters and were instructed to first touch numbers from lowest to highest followed by letters in alphabetical order (i.e., 5T3J – "35JT").

All responses were obtained using touch screen technology. After the stimulus was delivered participants tapped numbers or letters contained within circles on the

answer key. Each time a response was made the corresponding button tracked how long the response took from the time the iPad administered the auditory test trial. The number or letter inside the button was then highlighted for a quarter of a second, indicating the selection. Within each trial, the participant must wait until the entire string of digits or combined digits and letters was spoken before a response can be initiated. Selections cannot be repeated on the same button in a single trial screen.

In order to have the iPad read out the digit span to the patient, we implemented Apple's own text-to-speech synthesizer with the list of stimuli. There are many benefits to this as opposed to recording sound files. For one, the spoken stimuli can be easily changed by adjusting the test's corresponding JSON file. We can also change the voice and accent just as one can change Siri's voice settings on an Apple device. We can even adjust the speech rate and pitch of the voice. Finally, our approach enables us to leverage the culture-neutral nature of the PPST by easily changing the language of administration.

2.2 iPad Application

Our iPad application is programmed with other neuropsychological tests allowing practitioners to switch between tests within the main menu. Before starting the test, a unique patient ID is assigned along with any notes the practitioner wants to include. There are settings on the main menu allowing for different test parameters. Each test is defined by a unique JSON file which can be changed on-the-fly, allowing for quick and easy changes. For instance, if different digits need to be read out in PPST, a simple change to the JSON file in the correct position will suffice. No other changes are necessary. This open system allowed us to build and edit tests as necessary.

After a test is completed, the results are stored into a JSON file and sent to our database. From there, researchers can open our web client, enter their credentials and password, and then access the formatted spreadsheets. The spreadsheets are formatted using Python language library OpenPyxl [35] and pre-formatted spreadsheet templates. Using OpenPyxl we can have specific cells within a spreadsheet use Excel functions that access spreadsheet columns and rows of our choice. These spreadsheets are then exported through our AWS Lambda-based backend. The AWS Lambda service [36] enables software developers to run programs in AWS' cloud without maintaining a dedicated server. Deployment to AWS Lambda was handled using Python language serverless framework Chalice [37]. For our purposes it made sense to access our web client and spreadsheets only when requested. These design choices have allowed our research to remain lean and adjustable.

2.3 Virtual Reality Application

We are also interested in exploring the value that other digital measurements might add to neuropsychological testing. In addition to collecting time-based parameters such as response latencies, a small group of VR headsets are also capable of tracking eye position (eye tracking) and providing biometric data such as pulse, thereby supporting the collection of a richer dataset as compared to our iPad-based tests. We implemented several cognitive tests, including Cancellation and Angle Matching (see Appendix for descriptions), in a desktop application environment that uses virtual reality (VR). COVID-19 dictated social distancing and remote work constraints have delayed data collection on our VR-based tests. The results and discussion of forthcoming chapters pertain to the iPad-based tests. This section describes our VR work to date.

We assumed that VR assessments would be tolerated by patients from a younger age group. During prototyping, elderly family members were asked to help test the VR application and we discovered that VR-based testing was well-tolerated with only minor adjustments to ensure comfort. There are two straps on the headset that need to be adjusted to ensure it rests comfortably on the patient's head. Patients are seated and stationary during testing so there was no issue of nausea or discomfort. For patients with glasses, we asked that they remove them in order to prevent damage to the glasses as well as the lenses on the VR headset.

2.3.1 Hardware Acquisition

The first step in the VR research was to identify a research-grade VR headset that includes biometric tracking capabilities into one package and was also available off the shelf. At the time, the options that satisfied these criteria were extremely limited.

Thankfully, Hewlett Packard (HP) launched the Reverb G2: Omnicept Edition VR headset [38] in spring 2021. The base version of this product is meant simply for VR gaming and offers high resolution lenses and speakers to deliver a high-quality immersive and comfortable VR experience. The upgraded Omnicept version augments the base version with a variety of biometric sensors that measure muscle movement, eye gaze, pupil size, and pulse (heart rate) all in one package. The eye-tracking sensors are provided by Tobii AB [39], leading innovators in eye tracking technology. The next step after acquiring the proper hardware was to design and implement an application that can run our cognitive tests in a VR environment.

2.3.2 VR Implementation

With this complete package of sensors and a computer capable of running the headset smoothly, there is the potential for conducting neuropsychological tests that capture more information than the comparable iPad assessments. Over six months of development time was dedicated exploring our options with VR testing until we ultimately decided to put it aside and focused our research efforts on iPad assessment. This was time spent learning how to use Unity, and designing a test that would work within the limits of the hardware and software in our stack. For this reason, it is important that this is mentioned as many efforts were made to make this research possible.

Some challenges we faced when designing the VR tests are discussed in Chapter 5 (Conclusions and Future Work) and within the Appendix. In the following chapters, we will detail the assessments conducted on our iPad application and analyze and interpret the data that was collected.

Chapter 3

Methods

3.1 Participants

The current research recruited 51 patients from Rowan University, Department of Family Practice. These participants were recruited for the current study as part of their routine medical care. Inclusion/ exclusion criteria included (1) English is the participants' first language, (2) major medical/neurological illness, such as cancer or epilepsy, that might be expected to alter cognitive abilities, were absent; and (3) no dementia will be present. Demographics are listed in Table 1. The sample was middle-aged (52.37$\pm$ 14.27 years); relatively well educated (14.24$\pm$ 1.68 years); and the majority of participants were female (62.70%). Gross neurocognitive abilities were assessed with the Montreal Cognitive Assessment (MoCA) [40, 41]. Scores on this test range from 0 – 30. Mean MoCA performance was 24.42$\pm$ 2.66 (Table 1).

Table 1

Demographic and MoCA Test Performance

(N = 51)	**Age**	**Education years)**	**MoCA Score**
Range	56	6	11
Minimum	23	12	18
Maximum	79	18	29
Mean	52.37	14.24	24.41
Standard Deviation	14.27	1.68	2.66

3.2 Cardiovascular Risks

Cardiovascular risks (CVR) were coded including treatment for elevated cholesterol, diabetes, heart disease, and hypertension. Each CVR was scored as 0 when not present, 1 when present and treatment is ongoing, and 2 when present and no treatment is currently provided. Cardiovascular risks were summed to create a CVR index (range 0 – 4). Participants' systolic and diastolic blood pressure and A1C value was coded when available. The A1C test, a popular test used to diagnose prediabetes and diabetes, measures the average blood sugar level in a patient over the past three months.

3.3 Statistical Analysis

Statistical analysis was performed using IBM Statistical Product and Service Solutions (SPSS) [42].

3.3.1 PPST ANY and SERIAL ORDER and MoCA Test Performance

MoCA test performance was first analyzed using principal component analysis (PCA) to extract indices assessing executive, language, and verbal memory abilities. Relations between MoCA PCA derived indices and PPST ANY order and SERIAL order recall were assessed with Pearson correlational analyses. Correlational analyses were also used to assess relations between PPST ANY order and SERIAL order recall and the CVR index defined in Section 3.2.

3.3.2 PPST Latency Measures

The latency to generate individual PPST responses from both test conditions was correlated with PCA derived MoCA indices assessing executive, language, and verbal memory abilities. Separate regression analyses were conducted to assess the relation

between the CVR index and latencies to generate individual responses on the PPST numbers only and the combined number/letter test conditions. For these analyses, age was entered into the first block. Latencies for 3, 4, and 5-span responses were entered into the second block, respectively.

Chapter 4

Results

4.1 MoCA Principal Component Analysis

Table 2 lists the MoCA parameters subjected to PCA using varimax rotation. This analysis resulted in a 3–factor solution accounting for 68.31 percent of variance (Table 2). Factor 1 (executive control) contained three variables - time to completion for the MoCA Trail Test– Part B, time to completion for subtracting serial 7s, and MoCA clock drawing errors. Factor 2 (language) was related to performance on MoCA Naming and Similarities tests. Finally, factor 3 (memory) was related to verbal memory test performance as measured by the MoCA cued recall and recognition tests.

Table 2

MoCA Principal Component Analysis

	Factor 1	**Factor 2**	**Factor 3**
Trail Making Test– (time to completion)	**0.761**	-0.337	-0.136
Subtracting Serial 7s (time to completion)	**0.719**	-0.483	-0.087
Clock Drawing Errors	**-0.740**	-0.318	0.169
Naming Test Performance	-0.006	**0.821**	0.042
Similarities Test Performance	-0.170	**0.615**	0.226
Delayed Cued Recall	-0.004	0.132	**0.890**
Delayed Recognition	-0.322	0.112	**0.760**
Variance	37.44	16.16	14.69

4.2 PPST ANY and SERIAL ORDER and MoCA Test Performance

Descriptive statistics for PPST ANY and SERIAL ORDER recall can be found in Table 3 below, which shows that the PPST was well tolerated by the patients in our sample.

Table 3

PPST Percent ANY and SERIAL ORDER Correct

PPST Number Test Condition		
	mean	standard deviation
PPST 3-span ANY ORDER	99.86	0.96
PPST 3-span SERIAL ORDER	99.31	2.83
PPST 4-span ANY ORDER	99.90	0.72
PPST 4-span SERIAL ORDER	98.96	5.26
PPST 5-span ANY ORDER	97.62	4.31
PPST 5-span SERIAL ORDER	90.64	13.17
PPST Number/ Letter Test Condition		
	mean	standard deviation
PPST 3-span ANY ORDER	99.17	2.96
PPST 3-span SERIAL ORDER	96.67	8.91
PPST 4-span ANY ORDER	97.08	4.82
PPST 4-span SERIAL ORDER	91.69	13.76
PPST 5-span ANY ORDER	87.75	9.92
PPST 5-span SERIAL ORDER	74.88	17.85

The Pearson Product Moment Correlation (i.e., ‘r’) is a bivariate statistical analysis that measures the relation between two phenomena, namely, (1) the strength of the association between two variables, and, (2) the direction of that relationship. The Pearson correlation coefficient ranges between +1.00 and -1.00. A value of +1.00 or -1.00 suggests a perfect relationship. As the correlation coefficient values moves toward zero, the relationship between the two variables grows weaker. The direction of the

relationship is indicated by the sign of the coefficient. A positive relationship suggests that as value of one variable increase, the value of the second variable also increases. A negative relationship suggests that as the value of one variable increases, the value of the second variable decreases. In the neuropsychology field, a correlation value between ±0.3-0.5 is considered promising when attempting to establish concurrent validity between an established assessment and an untested assessment.

The p-value obtained with Pearson's correlation analysis describes the significance of the correlation in as far as how well it rejects the null hypothesis, meaning that there is no relationship between the two variables. Informally stated, a sufficiently low p-value means that the null hypothesis can be rejected. In our research, we set our p-value threshold at 0.05, so any values below this threshold have a very low probability of supporting the null hypothesis. Correlations with a p-value below 0.05 are statistically significant and any values below 0.01 are even more significant.

In the number test condition, Pearson correlation analyses found a correlation between MoCA PCA derived executive index and PPST 3-span SERIAL ORDER recall ($r= -0.348$; $p< 0.016$); PPST 4-span recall ($r= -0.376$, $p< 0.009$), and 5-span SERIAL ORDER recall ($r= -0.405$, $p< 0.005$), where slower time to completion and greater number of clock drawing errors were associated with worse PPST SERIAL ORDER recall.

For the number/letter test condition, the MoCA PCA derived executive index slower time to completion and a greater number of clock drawing errors correlated with reduced PPST 4-span ANY ORDER ($r= -0.309$, $p< 0.034$); 4-span SERIAL ORDER recall ($r= -0.366$, $p< 0.012$); and 5-span SERIAL recall ($r= -0.288$, $p< 0.049$).

4.3 PPST ANY and SERIAL ORDER Recall and Cardiovascular Risk

Analyses examining blood pressure and cholesterol were not significant. Additionally, no significant results were found for the number test condition. However, for the combined number/letter test condition, elevated A1C was correlated with reduced PPST 3-span ANY ORDER recall ($r= -0.362$, $p< 0.046$) and reduced PPST 4-span ANY ORDER recall ($r= -0.392$, $p< 0.029$). For the CVR index, greater evidence for cardiovascular risk was associated with reduced PPST 3-span ANY ORDER ($r= -0.531$, $p< 0.001$) and PPST 5-span ANY ORDER recall ($r= -0.356$, $p< 0.018$).

4.4 PPST Latency Data and Neuropsychological Performance

The latency to generate all responses for the PPST number and combined number/ letter test conditions are contained in Tables 4-6. Correlation between MoCA PCA derived neuropsychological indices and 3-span and 4-span latencies in the 3 and 4-span trials blocks were not significant. Nonetheless, in the 5-span number test condition, longer or *slower* 3rd and 5th response latency was associated with *better* performance on the MoCA PCA derived memory index ($r= 0.314$, $p< 0.038$; $r= 0.401$, $p< 0.007$, respectively). None of these analyses were significant for the combined number/letter test condition.

Table 4

PPST 3-Span Latency (Mean & Standard Deviations)

PPST Number Test Condition			
	PPST 3- span 1st latency	PPST 3- span 2nd latency	PPST 3- span 3rd latency
Mean	2.16 secs	0.86 secs	0.66 secs
Std. Deviation	0.75	0.62	0.27
PPST Number/ Letter Test Condition			
	PPST 3-span 1st latency	PPST 3-span 2nd latency	PPST 3-span 3rd latency
Mean	2.55 secs	1.21 secs	0.75 secs
Std. Deviation	1.33	0.57	0.28

Table 5

PPST 4-Span Latency (Mean & Standard Deviations)

PPST Number Test Condition				
	PPST 4-span 1st latency	PPST 4-span 2nd latency	PPST 4-span 3rd latency	PPST 4-span 4th latency
Mean	1.84 secs	2.72 secs	3.73 secs	4.36 secs
Std. Deviation	1.01	1.28	1.73	1.97
PPST Number/ Letter Test Condition				
	PPST 4-span 1st latency	PPST 4-span 2nd latency	PPST 4-span 3rd latency	PPST 4-span 4th latency
Mean	2.50 secs	0.87 secs	2.20 secs	0.91 secs
Std. Deviation	0.94	0.41	0.86	1.69

Table 6

PPST 5-Span Latency (Mean & Standard Deviations)

PPST Number Test Condition					
	PPST 5-span 1st latency	PPST 5-span 2nd latency	PPST 5-span 3rd latency	PPST 5-span 4th latency	PPST 5-span 5th latency
Mean	2.13 secs	0.99 secs	1.72 secs	1.30 secs	0.56 secs
Std. Deviation	1.28	0.54	1.37	0.99	0.34
PPST Number/ Letter Test Condition					
Mean	2.99 secs	0.85 secs	2.44 secs	1.92 secs	0.96 secs
Std. Deviation	1.61	0.55	2.91	1.23	1.38

4.5 PPST Latency and Cardiovascular Risk

Tables 4-6 above show means and standard deviation for latencies obtained in patient trials for PPST 3-span, 4-span and 5-span (number and number/letter conditions) respectively. To explore the interesting patterns that were observed, a series of regression analyses were conducted examining relationships between the CVR index and PPST latencies. In these analyses, age was entered in block 1 followed by latencies within each trial block entered in block 2. Analyses looking at 3, 4, and 5- span latencies in the numbers only test condition were not significant. The analyses looking at 3 and 4- span latencies in the combined number/letter test condition were also not significant. However, the analysis of the 5-span combined number/letter test condition was significant ($R= 0.594$, $R^2= 0.353$, $df= 5/ 30$, $F= 2.74$, $p< 0.037$). Only the latency for the fifth response entered the final model ($beta= 0.519$, $p< 0.003$), suggesting that for the PPST, *slower* latency is associated with *increasing* cardiovascular risk.

Chapter 5

Conclusions and Future Work

5.1 Research Accomplishments

The purpose of the current research was to demonstrate the efficacy of using a digital platform to administer and score neuropsychological tests. As found in previous research [31, 32] and in the present study, assessing neuropsychological abilities using digital technology permits the collection of highly nuanced data that previously has not been obtainable.

In the current research, the PPST was administered to a group of patients drawn from an ambulatory family medical practice. Patients were not suspected of having any neurocognitive issues and were assessed as part of their routine medical care. Along with the PPST, patients were administered the MoCA. The current research tested two hypotheses. Our first hypothesis was that PPST indices measuring auditory span (ANY ORDER recall) and mental manipulation (SERIAL ORDER recall) will be related to other neuropsychological indices that assess executive abilities. Our second hypothesis was to assess how well the PPST can screen for neuropsychological difficulties in the context of elevated cardiovascular risk.

In the current research, selected MoCA test parameters were analyzed using a principal component analysis. This analysis resulted in the compilation of three neuropsychological indices measuring executive abilities, language, and verbal episodic memory, respectively. Consistent with our first hypothesis, PPST SERIAL ORDER recall was statistically related to the MoCA PCA executive index. As described above,

on the PPST numbers test condition, correlational analyses found reduced MoCA executive test performance associated with reduced PPST 3, 4, and 5-span SERIAL ORDER recall. When the PPST combined number/letter test condition was analyzed, reduced MoCA executive abilities was found in relation to reduced PPST 4-span ANY ORDER, as well as 4 and 5-span SERIAL ORDER recall.

These data are consistent with previous research described by Lamar et al., [27, 28] and Libon et al., [26]. In these studies, memory clinic patients with dementia and MCI were studied. Analogous ANY and SERIAL recall indices correlated with other measures of executive abilities as well as posterior and anterior MRI white matter alterations. Moreover, in previous research, Emrani et al., [29] demonstrated how SERIAL ORDER scoring can more effectively differentiate mixed/dys-MCI patients from non-MCI and aMCI patients. Taken as a whole, the current findings along with past research provide support for the concurrent validity for the PPST as means to assess executive abilities.

Past research has found that cardiovascular risk such as hypertension, cholesterol, and diabetes are highly associated with reduced performance on neuropsychological tests. De Anda-Duran [43] studied a group of community dwelling research volunteers from the Bogalusa Heart Study. These researchers found that chronic cardiovascular risk, such as diabetes and hypertension, were highly associated with reduced performance on neuropsychological tests assessing executive and information processing speed abilities, as well as greater carotid stenosis. Consistent with this research and our second hypothesis, PPST test performance was also associated with greater cardiovascular risk. Elevated A1C was associated with reduced 3 and 4-span ANY ORDER recall. A greater

score on the cardiovascular risk index was also associated with reduced 3-span and 5-span ANY ORDER recall. When individual latencies were examined from correct PPST 5-span test trials in number test condition, greater cardiovascular risk was associated with slower latency to generate the second response. Greater cardiovascular risk was also associated with slower latency to generate the fifth response in the combined number/letter test condition.

5.2 Lessons and Challenges

Multiple applications were prototyped in order to test the limits of neuropsychological assessments on digital platforms. Between the VR and iOS platforms, there were major differences that led to the use of some tests and the rejection of others. For example: the Cancellation task was difficult for VR users as the eye tracking sensors are not precise enough to differentiate between two closely-placed objects. However, this test was well-tolerated on the iPad application as replication of this test was nearly identical to the original pen-and-paper version. The most effective test for VR and eye tracking was the Angle Matching test which has 9 oblique lines on the VR display. It was simple enough for eye tracking while still maintaining a level of difficulty.

While there are others to mention, more implementation challenges can be found in the Appendix. Due to my own lack of experience in 3D game engines, several months were spent on learning the technology that would power our digital assessments on our VR platform. Many challenges resulted from technical considerations of whether a feature was possible to implement on the hardware, and also whether that feature would be well-tolerated by the patients. A test that is too easy will not assess anything about the

patient, and a test that is too difficult will not be effective for data collection. There had to be a balance of difficulty in order for effective data collection.

Every test we implemented went through several iterations before reaching its finalized version. A test's user interface must be intuitive for both patient and test administrator. For instance, one iteration of the PPST focused on allowing the backtracking of trials in situations where the test administrator felt a trial should be re-administered. In this case, our data collection would only track the data of the most recent trial response. Multiple changes also had to be made to the voice synthesizing used in our application. Patient feedback was used to arrive at a synthesized voice that was well-tolerated.

5.3 Conclusion

A major strength of the current research is the engineering of a new digitally administered test able to assess executive abilities and the use of innovative technology to obtain latency data that was previously unobtainable. Several weaknesses of our research must be acknowledged. First, our sample was modest, and the number of cardiovascular risks were also modest. Additional research with a larger sample is necessary to replicate the data reported above. Second, to validate the relationship between PPST and neuropsychological test performance, PPST ANY and SERIAL recall, and the latency measures should be validated against a comprehensive neuropsychological protocol. Despite these weaknesses, the data described suggest that the PPST could be a parsimonious and effective test to screen for neurocognitive impairment in the context of elevated cardiovascular risk in an ambulatory primary care environment. In summary, we augmented classical neuropsychology with digital administration. We successfully

built a digital test using the PPST and we were able to provide evidence for how well it can assess executive function and CV risk. Additionally, we explored the limits of neuropsychological assessments in a VR setting and left a solid foundation for future research.

5.4 Future Work

This research is not complete in the sense that additional testing and evaluation needs to be performed to ensure that our assessments are in line with their traditional counterparts. Our limited population leaves an opportunity for more data collection for a wider population.

There is also the opportunity for data collection with our VR assessments. Constraints related to COVID-19 restrictions and delays caused by supply disruptions forced us to reel back some of our original research goals. The exploration of VR-based neuropsychological assessments is likely to continue. Based on our limited experience with building VR-based assessments, we posit that that these assessments are promising due to the rich amount of data that can be collected. In the future, we anticipate that more Computer Science researchers will be intrigued by the potential of this research and what it could mean for the well-being of future generations.

References

[1] Alzheimer's Association. *2021 Alzheimer's Disease Facts and figures*. Retrieved January 25, 2022, from https://www.alz.org/media/Documents/alzheimers-facts-and-figures.pdf

[2] Petersen, R. C., Doody, R., Kurz, A., Mohs, R. C., Morris, J. C., Rabins, P. V., Ritchie, K., Rossor, M., Thal, L., & Winblad, B. (2001). Current concepts in mild cognitive impairment. *Archives of Neurology*, 58(12), 1985. https://doi.org/10.1001/archneur.58.12.1985

[3] Apostolova, L. G. (2016). Alzheimer disease. *CONTINUUM: Lifelong Learning in Neurology*, 22(2, Dementia), 419–434. https://doi.org/10.1212/con.0000000000000307

[4] DeCarli, C. (2007). Qualitative estimates of medial temporal atrophy as a predictor of progression from mild cognitive impairment to dementia. *Archives of Neurology*, 64(1), 108. https://doi.org/10.1001/archneur.64.1.108

[5] Jack, C. R., Petersen, R. C., Xu, Y. C., O'Brien, P. C., Smith, G. E., Ivnik, R. J., Boeve, B. F., Waring, S. C., Tangalos, E. G., & Kokmen, E. (1999). Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment. *Neurology*, 52(7), 1397–1397. https://doi.org/10.1212/wnl.52.7.1397

[6] Tapiola, T., Pennanen, C., Tapiola, M., Tervo, S., Kivipelto, M., Hänninen, T., Pihlajamäki, M., Laakso, M. P., Hallikainen, M., Hämäläinen, A., Vanhanen, M., Helkala, E.-L., Vanninen, R., Nissinen, A., Rossi, R., Frisoni, G. B., & Soininen, H. (2008). MRI of hippocampus and entorhinal cortex in mild cognitive impairment: A follow-up study. *Neurobiology of Aging*, 29(1), 31–38. https://doi.org/10.1016/j.neurobiolaging.2006.09.007

[7] Fagan, A. M., Roe, C. M., Xiong, C., Mintun, M. A., Morris, J. C., & Holtzman, D. M. (2007). Cerebrospinal fluid tau/β-amyloid42 ratio as a prediction of cognitive decline in nondemented older adults. *Archives of Neurology*, 64(3), 343. https://doi.org/10.1001/archneur.64.3.noc60123

[8] Nordlund, A., Rolstad, S., Klang, O., Lind, K., Pedersen, M., Blennow, K., Edman, Å., Hansen, S., & Walin, A. (2008). Episodic memory and speed/attention deficits are associated with Alzheimer-typical CSF abnormalities in MCI. *Journal of the International Neuropsychological Society*, 14(4), 582–590. https://doi.org/10.1017/s135561770808079x

[9] Simonsen, A. H., McGuire, J., Hansson, O., Zetterberg, H., Podust, V. N., Davies, H. A., Waldemar, G., Minthon, L., & Blennow, K. (2007). Novel panel of cerebrospinal fluid biomarkers for the prediction of progression to Alzheimer dementia in patients with mild cognitive impairment. *Archives of Neurology*, 64(3), 366. https://doi.org/10.1001/archneur.64.3.366

[10] Morris, J. C., Storandt, M., Miller, J. P., McKeel, D. W., Price, J. L., Rubin, E. H., & Berg, L. (2001). Mild cognitive impairment represents early-stage Alzheimer disease. *Archives of Neurology*, 58(3). https://doi.org/10.1001/archneur.58.3.397

[11] Griffith, H. R., Netson, K. L., Harrell, L. E., Zamrini, E. Y., Brockington, J. C., & Marson, D. C. (2006). Amnestic mild cognitive impairment: Diagnostic Outcomes and clinical prediction over a two-year time period. *Journal of the International Neuropsychological Society*, 12(2), 166–175. https://doi.org/10.1017/s1355617706060267

[12] Ritchie, K., & Touchon, J. (2000). The heterogeneity of mild cognitive impairment: A neuropsychological analysis. *Lancet*, 335, 225-228.

[13] Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, 256(3), 183–194. https://doi.org/10.1111/j.1365-2796.2004.01388.x

[14] Petersen, R. C., & Morris, J. C. (2005). Mild cognitive impairment as a clinical entity and treatment target. *Archives of Neurology*, 62(7), 1160. https://doi.org/10.1001/archneur.62.7.1160

[15] Delano-Wood, L., Bondi, M. W., Sacco, J., Abeles, N., Jak, A. J., Libon, D. J., & Bozoki, A. (2009). Heterogeneity in mild cognitive impairment: Differences in neuropsychological profile and associated white matter lesion pathology. *Journal of the International Neuropsychological Society*, 15(6), 906–914. https://doi.org/10.1017/s1355617709990257

[16] Libon, D. J., Xie, S. X., Eppig, J., Wicas, G., Lamar, M., Lippa, C., Bettcher, B. M., Price, C. C., Giovannetti, T., Swenson, R., & Wambach, D. M. (2009). The heterogeneity of mild cognitive impairment: A neuropsychological analysis. *Journal of the International Neuropsychological Society*, 16(1), 84–93. https://doi.org/10.1017/s1355617709990993

[17] Lopez, O. L., Jagust, W. J., DeKosky, S. T., Becker, J. T., Fitzpatrick, A., Dulberg, C., & Kuller, L. H. (2003). Prevalence and classification of mild cognitive impairment in the Cardiovascular Health Study Cognition Study: part 1. *Archives of neurology*, 60(10), 1385-1389.

[18] Lopez, O. L., Kuller, L. H., Becker, J. T., Dulberg, C., Sweet, R. A., Gach, H. M., & DeKosky, S. T. (2007). Incidence of dementia in mild cognitive impairment in the cardiovascular health study cognition study. *Archives of Neurology*, 64(3), 416-420.

[19] Fischer, P., Jungwirth, S., Zehetmayer, S., Weissgram, S., Hoenigschnabl, S., Gelpi, E., Krampla, W., & Tragl, K. H. (2007). Conversion from subtypes of mild cognitive impairment to Alzheimer dementia. *Neurology*, 68(4), 288–291. https://doi.org/10.1212/01.wnl.0000252358.03285.9d

[20] Zanetti, M., Ballabio, C., Abbate, C., Cutaia, C., Vergani, C., & Bergamaschini, L. (2006). Mild cognitive impairment subtypes and vascular dementia in community-dwelling elderly people: A 3-year follow-up study. *Journal of the American Geriatrics Society*, 54(4), 580–586. https://doi.org/10.1111/j.1532-5415.2006.00658.x

[21] Busse, A., Hensel, A., Guhne, U., Angermeyer, M. C., & Riedel-Heller, S. G. (2006). Mild cognitive impairment: Long-term course of four clinical subtypes. *Neurology*, 67(12), 2176–2185. https://doi.org/10.1212/01.wnl.0000249117.23318.e1

[22] DeCarli, C. (2003). Mild cognitive impairment: Prevalence, prognosis, aetiology, and treatment. *The Lancet Neurology*, 2(1), 15–21. https://doi.org/10.1016/s1474-4422(03)00262-x

[23] Eppig, J., Wambach, D., Nieves, C., Price, C. C., Lamar, M., Delano-Wood, L., Giovannetti, T., Bettcher, B. M., Penney, D. L., Swenson, R., Lippa, C., Kabasakalian, A., Bondi, M. W., & Libon, D. J. (2011). Dysexecutive functioning in mild cognitive impairment: Derailment in temporal gradients. *Journal of the International Neuropsychological Society*, *18*(1), 20–28. https://doi.org/10.1017/s1355617711001238

[24] Wechsler, D. (1945). A standardized memory scale for clinical use. *The Journal of Psychology*, 19(1), 87–95. https://doi.org/10.1080/00223980.1945.9917223

[25] Wechsler, D. (2008). Wechsler Adult Intelligence Scale--fourth edition. *PsycTESTS Dataset*. https://doi.org/10.1037/t15169-000

[26] Libon, D. J., Bondi, M. W., Price, C. C., Lamar, M., Eppig, J., Wambach, D. M., Nieves, C., Delano-Wood, L., Giovannetti, T., Lippa, C., Kabasakalian, A., Cosentino, S., Swenson, R., & Penney, D. L. (2011). Verbal serial list learning in mild cognitive impairment: A profile analysis of interference, forgetting, and errors. *Journal of the International Neuropsychological Society*, 17(05), 905–914. https://doi.org/10.1017/s1355617711000944

[27] Lamar, M., Price, C. C., Libon, D. J., Penney, D. L., Kaplan, E., Grossman, M., & Heilman, K. M. (2007). Alterations in working memory as a function of leukoaraiosis in dementia. *Neuropsychologia*, 45(2), 245–254. https://doi.org/10.1016/j.neuropsychologia.2006.07.009

[28] Lamar, M., Catani, M., Price, C. C., Heilman, K. M., & Libon, D. J. (2008). The impact of region-specific leukoaraiosis on working memory deficits in dementia. *Neuropsychologia*, 46(10), 2597–2601. https://doi.org/10.1016/j.neuropsychologia.2008.04.007

[29] Emrani, S., Libon, D. J., Lamar, M., Price, C. C., Jefferson, A. L., Gifford, K. A., Hohman, T. J., Nation, D. A., Delano-Wood, L., Jak, A., Bangen, K. J., Bondi, M. W., Brickman, A. M., Manly, J., Swenson, R., & Au, R. (2018). Assessing working memory in mild cognitive impairment with serial order recall. *Journal of Alzheimer's Disease*, 61(3), 917–928. https://doi.org/10.3233/jad-170555

[30] Emrani S., Lamar M., Price C.C., Baliga S., Wasserman V., Matusz E., Saunders J., Gietka V., Strate J., Swenson R., Baliga G. and Libon D.J, Neurocognitive constructs underlying executive control in statistically-determined Mild Cognitive Impairment, *Journal of Alzheimer's Disease*, vol. 82(1), p.5-16, 2021. https://doi.org/10.3233/JAD-201125 (edited)

[31] Emrani, S., Lamar, M., Price, C. C., Baliga, S., Wasserman, V., Matusz, E., Swenson, R., Baliga, G., & Libon, D. J. (2021). Assessing the capacity for mental manipulation in patients with statically-determined mild cognitive impairment using digital technology. *Exploration of Medicine*. https://doi.org/10.37349/emed.2021.00034

[32] Libon, D. J., Baliga, G., Swenson, R., & Au, R. (2021). Digital Neuropsychological Assessment: New Technology for Measuring Subtle Neuropsychological Behavior. *Journal of Alzheimer's Disease*, 82(1), 1–4. https://doi.org/10.3233/jad-210513

[33] Corsi, P. M. (1972). Human memory and the medial temporal region of the brain.

[34] Vandierendonck, A., Kemps, E., Fastame, M. C., & Szmalec, A. (2004). Working memory components of the Corsi blocks task. *British Journal of Psychology*, 95(1), 57–79. https://doi.org/10.1348/000712604322779460

[35] Clark, C., & Gazoni, E. (2021). OpenPyxl - A Python library to read/write Excel 2010 xlsx/xlsm files. OpenPyxl. Retrieved May 21, 2022, from https://openpyxl.readthedocs.io/en/stable/

[36] AWS Lambda. Amazon. Retrieved May 20, 2022, from https://aws.amazon.com/lambda/

[37] Chalice. Amazon. Retrieved May 20, 2022, from https://aws.github.io/chalice/

[38] HP. (2021). HP Reverb G2 Omnicept Edition. Retrieved May 21, 2022, from https://www.hp.com/us-en/vr/reverb-g2-vr-headset-omnicept-edition.html

[39] Tobii AB. (2022, February 24). HP Reverb G2 Omnicept Edition. Tobii VR. Retrieved May 21, 2022, from https://vr.tobii.com/integrations/hp-reverb-g2-omnicept-edition/

[40] Blanco-Campal, A., Diaz-Orueta, U., Navarro-Prados, A. B., Burke, T., Libon, D. J., & Lamar, M. (2019). Features and psychometric properties of the Montreal Cognitive Assessment: Review and proposal of a process-based approach version (Moca-PA). *Applied Neuropsychology: Adult*, *28*(6), 658–672. https://doi.org/10.1080/23279095.2019.1681996

[41] Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MOCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

[42] IBM. (2021). SPSS software. IBM. Retrieved May 21, 2022, from https://www.ibm.com/analytics/spss-statistics-software

[43] De Anda-Duran, I., Alonso, C. F., Libon, D. J., Carmichael, O. T., Kolachalama, V. B., Suglia, S. F., Au, R. and Bazzano, L. A. (2022). Carotid Intima-media Thickness and Midlife Cognitive Function: Impact of Race and Social Disparities in the Bogalusa Heart Study. *Neurology*. https://doi.org/10.1212/WNL.0000000000200155

[44] Stack Overflow. (2022). *Where developers learn, share, & build careers.* Stack Overflow. Retrieved May 27, 2022, from https://stackoverflow.com/

[45] Unity Technologies. (2022). Unity. Retrieved May 21, 2022, from https://unity.com/

[46] Epic Games Inc. (2022). The most powerful real-time 3D creation tool. Unreal Engine. Retrieved May 21, 2022, from https://www.unrealengine.com/en-US

[47] Valve Software. (2021, February 24). OpenVR SDK. GitHub. Retrieved May 21, 2022, from https://github.com/ValveSoftware/openvr

[48] The Khronos Group. (2016, December 6). OpenXR - high-performance access to AR and VR platforms and devices. The Khronos Group. Retrieved May 21, 2022, from https://www.khronos.org/openxr/

Appendix

Challenges of Implementation

Apple iPad and iOS Programming

In their prior work on digital neuropsychological assessment, Drs. Baliga and Libon had chosen the iPad as the test administration platform. Our work considered two possible choices for software platform. The first choice was a cross-platform framework such as Facebook's React Native or Google's Flutter. The most important advantage of this approach was that it could lead to an app that works both on Android and iOS platforms. The second choice was to use Apple's Xcode and the Swift programming language to create a native iOS app. We adopted the second approach primarily because it allows for very fine-tuned programs and access to Apple's libraries that would not have been easily available had we adopted the first approach. For example, we would not be able to directly access information from the Apple Pencil. This information was useful in some of the tests that were implemented, where we used pressure data from the Apple Pencil to understand a patient's motor input.

Xcode uses a *storyboard* to graphically lay out how different views (or menus) flow from one to the next. There is no way to programmatically construct new views without having to do something in the GUI of Xcode. Thus, each view has to have code specific to how the app will interact in that view and you have to, by hand, lay out the flow of the app. This can quickly get out of hand depending on how many cognitive tests are included.

In many cases, there are times where the documentation from Apple is sparse. In these situations, I found myself consulting forums such StackOverflow [44] where such problems and solutions are discussed by other app developers. For example, every time a "touch" is detected on the screen, a new *Touch* object is created. However, we have to look into the coalesced touches for each patient stimulus selection in order to obtain the full gamut of data that was collected in that input. By doing this, we can get an accurate measure of the pressure, azimuth angle, and altitude applied to the Apple Pencil when it is used in the test. This detail was not readily apparent in Apple's documentation at the time of development.

Cancellation Test Challenges

2D collision detection is another tricky feat in Swift to accomplish. This was relevant for the Cancellation and Angle Matching tests. The Cancellation test, designed by Dr. Baliga and Dr. Libon, features three test conditions (presented in 3 screens), namely, a letter condition, a symbol condition and a symbol-letter combined condition. In the letter condition, the patient is provided with a screen comprising several hundred different letters located at different positions within the screen. The patient has to use the Apple Pencil to circle all occurrences of a specified target letter. In our implementation of cancellation this test, we need to be able to tell if the circle drawn by a patient is surrounding an object on the screen. This is easier with letters and symbols since we can just draw a *CGRect* around the element and adjust the size and dimensions as we please. Then we can easily check if a *CGRect* is colliding with another.

Angle Matching Test Challenges

The Angle Matching test is a visual test designed by Dr. Baliga and Dr. Libon that tests for a patient's ability to assess orientation of lines presented in a two-dimensional space. A stimulus in the Angle Matching test contains one target line and some candidate lines. Exactly one candidate line, i.e., the correct answer, has the same orientation as the target line and the rest of the candidate lines are foils. The patient's task is to select the correct answer. We implemented this test on the iPad and using Virtual Reality (see below). In the iPad implementation of this test, it is not possible to use the *CGRect* object functionality to detect which line was selected. This is because *CGRect* objects are bound to the coordinate axes. Rotating the graphical context will only change how the *CGRect* object appears but it does not change the space it occupies. A better solution to this problem was to check if the bounding box of the drawn path intersects with the bounding box of the angle. This prevents the bounding boxes of our angles from being too large in order to account for their full width and length, and allows patients to mark an angle simply by drawing a line through it rather than having to draw a circle around the entire line. Thus, we have a more precise method of detecting collisions between the Apple Pencil path and the angled line's bounding box.

In our iPad application, we desired a modular and easily adjustable system to display tests and track data. Very early in development, we agreed that all tests and trials should be dictated by a JSON file that includes all the parameters. This way, all visible stimuli can be adjusted without having to change any code, just the JSON file that controls the test.

Unity

Unity [45] is a 3D game engine that has seen use in a multitude of science projects from industry to educational research. Unity was chosen as our 3D engine because of its ease of use, which advertises a gentler learning curve as opposed to Epic Games' Unreal engine [46]. These two 3D engines are known as industry standards and their use can be seen across a wide range of media and technology. While both 3D engines are capable for this project and will work with HP's headset and biometric sensors, Unity allows the use of C# scripts which are less time-consuming to write as opposed to Unreal's use of C++ for scripting. Additionally, Unity's documentation and community proved to be helpful in a variety of design challenges.

There are a multitude of VR frameworks that interface with Unity's XR (extended reality) layer, including Windows Mixed Reality (WMR), Valve Software's OpenVR [47], and OpenXR [48] which uses a compatibility layer to work with the VR APIs listed as well as a few not listed. For the purposes of this test, I decided on the OpenXR loader since this will allow for new VR headsets and eye tracking protocols to be implemented without changing the VR display layer. This will also allow for additional input devices. This seems to be where the Unity community is moving towards, as more and more VR devices with new capabilities become available.

Handling User Displacement

In some versions of tests, I use a script to adjust the view of the patient. This is necessary when a patient is not looking directly forward at the test stimuli or for whatever reason is displaced while in the VR environment. To achieve this, I apply an offset to the

VR camera, and adjust its rotation if the view is too far away from the test stimuli. The animation is smooth and is not uncomfortable or nauseating for the user.

Adjusting to the VR environment

Due to the differences in how space is defined between iPad and Unity, I had to adjust how the stimuli is displayed in Unity, as our first implementations were on the iPad. The coordinate systems used in Unity and in the Swift language graphical API are different. In Swift, UI elements are drawn on the screen using the top left corner as the origin point. This means that the Y coordinate value increases as you go down. In Unity, this Y axis is flipped so the origin point is on the floor. In order to draw angled lines for our Angle Matching test, I flipped the *Canvas* object that displays the lines. These lines are parented to the *Canvas* object, so by flipping the *Canvas*, the lines are also flipped. Thus, the stimuli and JSON file dictating how UI elements are drawn did not have to be changed at all. Like the iPad application, the JSON files that dictate the structure of the test can be effortlessly swapped between iPad and Unity applications as long as the test itself has been implemented.

Early into development, we believed it was important to figure out how to make UI elements that did not appear for the VR user. I thought it would be helpful for the test giver to see certain data such as pulse and cognitive load as the test progresses. Additionally, seeing their progress in identifying the correct stimuli would also be helpful. It would not make sense for the test taker to see their own pulse as the test is being done. That would be potentially stressful and does not help the test taker. It makes more sense to only give this information to the test giver. This can be achieved by pointing UI elements to the test giver's monitor, and only pushing what is important to

the test taker. This separation of UI elements is important so that the test giver can dictate the parameters of the test while the test taker can wait patiently as they become acclimated to the VR environment.

Data Collection Challenges

There is no built-in feature in HP's or Tobii's libraries that allow you to track what objects have been hovered and any details tied to that hover (timestamp, what object was hovered on, etc.) This feature does exist, but requires users to pay over $1.5K for a research license to Tobii who developed the software. For my own purposes, I decided to program my own version of this software. This system uses collision objects surrounding each quadrant of the test screen, allowing us to know if a patient is looking in a certain area, but not at any stimuli. This system reports the timestamp of the gaze detected, where this gaze was, and for how long it lasted. We can then differentiate between regions of the space as well as how long they spent looking at the target angle.

With this solution we can effectively know when and where a patient was looking and what kind of decisions they were possibly making during that time. This solution also helps with a few other problems, as a more complex solution might have to know the user's head position and rotation to understand exactly where they are looking.

Design Challenges of the Eye Tracking Tests

In these VR versions of cognitive tests, patients must provide input by using an included controller with the VR headset. By clicking a button on the controller, a selection can be made on an object that a person is gazing at. This is a departure from the

way that similar Cancellation and Angle Matching tests are carried out on pencil-and-paper test, where the patient selects a stimulus presented on a page by circling it.

Various cognitive tests were implemented using Unity's 3D game engine and an assortment of libraries that seamlessly integrate eye tracking input and biometrics from HP's headset. The code needed to acquire real-time eye tracking data was provided by HP and Tobii and designed to interface with 3D game engines like Unity and Epic Games's Unreal engine.

Just like in our iPad tests, the test parameters are determined by parsing a JSON file that includes information about a trial's stimuli. Afterwards, 3D objects had to be designed to react when a user's eyes look at an object. Once that was put together, I built a system that would be able to report how long a certain object or area within the virtual environment was looked at and at what time interval during the trial (as described above). This is done simply by assigning collision objects to 3D objects. The patient's eye gaze is represented by a ray pointing from their eyes. If that ray collides with a 3D collision object that means it is being looked at. The 3D object can be programmed to visually react by changing color if desired. The patient's pulse and pupillometry can also be recorded for the duration of the test.

After integrating the Cancellation test into our application, we realized that the eye tracking and gaze detection capabilities of our VR platform are not sufficiently accurate and precise. This is mainly due to the distance from the VR user to the stimuli. In Cancellation, there are over 200 stimuli on the screen and thus the stimuli had to be properly distanced from the viewer. Tobii's eye tracking sensors model the viewer's eye gaze as a cone of vision. Given the number of stimuli on a Cancellation test's screen and

the distance to the viewer, that cone of vision does not always contain a single stimulus, making it impossible for our application to discern which letter is being viewed. In the VR environment, the Cancellation test can become overwhelming due to the number of elements on the screen, which led us to abandoning further development of this test focusing our attention on the Angle Matching test. Our implementation of this test provides latencies, eye position, and biometrics such as pulse during the entire duration of each response.