Validation of a digitally delivered visual paired comparison task: reliability and convergent validity with established cognitive tests

Joshua L. Gills · Jordan M. Glenn · Erica N. Madero ·
Nick T. Bott · Michelle Gray

Received: 21 March 2019 / Accepted: 14 August 2019 / Published online: 29 August 2019
© American Aging Association 2019

Abstract Alzheimer’s disease (AD) affects the memory and cognitive function of approximately 5.7 million Americans. Early detection subsequently allows for earlier treatment and improves outcomes. Currently, there exists a validated 30-min eye-tracking cognitive assessment (VPC-30) for predicting AD risk. However, a shorter assessment would improve user experience and improve scalability. Thus, the purposes were to (1) determine convergent validity between the 5-min web camera-based eye-tracking task (VPC-5) and VPC-30, (2) examine the relationship between VPC-5 and gold-standard cognitive tests, and (3) determine the reliability and stability of VPC-5. This prospective study included two healthy cohorts: older adults (65+ years, $n = 20$) and younger adults (18–46 years, $n = 24$). Participants were tested on two separate occasions. Visit 1 included the Montreal Cognitive Assessment (MoCA), Digit Symbol Coding test (DSC), NIH Toolbox Cognitive Battery (NIHTB-CB), VPC-30, and VPC-5. Visit 2 occurred at least 14 days later; participants completed the VPC-5, DSC, NIHTB-CB, and dual-task walking assessments (DT). VPC-30 significantly correlated

with VPC-5 at the first ($p < .001$) and second ($p = .001$) time points. VPC-5 and DSC ($p < .01$) and Pattern Comparison Processing Speed Test (PSPAC) ($p = .01$) were also correlated on day 1. Significant associations existed between VPC-5 and DSC ($p < .001$), Flanker Inhibitory Control Test ($p = .05$), PSPAC ($p < .001$), and Picture Sequence Memory Test ($p = .02$) during day 14 testing session. The test retest reliability of VPC-5 was significant ($p < .001$). VPC-5 displayed moderate convergent validity with the VPC-30 and gold-standard measures of cognition, while demonstrating strong stability, suggesting it is a valuable assessment for monitoring memory function.

Keywords Cognition · Visual paired comparison · Dementia · Alzheimer’s disease

Introduction

Alzheimer’s disease (AD) is a growing public health concern in the USA. As of 2018, an estimated 5.7 million Americans were diagnosed with AD (Alzheimer’s Association 2018). AD is the sixth leading cause of death in the USA, responsible for more than 80,000 deaths annually (Alzheimer’s Association 2015). Additionally, a new diagnosis occurs every 65 s, resulting in more than 1 million new cases each year (Alzheimer’s Association 2015). If this trajectory continues, there will be an added \$750 billion dollars in long-term care costs by 2050. Early detection of AD can save an estimated \$7.9 trillion in medical expenditures, with an estimated \$7.0 trillion saved

J. L. Gills · M. Gray (✉)
Exercise Science Research Center, University of Arkansas,
Fayetteville, AR, USA
e-mail: rgray@uark.edu

J. M. Glenn
VP Clinical Development, Neurotrack Technologies, Inc.,
University of Arkansas, Redwood City, CA, USA

E. N. Madero · N. T. Bott
Neurotrack Technologies, Inc., Redwood City, CA, USA

even under a partial early diagnosis projection (Alzheimer's Association 2018). As a result, early detection of mild cognitive impairment (MCI) and AD is imperative to allow for prompt intervention, potentially reducing the rising costs associated with these age-related cognitive changes.

The average age of MCI and/or AD detection is 70 years (Hanninen et al. 2002; Ward et al. 2012); however, negative neuronal and neurobiological changes begin approximately 15 years before apparent cognitive changes are recognized (Bott et al. 2018; Kremen et al. 2014). Thus, it is critical to administer reliable cognitive assessments that can detect cognitive deficits as early as possible in the disease's progression. Furthermore, these tests should aim to distinguish between normal changes seen with cognitive aging, such as decreases in recall, and those that are indicative of impairment, such as decreases in recognition memory (Danckert and Craik 2013; Sekuler et al. 2005). Many traditional in-clinic assessments, while reliable, lack adequate scalability to the general population because trained professionals are required to guide participants through the tests. Remotely delivered assessments hold the potential to screen large numbers of older adults for signs of impairment.

Visual paired comparison (VPC) tasks are eye tracking-based tests that have been validated as a discriminant cognitive assessment tool (Crutcher et al. 2009; Zola et al. 2013). VPC tasks assess visual memory recognition by dividing the amount of time spent viewing novel images by the total time spent viewing both novel and familiar images (Bott et al. 2017; Fagan 1970; Zola et al. 2013). Individuals with MCI or AD have impaired declarative memory for previously viewed images and tend to spend roughly equal amounts of time viewing both novel and previously viewed images. Conversely, individuals with normal cognitive function spend a greater proportion of time viewing the novel images (i.e., the images not previously shown). Healthy older individuals should not have notably lower scores on VPC tasks than younger individuals, as recognition memory tends to remain stable with normal cognitive aging (Danckert and Craik 2013). However, individuals with MCI or AD are expected to score lower than unimpaired individuals even among participants that may have preclinical changes in cognition. Importantly, previous literature demonstrates that the 30-min VPC (VPC-30) task can reliably predict onset of either MCI or AD within 3–6 years (Zola et al. 2013).

Traditionally, VPC tests have utilized commercial grade eye-trackers, which are high frame-rate cameras capable of capturing a number of complex visual features. These commercial trackers are a common method for tracing eye movements during VPC tasks (Crutcher et al. 2009; Lagun et al. 2011; Zola et al. 2013). However, this equipment is expensive and not readily available outside of research or clinical facilities, limiting scalability of the assessment. Recently, Bott and colleagues (Bott et al. 2018) validated the use of readily available web cameras for VPC tasks as an alternative to the traditional gold-standard eye-tracking equipment. With web cameras ubiquitously available on most standard computing/smart devices, this represents a more scalable method to gather population-level and longitudinal data on cognitive health. Since eye movements provide a non-invasive “window to the brain” and offer valuable insights into cognitive function, the recording of gaze with web cameras has become a trend in recent cognitive research (Hansen and Pece 2005; Lin et al. 2013; Petridis et al. 2013; Vivero et al. 2010).

The aforementioned VPC-30 is one such test that has been validated for use with standard built-in web camera to assess memory function (Bott et al. 2017). It is important to note that while the VPC-30 is a validated test, it is meant to screen for cognitive impairment associated with various forms of dementia, rather than provide a diagnosis. The VPC-30 is a passive paradigm, meaning participants complete the test in its entirety without explicit instructions on where they are supposed to focus their gaze. The integrity of the VPC-30 test is dependent on the user not knowing exactly what the test is measuring. As a result, a task utilizing a shorter active paradigm, in which participants are given specific instructions beforehand, has been developed to improve the user experience and increase the scalability of the assessment. This shorter test with an active paradigm allows for repeat testing over time without influencing the results. Therefore, the purpose of the current investigation was to validate an active web-based 5-min version of the VPC (VPC-5) test among healthy younger and older adults. The three primary objectives of this study were to (1) determine convergent validity between the VPC-5 and VPC-30, (2) examine the relationship between the VPC-5 and gold-standard cognitive tests, and (3) determine the stability and test-retest reliability of the VPC-5 over time.

Methods

Participants and procedures

This investigation included a prospective study design, in which participants were tested twice with at least 14 days between testing sessions. A group of 49 cognitively healthy participants were recruited through flyers, website announcements, and active speaking engagements. Cognitively healthy subjects, as opposed to individuals with cognitive impairment, were recruited so that the efficacy of the new technology could first be validated in a healthy population before expanding to include those with cognitive issues. However, we chose to include both younger and older participants in order to assess the feasibility and validate the test across a wider age range.

Of the 49 recruited participants, 44 completed all assessments. The five individuals that did not complete all exams during both testing sessions were excluded from final analyses. Eligible participants were between the ages of 18–46 years (younger adults) or above the age of 65 (older adults) and were able to read and understand English. Of the four participants excluded for incomplete data, three were older adults and one was a younger adult. Individuals were also excluded from the study if they met any of the following criteria: diagnosed with ADHD, disabling vision loss, inability to complete the calibration procedure for the web camera, history of substance abuse, or known learning disability, neurological illness (stroke, tumor), psychiatric illness, current MCI/AD diagnosis, or scoring less than 26 on the Montreal Cognitive Assessment (MoCA).

Participants reported for testing on two separate occasions. During the first visit, participants signed the informed consent document; completed a medical history questionnaire; and completed the MoCA, Wechsler Adult Intelligence Scale-Revised Digit Symbol Coding Test (DSC), NIH Toolbox Cognitive Battery (NIHTB-CB), VPC-30 test, and VPC-5 test (Fig. 1). The second testing session occurred at least 14 days later. During the second visit, subjects completed the VPC-5, DSC, NIHTB-CB, dual-task, and provided initial biometric measurements (height, weight, and body mass index). The NIHTB-CB and VPC tasks were randomized within subjects from day 1 to day 14 to ensure score differences were not due to fatigue resulting from cognitive load.

Biometric assessments

Biometric assessments included height, weight, and body mass index. Height was measured with a standing stadiometer (Seca; Hamburg, Deutschland). During this assessment, subjects were asked to remove their shoes and stand up as straight as possible. Height was recorded to the nearest 0.1 cm. Weight was measured with a balance-beam scale (Sunbeam Products, Inc., McCook, IL). Participants removed their shoes, any heavy clothing (sweaters, jackets, or coats), and emptied their pockets. Weight was measured to the nearest 0.1 kg. Body mass index was calculated as a ratio between weight and height (kg/m^2).

Cognitive assessments

Montreal cognitive assessment MoCA is a paper-pencil test that is commonly used in clinical settings as a screening tool for cognitive impairment. The test assesses the following cognitive domains: attention and concentration, executive function, memory, language, visuoconstructional skills, conceptual thinking, calculations, and orientation. This assessment has been described in detail elsewhere (Nasreddine et al. 2005). Briefly, the short-term memory recall task, worth 5 points, involves two learning trials of five nouns and delayed recall towards the end of the exam. Visuospatial abilities are assessed using a clock-drawing task, worth 3 points, and a three-dimensional cube copy worth 1 point. Multiple aspects of executive functions are assessed using an alternation task adapted from the Trail Making B task (1 point), a phonemic fluency task (1 point), and a two-item verbal abstraction task (2 points). Attention, concentration, and working memory are evaluated using a sustained attention task (detection using hand tapping; worth 1 point). Language is assessed using a three-time confrontation naming task with low-familiarity animals (lion, dromedary, rhinoceros; 3 points), repetition of two syntactically complex sentences (2 points), and the aforementioned fluency task. Orientation to date, day, and place are evaluated as the final task and is worth 6 points. MoCA performance was used to categorize subjects as cognitively intact (≥ 26) or cognitively impaired (< 26) (Nasreddine et al. 2005); subjects scoring < 26 were excluded from the study.

Fig. 1 Study testing flow. This figure illustrates the study flow for each participant. The use of the “dagger” symbol indicates that participants had to wait a minimum of 14 days before retesting (not all retesting trials took place at exactly 14 days after baseline). The use of the “number sign” symbol indicates that the NIH Toolbox Cognitive Battery (NIHTB-CB) and visual paired comparison (VPC) tasks were randomized within subjects on days 1 and 14

Digit Symbol Coding The DSC test is a paper-pencil test that assesses processing speed, working memory, visuospatial processing, and attention (Donohue et al. 2014). DSC consists of rows containing small blank squares, each paired with a randomly assigned number from one to nine. Above these rows is a printed key that pairs each number with a different symbol. Using the reference key, the subject has 90 s to pair specific numbers with the appropriate geometric figures. The score is determined by the number of symbols that are correctly paired with the corresponding numbers. This test is a valid and reliable measure for detecting early signs of cognitive decline (Crowe et al. 1999; Donohue et al. 2014) and predicting cognitive disorders (Best et al. 2016).

NIH Toolbox Cognitive Battery The NIHTB-CB is a computerized cognitive composite that assesses various functional domains. The four tests comprising the NIHTB-CB were completed on an iPad in this study (version 11.4, Apple Inc., Cupertino, CA). First, the Flanker Inhibitory Control and Attention Test (Flanker) measures inhibitory control and attention. The Flanker requires the participant to focus on a particular stimulus while inhibiting attention to the surrounding stimuli. Next, the Dimension Change Card Sort Test (DCCS) evaluates cognitive flexibility and attention. Two target pictures are presented that vary along two dimensions (e.g., shape and color). Participants are asked to match a series of bivalent test pictures (e.g., yellow balls and blue trucks) to the target pictures, first according to one dimension (e.g., color) and then, after a number of trials, according to the other dimension

(e.g., shape). The Pattern Comparison Processing Speed Test (PSPAC) measures processing speed. The test requires participants to discern whether two side-by-side pictures are identical or different. Finally, the Picture Sequence Memory Test (PSMT) measures episodic memory. Sequences of pictured objects and activities are presented in a particular order and participants are asked to reproduce the sequence of pictures shown on the screen. The NIHTB-CB composite has high reliability and good construct validity for evaluating cognition, indicating it can be an effective tool in epidemiologic and clinical studies (Heaton et al. 2014).

Visual paired comparison assessment

The VPC tests (VPC-30 and VPC-5) were completed on a laptop computer equipped with a factory-installed web camera. The VPC test construction is explained in detail elsewhere (Bott et al. 2018). Briefly, VPC tasks use eye-tracking data to assess declarative memory by quantifying the amount of time a participant spends viewing novel images compared to previously viewed images (Crutcher et al. 2009; Zola et al. 2013; Bott et al. 2018). The VPC-5 test is an active paradigm. At the beginning of the test, participants are instructed to remember the images from the familiarization phase and to focus on the novel images during the test phase rather than the previously viewed images. Conversely, the VPC-30 test is a passive paradigm in which participants are only instructed to watch the images as if they were watching television.

During the familiarization phase of both the VPC-30 and the VPC-5 assessments, participants are presented

with pairs of identical visual stimuli for 5 s. After the familiarization phase of the VPC-30, a 2-s or 2-min delay occurred before the test phase, allowing for the evaluation of both instant and delayed recognition memory. After the familiarization phase of the VPC-5, a continuous and cumulative delay occurred across each test trial. During the test phase, subjects were presented with pairs of visual stimuli (images), including one from the familiarization phase and one novel image. The proportion of time a participant spent gazing at the novel image relative to the total viewing time produced a novelty preference score, with higher scores representing better declarative memory and lower scores indicating impaired cognitive function (Bott et al. 2018). The assessments were recorded and a video of the participant's face was stored on a secure server. Eye movements were tracked and scored. Detailed scoring information is published elsewhere (Bott et al. 2018). The VPC-30 assessment was completed on day 1 and the VPC-5 assessment was completed on days 1 and 14. VPC-30 was not re-administered in order to mitigate performance contamination with VPC-5.

Dual-task

The dual-task walking assessment is a measure of attention and executive function (Brustio et al. 2017; Yogev-Seligmann et al. 2008). Dual-task assessments vary in protocol. For this investigation, we instructed participants to walk a distance of 10 m at their usual speed without a dual-task condition. There was a 5-m distance before and after the 10-m distance to account for acceleration and deceleration (Glenn et al. 2015). For the next part of the assessment, participants were instructed to walk as quickly and safely as possible without running. These two assessments were used as the baseline tests. For the dual-task conditions, participants were instructed to perform the same walking conditions and simultaneously perform serial subtractions (Hausdorff et al. 2001). A random 3-digit number was selected (100–999) and participants were instructed to subtract three from each number while performing each walking condition (usual and fast). The walking speed trials were averaged separately and used for all analyses. Dual-task is a valid and highly reliable method for assessing working memory in young and older adults (Montero-Odasso et al. 2009; McCulloch et al. 2009). Dual-task was completed strictly on day 14 because it is a

physically active task. All physical tasks and biometric assessments were completed on day 14 (Fig. 1).

Data analysis

Pearson's correlation coefficient analyses were used to determine the convergent validity between the VPC-5 and VPC-30 tests, test-retest reliability of VPC-5 between testing days, and relationships between the VPC-5 and the other standard cognitive tests (DSC, NIHTB-CB, and dual-task). Normative scores were utilized during NIHTB-CB analyses. A Cronbach's alpha intraclass correlation (ICC) test was performed to determine internal consistency across the trials of the VPC-5 on day 1, as well as the trials of the VPC-5 from days 1 and 14. Repeated-measures analyses of variance (ANOVA) was performed to detect changes over time for the VPC-5, NIHTB-CB, and DSC. Statistical significance was set at $\alpha = 0.05$ for all analyses.

Results

Participant characteristics

Of the 44 subjects who completed the study, mean age was 50.5 ± 27.5 years (range 21–89 years; Table 1). The overall sample was 64% female (28 subjects) and over 75% of participants were college or trade school graduates. Due to technical difficulties (recording quality, network connectivity, glare from glasses, and/or low light in the testing room), not all web camera VPC data could be scored. Of the 44 subjects who completed the assessments, 42 of the VPC-30 scores were evaluated, 42 of the day 1 VPC-5 scores were analyzed, and 37 of the day 14 VPC-5 scores were examined. The second testing session took place an average of 22.6 days after the initial testing session.

Relationship between 30-min and 5-min visual paired comparison tests

Analysis revealed a significant positive association between the VPC-30 and VPC-5 tests, indicating that VPC-5 is also a valid measure of declarative memory. The VPC-30 was significantly correlated with the VPC-5 at the first ($r = .55$; $p < .001$; Fig. 2a) and second ($r = .51$; $p = .001$; Fig. 2b) time points. It is important to note that due to the single

Table 1 Characteristics of the study cohort ($n = 44$)

	Young adults ($n = 24$)	Older adults ($n = 20$)	p values
Average age (SD)	26.5 years (7.4)	79.3 years (6.4)	$p = .00$
Sex			$p = .87$
Female	62.5%	65.0%	
Education			$p = .08$
High school graduate	4.2%	5.0%	
Some college	29.2%	10.0%	
College graduates or higher	66.7%	85.0%	
Race			$p = .21$
European-American	83.3%	95.0%	
Other	16.7%	5.0%	
Biometric			
Height (SD)	170.1 cm (8.6)	166.5 cm (8.6)	$p = .38$
Weight (SD)	74.5 kg (17.8)	73.2 kg (14.3)	$p = .96$
Body mass index (SD)	25.8 kg/m ² (6.4)	26.3 kg/m ² (3.9)	$p = .76$

SD standard deviation

administration of the VPC-30 (due to concern regarding performance contamination given the

assessment's passive paradigm, the VPC-5 at day 1 and day 14 are correlated to the VPC-30 at day 1).

Fig. 2 Relationship between the 30-min visual paired comparison task (VPC-30) and the 5-min visual paired comparison task (VPC-5) at **a** day 1 and **b** day 14. Black circles represent older subjects and red circles represent younger subjects

Reliability of VPC-5 test

Performance on VPC-5 test on day 1 was positively correlated with performance on the VPC-5 on day 14 ($r = .73$; $p < .001$; Fig. 3) indicating acceptable test-retest reliability between assessments. Additionally, there was no difference between day 1 and day 14 for VPC-5 performance ($t = 1.719$, $p = .09$). When evaluating intraclass correlations between trials on day 1, Cronbach's alpha revealed stability over time ($\alpha = .90$). Additionally, longitudinal stability of VPC-5 was assessed using data from both days 1 and 14. All trials were examined using Cronbach's alpha; results revealed strong reliability between testing days ($\alpha = .91$), demonstrating VPC-5 is a stable assessment of cognition over time.

Associations between VPC-5 test and cognitive assessments

Cognitive assessments This study also examined the correlation between VPC-5 task performance and domain-specific cognitive function. The NIHTB-CB age-based normative percentile scores for each task were used as a basis of comparison for VPC-5

performance. On day 1 (Table 1), significant relationships existed between the VPC-5 and PSPAC ($r = .44$; $p = .01$; Fig. 4a), but not the PSMT ($r = .10$; $p = .54$; Fig. 4b), Flanker ($r = .20$; $p = .19$; Fig. 4c), or DCCS ($r = .18$; $p = .24$). On day 14 (Table 2), significant associations were observed between VPC-5 and PSPAC ($r = .56$; $p < .001$; Fig. 5a), PSMT ($r = .38$; $p = .02$; Fig. 5b), and Flanker ($r = .32$; $p = .05$; Fig. 5c); however, no relationship was observed with the DCCS ($r = .22$; $p = .19$). MoCA scores were not significantly correlated with the VPC-5 at day 1 ($r = .11$; $p = .47$; Fig. 6a) or day 14 ($r = .19$; $p = .26$; Fig. 6b). Finally, scores on the DSC were significantly correlated with the VPC-5 on day 1 ($r = .38$; $p = .01$; Fig. 4d) and day 14 ($r = .41$; $p = .01$; Fig. 5d).

Dual-task We also evaluated functional measures of cognition by comparing DT-HS and DT-MS scores to the VPC-5 at day 1 and day 14. Neither DT-HS ($r = -.19$; $p = .23$; Fig. 7a) or DT-MS ($r = -.16$; $p = .29$; Fig. 7b) were significantly correlated with VPC-5 at day 1, and the results were similar between the Day 14 VPC-5 scores and DT-HS ($r = -.19$; $p = .26$; Fig. 8a), as well as DT-MS ($r = -.13$; $p = .43$; Fig. 8b).

Fig. 3 Relationship between the 5-min visual paired comparison task (VPC-5) at day 1 and day 14. Black circles represent older subjects and red circles represent younger subjects

Fig. 4 Relationships between the 5-min visual paired comparison task (VPC-5) and the **a** Pattern Comparison Processing Speed Test (PSPAC), **b** Picture Sequence Memory Test (PSMT), **c** Flanker

Inhibitory Control Test (Flanker), and **d** Digit Symbol Coding Test (DSC) on day 1. Black circles represent older subjects and red circles represent younger subjects

Discussion

The three primary objectives of this study were to (1) demonstrate convergent validity between the active VPC-5 test paradigm and the passive VPC-30 test paradigm, (2) examine the relationship between the VPC-5 and gold-standard cognitive tests, and (3) determine the stability and test-retest reliability of the VPC-5 over time. For the first objective, our study found moderate

convergent validity between the VPC-5 and VPC-30 tests, indicating the VPC-5 is a viable alternative to the VPC-30 for measuring visual recognition memory. The active nature of the VPC-5 (i.e., participants are given instructions beforehand) differs from the passive paradigm of the VPC-30. The inherent benefit of the active nature of the VPC-5 is that it is meant to be taken repeatedly over time, whereas the passive VPC-30 is meant to be taken only once to obtain a baseline score.

Table 2 Correlations between VPC-5 test data from day 1 and cognitive assessments from day 1

	VPC-5	VPC-30	Flanker	DCCS	PSPAC	PSMT	MoCA	DSC
VPC-5	—							
VPC-30	0.55*	—						
Flanker	0.20	−0.16	—					
DCCS	0.18	−0.05	0.56*	—				
PSPAC	0.44*	0.26	0.44*	0.44*	—			
PSMT	0.10	0.12	0.32*	0.28	0.28*	—		
MoCA	0.11	0.10	0.45*	0.47*	0.16	0.62*	—	
DSC	0.38*	0.27	0.40*	0.13	0.54*	0.38*	0.16	—

Flanker Flanker Inhibitory Control test, *DCCS* Dimension Change Card Sort test, *PSPAC* Pattern Comparison Processing Speed test, *PSMT* Picture Sequence Memory Test, *MoCA* Montreal Cognitive Assessment, *DSC* Digit Symbol Code test

* $p < .05$

Fig. 5 Relationship between the 5-min visual paired comparison task (VPC-5) and the **a** Pattern Comparison Processing Speed Test (PSPAC), **b** Picture Sequence Memory Test (PSMT), **c** Flanker

Inhibitory Control Test (Flanker), and **d** Digit Symbol Coding Test (DSC) on Day 14. Black circles represent older subjects and red circles represent younger subjects

Fig. 6 Relationship between the Montreal Cognitive Assessment (MoCA) and the 5-min visual paired comparison task (VPC-5) at **a** day 1 and **b** day 14. Black circles represent older subjects and red circles represent younger subjects

Fig. 7 Relationship between the 5-min visual paired comparison task (VPC-5) at day 1 and **a** Dual-task habitual and **b** Dual-task maximal scores. Black circles represent older subjects and red circles represent younger subjects

Fig. 8 Relationship between the 5-min visual paired comparison task (VPC-5) at day 14 and **a** Dual-task habitual and **b** Dual-task maximal scores. Black circles represent older subjects and red circles represent younger subjects

Table 3 Correlations between VPC-5 test data from day 14 data and cognitive assessments from day 14

	VPC-5	VPC-30	Flanker	DCCS	PSPAC	PSMT	MoCA	DSC	DT-HS	DT-MS
VPC-5	–									
VPC-30	0.51*									
Flanker	0.32	0.05	–							
DCCS	0.22	–0.09	0.54*	–						
PSPAC	0.56*	0.18	0.44*	0.51*	–					
PSMT	0.38*	0.16	–0.07	0.16	0.16*	–				
MoCA	0.19	0.10	0.17	0.25	0.33*	0.36*	–			
DSC	0.42*	0.23	0.46*	0.32*	0.60*	0.32*	0.19	–		
DT-HS	–0.19	–0.12	–0.14	–0.08	–0.27	–0.29	–0.01	–0.35*	–	
DT-MS	–0.16	–0.13	–0.24	–0.21	–0.46*	–0.16	–0.03*	–0.47*	0.86*	–

Flanker Flanker Inhibitory Control test, *DCCS* Dimension Change Card Sort test, *PSPAC* Pattern Comparison Processing Speed test, *PSMT* Picture Sequence Memory Test, *MoCA* Montreal Cognitive Assessment, *DSC* Digit Symbol Code test, *DT-HS* Dual-task habitual, *DT-MS* Dual-task maximal speed

* $p < .05$

Additionally, the active nature of the VPC-5 prevents the need to conceal what the test is measuring from the user, enhancing repeatability. Finally, the shorter duration of the VPC-5 lowers the burden on the user, thereby improving scalability.

The second objective of this investigation was to compare performance on the VPC-5 test to traditional gold-standard cognitive assessments. The VPC-5 showed significant correlations with traditional composites (Tables 2 and 3), suggesting that the VPC-5 may represent an alternative to traditional, longer cognitive assessments. Similar to the results from the Bott et al. (2018) study comparing the VPC-30 test to the NIHTB-CB and various paper-pencil tests, we found the VPC-5 test correlated significantly with measures of certain cognitive domains. Specifically, the VPC-5 had higher correlations with measures of inhibitory control and attention (Flanker), processing speed (PSPAC), visual episodic memory (PSMT), and executive function (DSC). These findings are consistent with previous research suggesting VPC tests have strong relationships with visual episodic memory (Bott et al. 2018) and processing speed, which is recognized to influence a variety of cognitive tasks, including those assessing executive function (Rose 1980; Salthouse 1996). As expected, there were no statistically significant differences between younger and older adults without cognitive impairment on the VPC-5, since recognition memory is a stable construct in normal aging. Future studies will examine if the VPC-5 can distinguish between cognitively healthy and impaired subjects.

A major issue with cognitive assessments revolves around learning effects associated with repeated administrations (Goldberg et al. 2015). While the NIHTB-CB only has one version of each assessment, making it susceptible to learning effects (Heaton et al. 2014), it should be noted that learning effects in the VPC-5 were mitigated by utilizing alternate forms of the test on days 1 and 14 (as recommended by Goldberg et al. 2015). On day 1, the VPC-5 was significantly correlated with only one out of the four NIHTB-CB tests; however, during assessments completed on day 14, three out of the four cognitive tests from the NIHTB-CB showed significant correlations with the VPC-5 test. Consequently, the Flanker and PSPAC out of the four of the NIHTB-CB tests demonstrated significant improvement from day 1 to day 14 (Flanker, PSPAC, and PMST). It is unlikely that these differences were related to changes in cognitive ability (Goldberg et al. 2015) and therefore suggest a learning effect associated with test administration. The VPC-5 was stable over time, resulting in no differences between time points ($p = .09$), while demonstrating significant intraclass correlations (the third objective of this investigation). This indicates a level of test-retest reliability that is higher than the NIHTB-CB. In parallel, this helps explain the significant correlations between the VPC-5 and NIHTB-CB at follow-up that were not present at baseline. We recommend that if using the NIHTB-CB, a practice trial should be provided for all participants to mitigate learning effects in future administrations; however, this does not appear to be necessary for the VPC-5.

One limitation to the current investigation is the issue with data quality from some of the VPC tests. Seven participants' assessments were omitted from scoring due to glare from glasses, low light in the testing room, and/or electronic errors. The tests that were omitted included one young adult on for VPC-5 day 1 and five young adults for VPC-5 day 14. The subjects were still included in the overall analyses because they completed both visits at the testing center. An additional limitation is that all assessments were completed in a research setting instead of the participant's home, limiting the ecological validity of the results. However, the validation of this assessment needed to take place in a research setting in order to ensure complete compliance and to gather high-quality data before remote testing could be deployed. In the present investigation, all participants were cognitively intact. Future studies should include individuals with different levels of cognition, as well as adults with known cognitive impairment to establish the ability of VPC-5 to discriminate between individuals with and without cognitive impairment. Finally, this study had a limited total sample size ($n=44$). It is warranted for these results to be further investigated in a larger sample, including individuals of different cognitive states (mild cognitive impairment, dementia, Alzheimer's disease, etc.) in an effort to further expand the generalizability of the findings.

Conclusion

Overall, this study demonstrated moderate convergent validity between the VPC-5 and VPC-30. Results also showed significant associations between the VPC-5 and standard cognitive assessments, as well as strong test-retest reliability and stability. As a result, the VPC-5 test may be a valuable and reliable assessment for monitoring memory function.

Current clinical practice primarily utilizes in-person methods of cognitive assessment which, while effective, require an exorbitant amount of time and effort and thereby limit scalability (Harvey 2012). While there are many validated assessments available today that evaluate cognitive function (Bland et al. 2016; Maruff et al. 2009; Smith-Ray et al. 2016), it is impossible to evaluate the reliability of these assessments in a remote environment,

given the lack of insight clinicians have into the testing environment and conditions. The VPC-5 provides a unique opportunity to remotely screen individuals before entering a clinical neuropsychological testing environment, as the webcam-based administration allows the clinician to have "eyes on the patient" and assess the validity of the results. Other remote testing options lack this feature (Bland et al. 2016; Maruff et al. 2009; Smith-Ray et al. 2016), positioning the VPC-5 as a more efficacious option for remote testing.

A final comment should be made about test scalability. Due to shorter duration, easier participant experience, and digital delivery, the scalability of the active VPC-5 exam may be greater than that of other cognitive assessments. The gold-standard assessments are reliable methods for detecting preclinical cognitive decline, but their scalability is limited because they need to be taken in person with a trained professional guiding them through various sets of instructions (Heaton et al. 2014). The VPC-5 represents a reliable, language-agnostic assessment requiring minimal instruction, with the ability to be completed anywhere (Bott et al. 2018). In all, this represents a promising way to screen and monitor memory function in healthy individuals over time without the need for repeated and/or burdensome in-clinic visits. Future studies will explore the utility of the VPC-5 in cognitively impaired populations as well as the efficacy for testing in remote (in-home) environments.

References

- Alzheimer's Association (2018) 2018 Alzheimer's disease facts and figures. *Alzheimers Dement* 14(3):367–429
- Alzheimer's Association (2015) 2015 Alzheimer's disease facts and figures. *Alzheimers Dement* 11(3):332–384. <https://doi.org/10.1016/j.jalz.2015.02.003>
- Best JR, Liu-Ambrose T, Boudreau RM, Ayonayon HN, Satterfield S, Simonsick EM et al (2016) An evaluation of the longitudinal, bidirectional associations between gait speed and cognition in older women and men. *J Gerontol Ser A Biol Sci Med Sci* 71(12):1616–1623. <https://doi.org/10.1093/gerona/glw066>
- Bland A, Roiser J, Mehta M, Schei T, Boland H, Campbell-Meiklejohn D et al (2016) EMOTICOM: a neuropsychological test battery to evaluate emotion. *Motiv Impulsiv Soc Cogn* 10. <https://doi.org/10.3389/fnbeh.2016.00025>
- Bott N, Madero EN, Glenn J, Lange A, Anderson J, Newton D, Brennan A, Buffalo EA, Rentz D, Zola S (2018) Device-

- embedded cameras for eye tracking–based cognitive assessment: validation with paper-pencil and computerized cognitive composites. *J Med Internet Res* 20(7):e11143. <https://doi.org/10.2196/11143>
- Bott NT, Lange A, Rentz D, Buffalo E, Clopton P, Zola S (2017) Web camera based eye tracking to assess visual memory on a visual paired comparison task. *Front Neurosci* 11 <https://doi.org/10.3389/fnins.2017.00370>
- Brustio PR, Magistro D, Zecca M, Rabaglietti E, Liubicich ME (2017) Age-related decrements in dual-task performance: comparison of different mobility and cognitive tasks. A cross sectional study. *PLoS ONE* 12(7):e0181698. <https://doi.org/10.1371/journal.pone.0181698>
- Crowe SF, Benedict T, Enrico J, Mancuso N, Matthews C, Wallace J (1999) Cognitive determinants of performance on the digit symbol-coding test, and the symbol search test of the Weis-III, and the symbol digit modalities test: an analysis in a healthy sample. *Aust Psychol* 34(3):204–210. <https://doi.org/10.1080/00050069908257455>
- Crutcher MD, Calhoun-Haney R, Manzanares CM, Lah JJ, Levey AI, Zola SM (2009) Eye tracking during a visual paired comparison task as a predictor of early dementia. *Am J Alzheimers Dis Other Dement* 24(3):258–266. <https://doi.org/10.1177/1533317509332093>
- Danckert SL, Craik FIM (2013) Does aging affect recall more than recognition memory? *Psychol Aging* 28(4):902–909. <https://doi.org/10.1037/a0033263>
- Donohue MC, Sperling RA, Salmon DP, Rentz DM, Raman R, Thomas RG et al (2014) The preclinical Alzheimer cognitive composite: measuring amyloid-related decline. *JAMA Neurol* 71(8):961–970. <https://doi.org/10.1001/jamaneurol.2014.803>
- Fagan JF (1970) Memory in the infant. *J Exp Child Psychol* 9(2): 217–226. [https://doi.org/10.1016/0022-0965\(70\)90087-1](https://doi.org/10.1016/0022-0965(70)90087-1)
- Glenn J, Vincenzo J, Canella C, Binns A, Gray M (2015) Glenn 2015 - Habitual and maximal dual-task gait speeds among sedentary, recreationally active, and masters athlete late-middle aged adults. *J Aging Phys Act* 23(3):433–437
- Goldberg TE, Harvey PD, Wesnes KA, Snyder PJ, Schneider LS (2015) Practice effects due to serial cognitive assessment: implications for preclinical Alzheimer's disease randomized controlled trials. *Alzheimer's & Dementia* 1(1):103–111. <https://doi.org/10.1016/j.dadm.2014.11.003>
- Hanninen T, Hallikainen M, Tuomainen S, Vanhanen M, Soininen H (2002) Prevalence of mild cognitive impairment: a population-based study in elderly subjects. *Acta Neurol Scand* 106(3):148–154. <https://doi.org/10.1034/j.1600-0404.2002.01225.x>
- Hansen DW, Pece AEC (2005) Eye tracking in the wild. *Comput Vis Image Underst* 98(1):155–181. <https://doi.org/10.1016/j.cviu.2004.07.013>
- Harvey PD (2012) Clinical applications of neuropsychological assessment. *Dialogues Clin Neurosci* 14(1):91–99
- Hausdorff JM, Rios DA, Edelberg HK (2001) Gait variability and fall risk in community-living older adults: a 1-year prospective study. *Arch Phys Med Rehabil* 82(8):1050–1056. <https://doi.org/10.1053/apmr.2001.24893>
- Heaton RK, Akshoomoff N, Tulsky D, Mungas D, Weintraub S, Dikmen S, Beaumont J, Casaletto KB, Conway K, Slotkin J, Gershon R (2014) Reliability and validity of composite scores from the NIH toolbox cognition battery in adults. *J Int Neuropsychol Soc* 20(06):588–598. <https://doi.org/10.1017/S1355617714000241>
- Kremen WS, Jak AJ, Panizzon MS, Spoon KM, Franz CE, Thompson WK, Jacobson KC, Vasilopoulos T, Vuoksimaa E, Xian H, Toomey R, Lyons MJ (2014) Early identification and heritability of mild cognitive impairment. *Int J Epidemiol* 43(2):600–610. <https://doi.org/10.1093/ije/dyt242>
- Lagun D, Manzanares C, Zola SM, Buffalo EA, Agichtein E (2011) Detecting cognitive impairment by eye movement analysis using automatic classification algorithms. *J Neurosci Methods* 201(1):196–203. <https://doi.org/10.1016/j.jneumeth.2011.06.027>
- Lin Y-T, Lin R-Y, Lin Y-C, Lee GC (2013) Real-time eye-gaze estimation using a low-resolution webcam. *Multimed Tools Appl* 65(3):543–568. <https://doi.org/10.1007/s11042-012-1202-1>
- Maruff P, Thomas E, Cysique L, Brew B, Collie A, Snyder P, Pietrzak R (2009) Validity of the CogStat Brief battery: relationship to standardized tests and sensitivity to cognitive impairment in mild traumatic brain injury, Schizophrenia, and AIDS Dementia Complex. *Arch Clin Neuropsychol* 24: 165–1178. <https://doi.org/10.1093/arclin/acp010>
- McCulloch KL, Mercer V, Giuliani C, Marshall S (2009) Development of a clinical measure of dual-task performance in walking: reliability and preliminary validity of the walking and remembering test. *J Geriatr Phys Ther* 32:8
- Montero-Odasso M, Casas A, Hansen KT, Bilski P, Gutmanis I, Wells JL, Borrie MJ (2009) Quantitative gait analysis under dual-task in older people with mild cognitive impairment: a reliability study. *J NeuroEng Rehabil* 6(1):35. <https://doi.org/10.1186/1743-0003-6-35>
- Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I et al (2005) The Montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment: MOCA: a Brief Screening Tool For MCI. *J Am Geriatr Soc* 53(4):695–699. <https://doi.org/10.1111/j.1532-5415.2005.53221.x>
- Petridis S, Giannakopoulos T, Spyropoulos CD (2013) Unobtrusive low cost pupil size measurements using web cameras. *ArXiv:1311.7327 [Cs]*. Retrieved from <http://arxiv.org/abs/1311.7327>
- Rose S (1980) Enhancing visual recognition memory in preterm infants. *Dev Psychol* 16(2):85–92. <https://doi.org/10.1037/0012-1649.16.2.85>
- Salthouse TA (1996) The processing-speed theory of adult age differences in cognition. *Psychol Rev* 103(3):403–428
- Sekuler R, Kahana MJ, McLaughlin C, Golomb J, Wingfield A (2005) Preservation of episodic visual recognition memory in aging. *Exp Aging Res* 31(1):1–13. <https://doi.org/10.1080/03610730590882800>
- Smith-Ray RL, Irmier C, Boulter K (2016) Cognitive training among cognitively impaired older adults: a feasibility study assessing the potential improvement in balance 4. <https://doi.org/10.3389/fpubh.2016.00219>
- Vivero V, Barreira N, Penedo MG, Cabrero D, Remeseiro B (2010) Directional gaze analysis in webcam video sequences. In: Campilho A, Kamel M (eds) *Image Analysis and Recognition*, vol 6112, pp 316–324. https://doi.org/10.1007/978-3-642-13775-4_32
- Ward A, Arrighi HM, Michels S, Cedarbaum JM (2012) Mild cognitive impairment: disparity of incidence and prevalence

- estimates. *Alzheimers Dement* 8(1):14–21. <https://doi.org/10.1016/j.jalz.2011.01.002>
- Yogev-Seligmann G, Hausdorff JM, Giladi N (2008) The role of executive function and attention in gait. *Mov Disord* 23(3):329–342. <https://doi.org/10.1002/mds.21720>
- Zola SM, Manzanares CM, Clopton P, Lah JJ, Levey AI (2013) A behavioral task predicts conversion to mild cognitive impairment and Alzheimer's disease. *Am J Alzheimers Dis Other Dement* 28(2):179–184. <https://doi.org/10.1177/1533317512470484>
- Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.