A short digital eye-tracking assessment predicts cognitive status among adults

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

Abstract Current cognitive assessments suffer from limited scalability and high user burden. This study aimed to (1) examine the relationship between a brief eye-tracking-based visual paired-comparison (VPC) and gold standard cognitive assessments, (2) examine longitudinal stability of the VPC task, (3) determine the ability of the VPC task to differentiate between cognitively normal (CN) individuals and individuals with mild cognitive impairment (MCI). Fifty-five adults ($n = 44$ CN, $n = 11$ MCI; 56.4 ± 26.7 years) were tested on two occasions, separated by at least 14 days. Visit 1 included VPC, Montreal Cognitive Assessment (MoCA), Digit Symbol Coding test (DSC), and NIH Toolbox Cognitive Battery (NIHTB-CB). Visit 2 included VPC, DSC, NIHTB-CB, and dual-task (DT). Significant differences existed between baseline VPC scores for CN and MCI groups ($p < .001$). VPC scores remained stable over time in both groups ($p < .05$). Significant associations existed between VPC and MoCA ($p < .01$), DSC ($p < .001$), and various NIHTB-CB subtests at both time points. The VPC test significantly predicts cognitive outcomes ($p < .05$), with age and VPC being the only significant predictors. Additionally, area under the curve (receiver operator characteristic = 0.80) for VPC scores demonstrated good classification accuracy. VPC reliably predicted cognitive status while remaining stable over time and displayed significant associations with gold standard cognitive assessments. VPC is a less burdensome and more scalable assessment than traditional tests, enabling longitudinal monitoring of cognitive status in resource-limited environments.

Keywords Cognition · Eye-tracking · Mild cognitive impairment · Dementia

Introduction

Approximately one in five adults over the age of 65 currently suffer from Alzheimer's disease (AD) or related dementias [1]. This number has increased drastically since 2000, with diagnoses rising 145% among Americans over the last 19 years [1]. AD accounts for one in three deaths among this population, making it the sixth leading cause of death in the USA. Not only is AD prevalent among older Americans, it is also one of the world's costliest diseases. In 2018, total lifetime costs were more than $350,000 per individual, translating to $2 trillion dollars of added burden on the healthcare system [1]. Assuming this trajectory continues, there will be an added $2.5 trillion dollars in long-term care costs by 2050 [1]. While there is no known cure for AD, delaying its onset by a single year could reduce future prevalence by 11% and cut annual health care costs by $219 billion [30]. Early detection of cognitive decline is valuable, allowing for faster intervention and treatment

J. L. Gills · J. M. Glenn · M. Gray (✉)
University of Arkansas, 1 University of Arkansas, Fayetteville, AR 72701, USA
e-mail: rgray@uark.edu

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

planning, both of which can reduce the costs connected with the disease.

Although 82% of adults over the age of 65 agree that testing their memory is important, only 16% receive regular cognitive assessments [1]. This lack of early screening limits effective treatment and intervention strategies. Thus, it is critical to administer cognitive assessments that have high reliability and low user burden to detect negative cognitive changes as early as possible. Current preclinical AD detection methods include neuroimaging and biomarker assessment, such as amyloid-β and tau proteins [17, 7], but these assessments are invasive, time-consuming, and expensive [1, 18]. A number of cognitive assessments are also available, but many of these tests are time-consuming, taxing to the user, and not widely scalable [4, 19, 26].

A unique cognitive assessment that mitigates some of these challenges is the visual paired comparison (VPC) task. The VPC task has been validated as a widely scalable discriminant digital cognitive assessment tool [4, 9, 31] for detecting cognitive decline. VPC tasks are entirely based on eye-tracking data collected by web cameras, allowing for assessment of patient populations with limited mobility and/or manual dexterity. VPC tasks assess declarative memory by comparing the proportion of time spent viewing a novel (not previously viewed) image compared to a previously viewed image [3, 11, 31]. Individuals with normal cognitive function spend a greater proportion of time gazing at the novel images than the familiar images. Conversely, individuals with impaired declarative memory spend equal amounts of time viewing both novel and previously viewed images.

A previously validated 30-min VPC task has been shown to reliably predict the onset of either mild cognitive impairment (MCI) or AD within 3–6 years [31]. This assessment originally used infrared eye-tracking equipment to analyze gaze, but Bott et al. [4] later validated the test with the use of readily available web cameras. The ubiquity of web cameras on most standard smart devices represents a more scalable technique for collecting population-level and longitudinal data on cognitive status. However, the passive nature of the 30-min VPC assessment and the associated time requirements are not optimal for rapid and scalable cognitive screening [4]. Thus, a brief 5-min VPC task utilizing an active paradigm allowing for a rapid and scalable cognitive assessment was developed.

VPC utilizes a short, active paradigm in which participants are given specific instructions to focus their gaze on the new/novel image before the testing phase begins. This allows the test to be repeated without influencing the results while concurrently reducing the burden on the user. While this test is validated to evaluate declarative memory function in healthy individuals [3, 4, 31], it remains unknown whether this exam can accurately discriminate between cognitively normal (CN) and cognitively impaired individuals. Therefore, the current investigation aimed to determine the ability of the active web-based 5-min VPC test to distinguish between cognitively normal and cognitively impaired adults. The three primary objectives of this study were to (1) examine the relationship between VPC and gold standard cognitive tests, (2) evaluate the stability of the VPC task over time, and (3) determine the ability of a brief VPC task to differentiate between cognitively normal and impaired individuals.

Methods

Participants and procedures

The current investigation was a prospective study, in which participants were tested twice with at least 14 days between testing sessions. This period was chosen in order to facilitate short-term test-retest reliability, which is standard for this type of analysis [23]. Eligible participants were between the ages of 18 and 46 years (younger adults) or above the age of 65 (older adults) and were able to read and understand English. Individuals were excluded from the study if they met any of the following criteria: diagnosed with attention-deficit/hyperactivity disorder (ADHD); known learning disability; disabling vision loss; inability to complete the calibration procedure for the web camera; or history of substance abuse, neurological illness, or psychiatric illness. Individuals with MCI, based on Montreal Cognitive Assessment (MoCA) scores < 26 were included in the present investigation; however, inclusion was only granted after consent was provided from the individual's primary caregiver. A group of 59 participants were recruited through flyers, website announcements, and word of mouth. Of the 59 recruited participants, 55 completed all assessments. The four individuals

that did not complete all exams during both testing sessions were excluded from all analyses.

Participants visited the study site on two occasions. During the first visit, participants signed the informed consent document, completed a medical history questionnaire, and completed the MoCA, Digit Symbol Coding (DSC) Test, NIH Toolbox Cognitive Battery (NIHTB-CB), and the VPC test. During the second visit, participants completed the VPC, DSC, NIHTB-CB, dual-task, and provided biometric measurements (height, weight, and body mass index). The NIHTB-CB and VPC tasks were randomized within participants from visit 1 to visit 2 to reduce the learning effect and ensure score differences were not due to cognitive fatigue. The cognitive tests were performed in the same room for both visits. The designated testing room was quiet and secluded from distracting movements and sounds.

Biometric assessments

Biometric assessments included height, weight, and body mass index. Height was measured with a standing stadiometer (Seca; Hamburg, Germany). During this assessment, participants 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 (Detecto, Webb City, MO). 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 attention and concentration, executive function, memory, language, visuoconstructional skills, conceptual thinking, mental calculation, and orientation. Details of this assessment have been described in detail elsewhere [22]. Briefly, MoCA is comprised of 11 tasks that provide an overall score on cognitive function. Scores range from 0 to 30, with higher scores representing greater cognitive ability. Adults with a score of 26 or greater were categorized as cognitively normal and individuals with scores below 26 were

categorized as MCI. If participants were deemed cognitively impaired by MoCA and/or a prior clinical diagnosis, informed consent from their legal guardian was obtained.

Digit Symbol Comparison The DSC test is a paper-pencil test assessing processing speed, working memory, visuospatial processing, and attention [10]. This assessment has been described in detail elsewhere. Briefly, 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 participant has 90 s to pair specific numbers with given 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 [8, 10] and predicting cognitive disorders [2].

NIHTB-CB The NIHTB-CB is a computerized cognitive composite battery assessing various functional domains. The four tests comprising the NIHTB-CB were completed on an iPad (11.4, Apple Inc., Cupertino, CA). These assessments have been described in detail elsewhere. First, the Flanker Inhibitory Control and Attention Test (Flanker) measured inhibitory control and attention. The Flanker required the participant to focus on a particular stimulus while inhibiting attention to the stimuli surrounding it. Next, the Dimension Change Card Sort Test (DCCS) evaluated cognitive flexibility and attention. Two target pictures were presented that vary along two dimensions (e.g., shape and color). Participants were 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) measured processing speed. The test required participants to discern whether two side-by-side pictures were identical or different. Finally, the Picture Sequence Memory Test (PSMT) measured episodic memory. Sequences of pictured objects and activities were presented in a particular order and participants were 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 [15]. NIHTB-CB normative scores were used for all analyses. Per NIH recommendations, a different version of each battery was used for each visit.

VPC assessment This investigation utilized a previously validated 5-min VPC task (Neurotrack Technologies, Inc., Redwood City, CA). The VPC task was completed on a laptop computer equipped with a factory-installed web camera (MacBook Air, Apple Inc., Cupertino, CA). The VPC test construction is explained in detail elsewhere [4]. In summary, VPC tasks use eye-tracking data to assess declarative memory function [4, 9, 31]. The VPC test used in this study is an active paradigm. At the beginning of the test, participants were instructed to remember the images from the familiarization phase and spend more time looking at the novel images during the test phase, as opposed to the familiar images. During the familiarization phase of VPC, participants were presented with 20 pairs of identical visual stimuli displayed for 5 s each. After the familiarization phase, a continuous and cumulative delay occurred across each test trial. Before the testing phase began, participants were instructed to look at the novel or unfamiliar image. During the test phase, participants were presented with pairs of 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 declarative memory function [4]. Eye movements were tracked throughout the task and scored. Detailed scoring information is published elsewhere [4]. Separate versions of the VPC assessment were administered during visits 1 and 2 to prevent a potential learning effect. In addition to proportion of time spent viewing the novel image, additional extraction of specific eye-tracking features provided the opportunity to characterize and investigate differences between populations. Interstimulus oscillations (ISO) is a metric defined as the number of saccades a participant makes between competing stimuli. The ISO count per trial is averaged across each of the 20 test trials and a mean ISO is generated. Similar saccadic-based measures have been used in other eye-tracking-based cognitive paradigms and have shown the ability to distinguish cognitive status including mild cognitive impairment, Alzheimer's disease and Parkinson's disease [6, 13, 16, 27, 28].

Dual-task The dual-task walking assessment evaluated attention and executive function [5, 29]. This assessment has been described in detail elsewhere [12]. Dual-task assessments vary in protocol, but for the purposes of this study, participants were instructed to walk 20 m at their usual speed while time was recorded by the researcher. There was a 5-m distance before and after the 10-m distance to account for acceleration and deceleration [12]. 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 [14]. A random 3-digit number between 100 and 999 was selected and participants were instructed to subtract three from each number while performing each walking condition. Four testing trials were completed, two at usual speed (dual-task habitual speed; DT-HS) and two at their maximal speed (dual-task maximal speed; DT-MS). Dual-task decrement was calculated as the difference between the walk trial while performing serial subtractions and the trial without subtraction. The walking speed trials (DT-HS and DT-MS) 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 [20, 21].

Data analysis

Participant characteristics were assessed between cognitive groups and sexes with an independent samples *t* test. Spearman's rho correlation coefficient analyses were used to determine associations between the VPC and the other standard cognitive tests (DSC, NIHTB-CB, and dual-task). Independent samples *t* tests were used to assess differences between cognitive group's VPC scores. Two one-sided tests (TOST) or equivalence testing was utilized to ensure the VPC paradigm was stable across testing trials for both CN and MCI participants. TOST equivalence bounds were set at 0.8 and 1.25 assuming a lognormal distribution. Additionally, logistic regression was utilized to determine how well the VPC task predicted cognitive outcomes, with cognitive scores as the dependent variable (CN vs. MCI) and VPC score, age, sex, and education as predictor variables (model 1). To account for the aging process,

another logistic regression was employed using only older adults that were cognitively intact and impaired (model 2). MoCA scores were used as the dependent variable and VPC score, age, sex, and education as predictor variables. Lastly, a receiver operating characteristic (ROC) classification was utilized to examine the sensitivity and specificity of the overall VPC score and interstimulus oscillations data in classifying subjects in cognitive groups based on MoCA scores. Statistical outliers were determined through box and whisker plots. Data points were removed if they were beyond 3 times the interquartile range. Statistical significance was set at $\alpha = .05$ for all analyses.

Results

Participant characteristics

The overall sample was 69% female ($n = 39$ participants) with over 74% of participants being college or trade school graduates (Table 1). There were no significant differences observed in VPC scores between sexes ($p = .12$). The mean age of the 55 participants was 55.9 ± 26.8 years (range 21–89 years; Table 1). Not all VPC eye-tracking data were scored due to technical difficulties (recording quality, network connectivity, glare from glasses, and/or low light in the testing room). Of the 55 participants who completed the assessments, 53 of the visit 1 VPC tests (two MCI subjects excluded) were analyzed and 51 of the visit 2 VPC tests (one MCI subject and five cognitively intact subjects excluded) were examined. Statistical outliers were removed from all analyses: two MCI participants and one CN participant were removed from the logistic regression models, one MCI participant was removed from the validity and reliability testing analyses, and one MCI participant was removed from the Pearson's correlation analysis (MoCA data). The second testing session took place an average of 23.7 ± 12.4 days after the initial testing session.

Relationships between VPC and cognitive assessments

Cognitive assessments MoCA scores were significantly correlated with VPC results at visit 1 ($r = .47$; $p < .001$; Fig. 1a) and visit 2 ($r = .47$; $p = .001$; Fig. 1b). DSC scores were also significantly correlated with VPC scores at visit 1 ($r = .57$; $p < .001$; Fig. 1c) and visit 2 ($r = .61$; $p < .001$; Fig. 1d). Additionally, this study

examined the relationships between the VPC task and domain-specific cognitive assessments. The NIHTB-CB age-based normative percentile scores for each task were used as a basis of comparison for VPC test performance. At visit 1, significant associations were found between VPC and DCCS ($r = .32$; $p = .02$), PSPAC ($r = .28$; $p = .05$; Fig. 1e), and PSMT ($r = .33$; $p = .02$) but not the Flanker ($r = .25$; $p = .08$). Visit 1 correlations are shown in Table 2. At visit 2, significant relationships were identified between VPC and PSPAC ($r = .53$; $p < .001$; Fig. 1f), and PSMT ($r = .44$; $p = .003$). A trend was observed between VPC scores and Flanker ($r = .29$; $p = .06$), as well as VPC scores and the DCCS ($r = .29$; $p = .06$) at visit 2. Correlations for visit 2 data are found in Table 2.

Dual-task Lastly, we evaluated functional measures of cognition by comparing DT-HS and DT-MS scores to VPC at visit 1 and visit 2. Significant correlations were found between VPC visit 1 and DT-MS ($r = -.34$; $p = .02$) and DT-HS ($r = -.43$; $p = .003$). Moreover, results revealed significant associations between visit 2 VPC scores and DT-MS ($r = -.39$; $p = .007$), as well as VPC scores and DT-HS ($r = -.40$; $p = .005$).

Validity and reliability of the VPC assessment

This sample was separated into two cohorts: CN ($n = 44$ participants) and MCI ($n = 11$ participants) (Table 1). An independent samples t test showed significant differences ($p < .001$) in VPC baseline (visit 1) scores (Fig. 2) and mean VPC interstimulus oscillation count ($p = .03$) (Fig. 2b). We analyzed reliability of our data using TOST analysis. We found that VPC scores for both CN ($p < .001$) and MCI ($p = .006$) participants from baseline to visit 2 were statistically equivalent (Fig. 2c). The mean ratio of VPC baseline scores and visit 2 scores was $1.0\% \pm 2.0\%$ for both CN and MCI participants. This demonstrates that VPC scores for baseline and visit 2 for both cognitive groups were within 2%, indicating statistical equivalence.

Logistic regression classification

CN participants had greater VPC values when compared to MCI participants ($83.84 \pm 7.77\%$ vs. $66.64 \pm 12.9\%$, respectively; $p = .01$; Fig. 2). Model 1 revealed that VPC scores, age, sex, and education status explained 57% of the variance in the MoCA cognitive classification, with

Table 1 Characteristics of the study cohort

	Cognitively intact participants ($n = 44$)	MCI participants ($n = 11$)	p values
Average age (SD)	50.32 years (27.6)	78.6 years (6.4)	$p < .001$
Sex			$p < .001$
Female	65.9%	90.9%	
Education			$p = .51$
High school graduate	4.5%	9.1%	
Some college	20.5%	18.2%	
College graduates or higher	75.0%	72.7%	
Race			$p = .55$
European-American	88.6%	81.8%	
Other	11.4%	18.2%	
Biometric			
Height (SD)	169.1 cm (8.4)	157.9 cm (11.5)	$p = .001$
Weight (SD)	74.2 kg (15.7)	65.5 kg (15.3)	$p = .12$
BMI (SD)	26.0 kg/m^2 (5.4)	26.2 kg/m^2 (4.3)	$p = .88$

Values are presented at means (SD). p values represented independent samples t test differences between groups. *BMI* body mass index

the model properly identifying 88.4% (95% of CN group; 56% of MCI group) of adults into either the CN or the MCI group ($p < .001$). Among all individuals, the VPC score ($p = .008$) was the only significant predictor of MoCA classification. The cutoff value for the VPC score was 0.75 or 75%.

Model 2 used older adults only classified by cognitive status: CN ($n = 20$ participants) and MCI ($n = 11$ participants). Model 2 containing all predictors was statistically significant ($p = .02$), indicating that the model was able to distinguish between older adults with and without MCI. The model as a whole explained 48% of MoCA cognitive classifications, with the model properly identifying 80% (91% of CN group; 56% of the MCI group) of adults in either group. VPC performance ($p = .01$) was a significant predictor of MoCA classification in older adults.

Receiver operator characteristic classification

Finally, we examined the sensitivity and specificity of the overall VPC score as well as eye-tracking-based interstimulus oscillations in classifying cognitive impairment as defined by the MoCA threshold of 26. Overall VPC score area under the ROC was 0.80, with a sensitivity of 90% and a specificity of 65% (Fig. 3a). Interstimulus oscillations area under the ROC was 0.75, with a sensitivity of 80% and a specificity of 71%. (Fig. 3b).

Discussion

The primary purposes of this investigation were to (1) examine the relationship between VPC and gold standard cognitive tests, (2) evaluate the stability of the VPC task over time, (3) determine the ability of a brief VPC task to differentiate between cognitively normal and impaired individuals. Our results demonstrate the VPC task is a useful screening tool for cognitive impairment and can be used as a widely scalable and accurate assessment of memory function. Moreover, the VPC task successfully discriminates between cognitively impaired and cognitively normal individuals, regardless of age, and is stable over time. Lastly, the VPC task is significantly correlated with several traditional in-clinic cognitive assessments.

This current study demonstrates that the VPC task can predict cognitive status, as determined by MoCA scores. Crutcher and associates [9] found similar results, albeit using a longer VPC test, between the control and MCI group. Their results demonstrated that the control group viewed the novel images 74% of the time, while viewing the familiar image 26% of the time. Conversely, they found the MCI group spent only 53% of the total time viewing the novel images and 47% of the total time looking at the familiar stimuli. In the current investigation, CN participants viewed the novel stimuli 84% of the time and the familiar images 16% of the time. The MCI participants viewed the novel images 67% of the
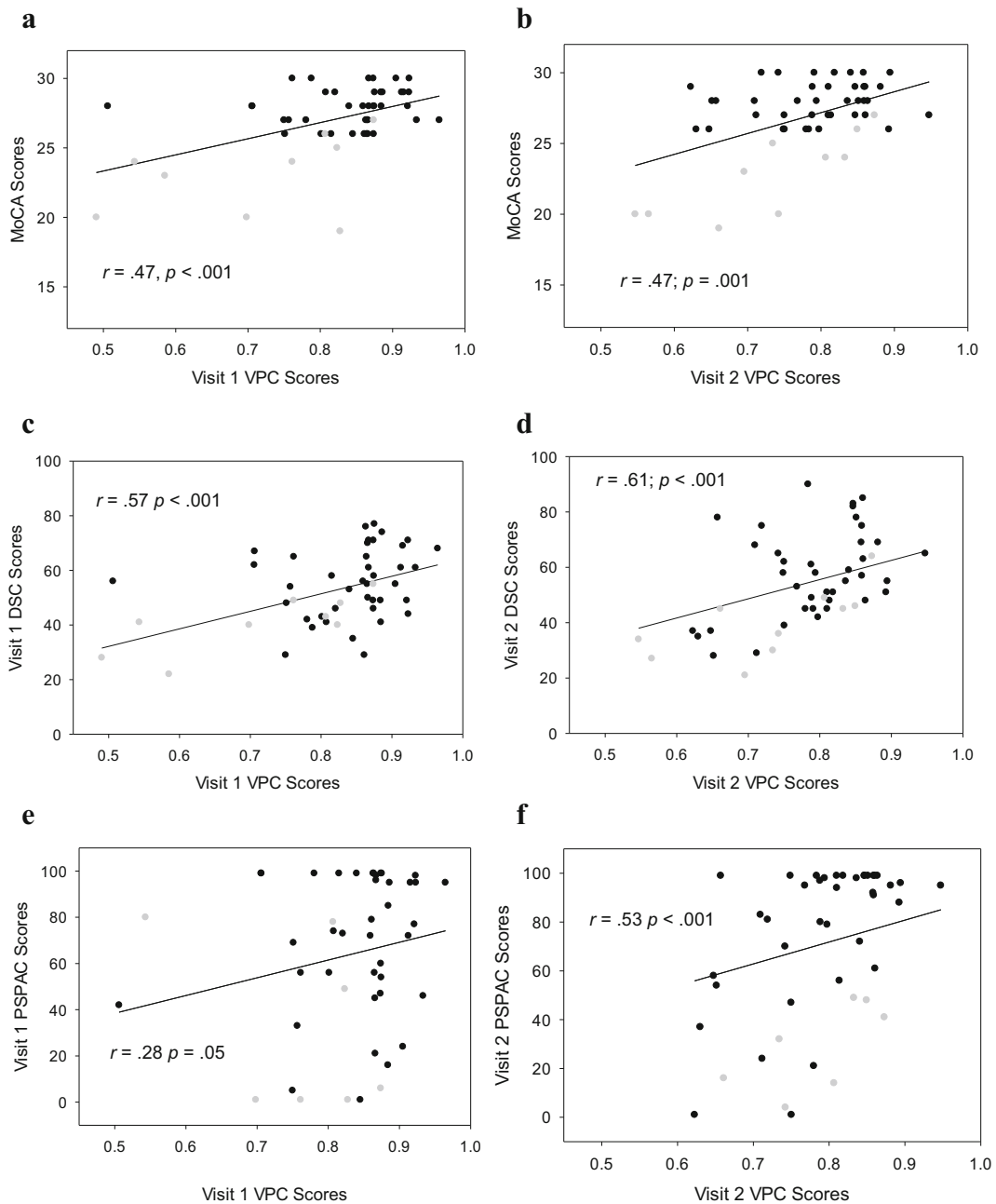
Fig. 1 **a–f** Correlations between gold standard cognitive assessments and visual paired comparison scores during visits 1 and 2. Gray-filled circles = MCI subjects; black-filled circles = cognitively intact subjects. Spearmans's rho correlation was utilized to determine this outcome

total time and spent 33% of total time viewing the familiar stimuli. These data suggest appropriate convergent validity between the active and passive paradigms.

This study also demonstrates the role of novel eye-tracking feature metrics in differentiating cognitive status. In the current investigation, the CN participants had a mean of 1.9 interstimulus oscillations per trial. The

MCI participants had a mean of 2.48 interstimulus oscillations per trial. Despite the relatively small sample size in the current study, the effect size of this difference is large (Cohen's $d = 0.80$).

ROC analyses demonstrated good classification accuracy of cognitively intact and cognitively impaired participants utilizing a MoCA cutpoint of 26.

Table 2 Correlations between VPC test data and cognitive assessments

	VPC—visit 1	VPC—visit 2
Flanker	.25	.29
DCCS	.32*	.29
PSPAC	.28*	.53*
PSMT	.33*	.44*
MoCA	.47*	.47*
DSC	.57*	.61*
DT-HS	− .45*	− .39*
DT-MS	− .43*	− .40*

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

Interestingly, utilizing ISO as the classification variable provided adequate classification between cognitively normal and cognitively intact participants. In future studies, composite classification variables in larger samples may provide enhanced classification accuracy.

We hypothesized participants in the current investigation would have higher viewing times for novel stimuli and lower viewing times for familiar images for both groups due to the active nature of the task ([9] utilized a passive paradigm). Unlike the original passive version, the current active version of the test provides instructions on which images to focus on before beginning the testing phase. The current VPC task was built this way to allow for repeat testing, whereas the longer passive test used in the previous study is only meant to produce a baseline score.

In addition to being able to discriminate between cognitively normal and cognitively impaired
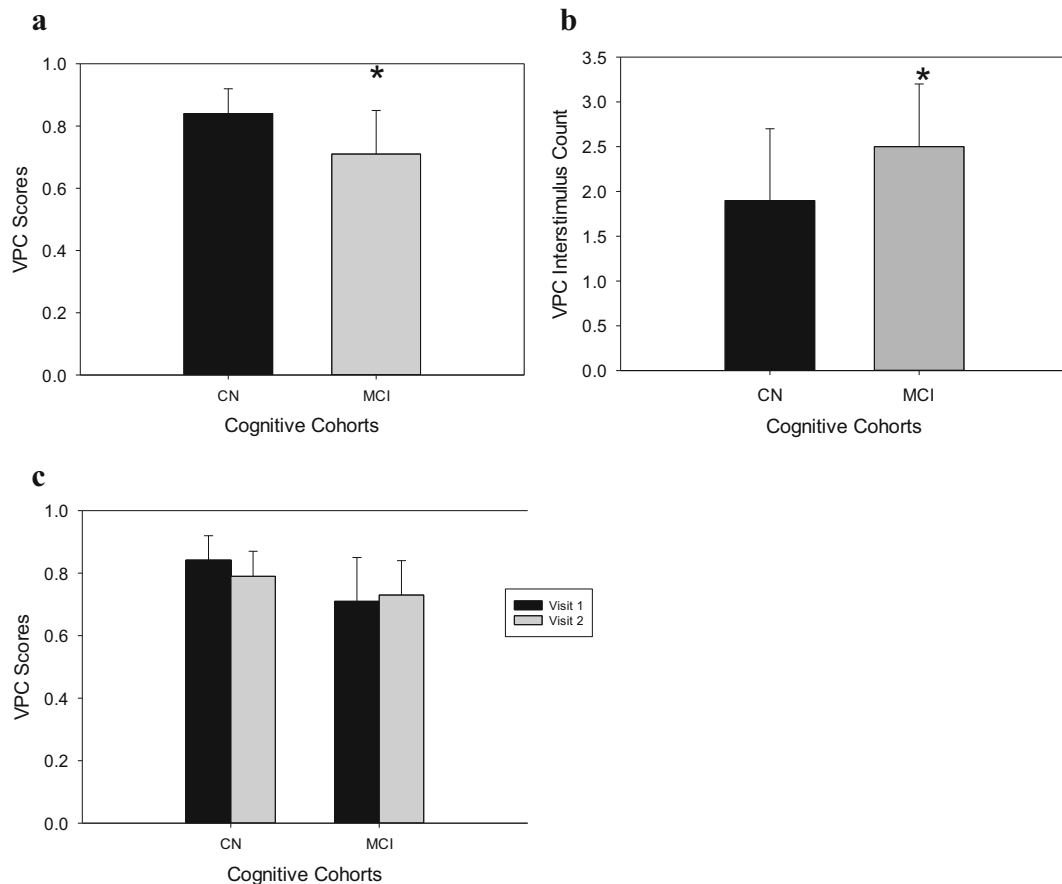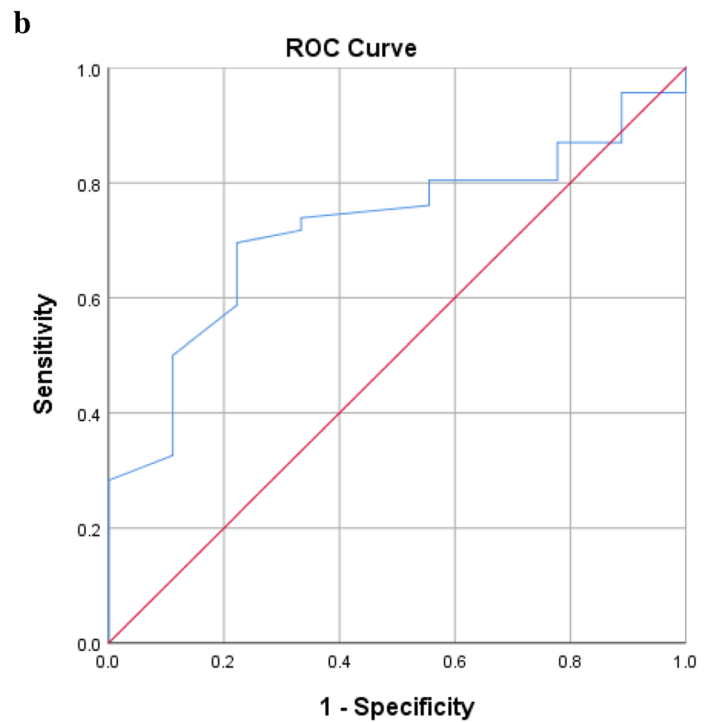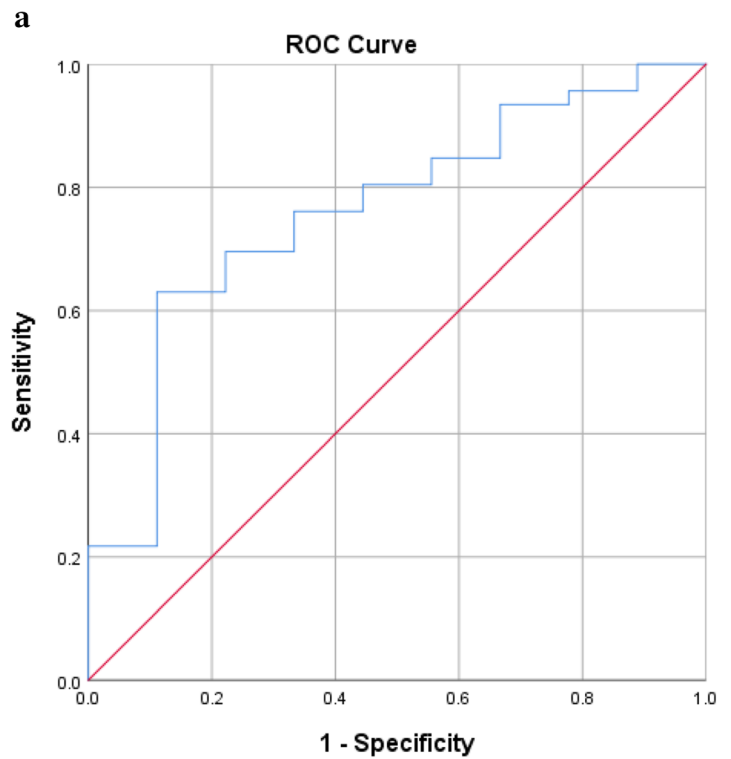
Fig. 2 **a–c** VPC overall and oscillation scores between MCI and CN individuals. **a**, **b** An independent samples *t* test was used to determine this outcome. c Two one-sided *t* tests were used to determine this outcome. CN, cognitively normal individuals; MCI, individuals with mild cognitive impairment. * indicates significant difference from CN and MCI groups; *p < .05*

Fig. 3 **a** Overall VPC score ROC analysis. **b** Interstimulus oscillations ROC analysis

individuals, the VPC test proved to be a reliable test by maintaining similar results for each cohort from visit 1 to visit 2 while using different forms of the test. Gills and colleagues revealed similar results; however, that analysis did not include MCI participants. Moreover, the scalability of this version of the VPC task is higher due to the shorter duration of the test, which lessens the burden on users. With digital cognitive examinations increasing in popularity [3, 18], a valid and reliable version of the VPC task with an active paradigm may be a valuable screening tool for assessing and tracking cognitive status over time.

These data from the present investigation also suggest the VPC test has significant relationships with traditional gold standard paper-pencil and digital cognitive assessments. In particular, the correlations with the MoCA are noteworthy as the MoCA is a standard cognitive assessment evaluating multiple domains of cognition, often used to determine if an individual may have mild or severe cognitive impairment. Nevertheless, it should be noted that this study compared the paper-pencil administration of the MoCA to the digital tasks and future research comparing the digital delivery of the MoCA to the VPC is warranted. Similar to previous studies in healthy adults [4], this study found that the VPC task had significant correlations with individual NIHTB-CB tasks (i.e., those measuring inhibitory control and attention, processing speed, and visual episodic memory). These outcomes are similar to previous research findings showing that VPC tests have significant associations with visual episodic memory [4] and processing speed. Importantly, both domains are known to influence a range of cognitive task/abilities, including executive functioning tasks [24, 25]. When compared to previous investigations, the correlations in the current study may be increased due to the addition of MCI participants (Table 2) as previous study only included healthy older adults [4].

The final task was a functional measure (dual-task), which is closely related to executive functioning along with the VPC task [4, 12]. A significant negative relationship was demonstrated between the dual-task assessments and the VPC assessments. The negative correlation indicates participants slowed their gait speed in order to successfully complete the cognitive task [20, 21].

Although this functional exam is challenging for most participants, these results reveal this task provides a greater struggle for individuals with MCI. This was exhibited in our results by an overall 2% increase in time to complete the DT-HS from habitual speed and a 5% increase while examining DT-MS from maximal walking speed. Moreover, in cognitively normal gait analyses, there was a 1% and 3% increase in time to complete the task for normal and maximal walking speeds, respectively, to dual-task trials. Conversely, in individuals with cognitive impairment, there was a 7% and 13% time increase to complete DT-HS and DT-MS from normal walking tasks. This indicates that negative changes in functional fitness may be a harbinger for cognitive decline in older adults and further demonstrates the importance for aging populations to maintain functional fitness levels to potentially delay the cognitive aging process [12, 20, 21].

Any digital examination is susceptible to technological issues, such as a weak WiFi signal and user errors. For this current investigation, one limitation was data quality. Eight participants' assessments (two MCI participants at visit 1, five CN participants at visit 2, and one MCI participant at visit 2) were omitted from VPC scoring due to electronic errors, glare from glasses, or low lighting in the room. These participants were still included in every other data point besides the VPC test because they completed both visits at the testing center. Furthermore, all assessments were completed in a research setting instead of the participant's home to ensure compliance and gather high-quality data, which limits the ecological validity of the results. We had a limited number of MCI participants (23%) compared to CN participants (77%). Future studies should examine larger sample sizes of these individuals and further assess differences between groups in cognitive performance over time. Lastly, subjects with diagnosed psychiatric illnesses such as depression were excluded but subjects may have had depressive symptomology that was not screened for and possibly could have been included in the sample.

In conclusion, the VPC task appears to be an appropriate test to discriminate cognitive status and may be a useful brief and scalable screening tool for cognitive impairment. The VPC task successfully discriminated

between CN and impaired individuals while demonstrating high test-retest reliability. Lastly, the inclusion of MCI participants in this study strengthens the previously reported relationships between the VPC and standard digital and paper-pencil cognitive assessments. The VPC task can be taken remotely in a user's home, which enables more scalable longitudinal assessment and tracking of cognitive changes.

References

1. Alzheimer's Association. 2019 Alzheimer's disease facts and figures. Alzheimers Dement. 2019;15(3):321–87.

2. Best JR, Liu-Ambrose T, Boudreau RM, Ayonayon HN, Satterfield S, Simonsick EM, et al. An evaluation of the longitudinal, bidirectional associations between gait speed and cognition in older women and men. J Gerontol A Biol Sci Med Sci. 2016;71(12):1616–23. https://doi.org/10.1093/gerona/glw066.

3. Bott NT, Lange A, Rentz D, Buffalo E, Clopton P, Zola S Web camera based eye tracking to assess visual memory on a visual paired comparison task Front Neurosci. 2017; 11. https://doi.org/10.3389/fnins.2017.00370.

4. Bott N, Madero EN, Glenn J, Lange A, Anderson J, Newton D, et al. Device-embedded cameras for eye tracking–based cognitive assessment: validation with paper-pencil and computerized cognitive composites. J Med Internet Res. 2018;20(7):e11143. https://doi.org/10.2196/11143.

5. Brustio PR, Magistro D, Zecca M, Rabaglietti E, Liubicich ME. Age-related decrements in dual-task performance: comparison of different mobility and cognitive tasks. A cross sectional study. PLoS One. 2017;12(7):e0181698. https://doi.org/10.1371/journal.pone.0181698.

6. Chehrehnegar N, Nejati V, Shati M, Esmaeili M, Rezvani Z, Haghi M, et al. Behavioral and cognitive markers of mild cognitive impairment: diagnostic value of saccadic eye movements and Simon task. Aging Clin Exp Res. 2019;31(11):1591–600. https://doi.org/10.1007/s40520-019-01121-w.

7. Chen-Chen T, Jin-Tai Y, Lan T Biomarkers for preclinical Alzheimer's disease. J Alzheimer's Dis. 2014;42(4):1051–1069.

8. Crowe SF, Benedict T, Enrico J, Mancuso N, Matthews C, Wallace J. Cognitive determinants of performance on the digit symbol-coding test, and the symbol search test of the Wais-III, and the symbol digit modalities test: an analysis in a healthy sample. Aust Psychol. 1999;34(3):204–10. https://doi.org/10.1080/00050069908257455.

9. Crutcher MD, Calhoun-Haney R, Manzanares CM, Lah JJ, Levey AI, Zola SM. Eye tracking during a visual paired comparison task as a predictor of early dementia. Am J Alzheimers Dis Other Dement. 2009;24(3):258–66. https://doi.org/10.1177/1533317509332093.

10. Donohue MC, Sperling RA, Salmon DP, Rentz DM, Raman R, Thomas RG, et al. The preclinical Alzheimer cognitive composite: measuring amyloid-related decline. JAMA Neurol. 2014;71(8):961–70. https://doi.org/10.1001/jamaneurol.2014.803.

11. Fagan JF. Memory in the infant. J Exp Child Psychol. 1970;9(2):217–26. https://doi.org/10.1016/0022-0965(70)90087-1.

12. Glenn JM, Vincenzo J, Canella CK, Binns A, Gray M. Habitual and maximal dual-task gait speeds among sedentary, recreationally active, and masters athlete late middle-aged adults. J Aging Phys Act. 2015;23(3):433–7. https://doi.org/10.1123/japa.2014-0069.

13. Hannula DE, Althoff RR, Warren DE, Riggs L, Cohen NJ, Ryan JD. Worth a glance: using eye movements to investigate the cognitive neuroscience of memory. Front Hum Neurosci. 2010;4:166. https://doi.org/10.3389/fnhum.2010.00166.

14. Hausdorff JM, Rios DA, Edelberg HK. Gait variability and fall risk in community-living older adults: a 1-year prospective study. Arch Phys Med Rehabil. 2001;82(8):1050–6. https://doi.org/10.1053/apmr.2001.24893.

15. Heaton RK, Akshoomoff N, Tulsky D, Mungas D, Weintraub S, Dikmen S, et al. Reliability and validity of composite scores from the NIH toolbox cognition battery in adults. J Int Neuropsychol Soc. 2014;20(06):588–98. https://doi.org/10.1017/S1355617714000241.

16. Henderson JM, Hollingworth A. Eye movements and visual memory: detecting changes to saccade targets in scenes. Percept Psychophys. 2003;65(1):58–71. https://doi.org/10.3758/BF03194783.

17. Ho JK, Nation DA, for the Alzheimer's Disease Neuroimaging Initiative. Neuropsychological profiles and trajectories in preclinical Alzheimer's disease. J Int Neuropsychol Soc. 2018;24(7):693–702. https://doi.org/10.1017/S135561771800022X.

18. Koo BM, Vizer LM. Mobile technology for cognitive assessment of older adults: a scoping review. Innov Aging. 2019;3(1). https://doi.org/10.1093/geroni/igy038.

19. Lagun D, Manzanares C, Zola SM, Buffalo EA, Agichtein E. Detecting cognitive impairment by eye movement analysis using automatic classification algorithms. J Neurosci Methods. 2011;201(1):196–203. https://doi.org/10.1016/j.jneumeth.2011.06.027.

20. McCulloch KL, Mercer V, Giuliani C, Marshall S. 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. 2009;32(1):2–9.

21. Montero-Odasso M, Casas A, Hansen KT, Bilski P, Gutmanis I, Wells JL, et al. Quantitative gait analysis under dual-task in older people with mild cognitive impairment: a reliability study. J Neuroeng Rehabil. 2009;6(1):35. https://doi.org/10.1186/1743-0003-6-35.

22. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc. 2005;53(4):695–9. https://doi.org/10.1111/j.1532-5415.2005.53221.x.

23. O'Brien AM, Casey JE, Salmon RM. Short-term test–retest reliability of the ImPACT in healthy young athletes. Appl Neuropsychol Child. 2018;7(3):208–16. https://doi.org/10.1080/21622965.2017.1290529.

24. Rose S. Enhancing visual recognition memory in preterm infants. Child Development. 1980;16(2), 85–92. https://doi.org/10.1037/0012-1649.16.2.85.

25. Salthouse TA. The processing-speed theory of adult age differences in cognition. Psychol Rev. 1996;103(3):403–28.

26. Wadsworth HE, Dhima K, Womack KB, Hart J, Weiner MF, Hynan LS, & Cullum CM Validity of teleneuropsychological assessment in older patients with cognitive disorders. Arch Clin Neuropsychology, 2018;33(8):1040–1045.

27. Waldthaler J, Tsitsi P, Svenningsson P. Vertical saccades and antisaccades: complementary markers for motor and cognitive impairment in Parkinson's disease. NPJ Parkinsons Dis. 2019;5(1):1–6. https://doi.org/10.1038/s41531-019-0083-7.

28. Wilcockson TD, Mardanbegi D, Xia B, Taylor S, Sawyer P, Gellersen HW, et al. Abnormalities of saccadic eye movements in dementia due to Alzheimer's disease and mild cognitive impairment. Aging (Albany NY). 2019;11(15):5389. https://doi.org/10.18632/aging.102118.

29. Yogev-Seligmann G, Hausdorff JM, Giladi N. The role of executive function and attention in gait. Mov Disord. 2008;23(3):329–42. https://doi.org/10.1002/mds.21720.

30. Zissimopoulos J, Crimmins E, St.Clair P. The value of delaying Alzheimer's disease onset. Forum Health Econ Policy. 2014; 0(0). https://doi.org/10.1515/fhep-2014-0013.

31. Zola SM, Manzanares CM, Clopton P, Lah JJ, Levey AI. A behavioral task predicts conversion to mild cognitive impairment and Alzheimer's disease. Am J Alzheimers Dis Other Dement. 2013;28(2):179–84. https://doi.org/10.1177/1533317512470484.