Journal of Alzheimer's Disease 59 (2017) 1227–1235
DOI 10.3233/JAD-170295
IOS Press

Validity of a Newly Developed Measure of Memory: Feasibility Study of the Virtual Environment Grocery Store

Thomas D. Parsons* and Michael Barnett
Department of Psychology, University of North Texas, Denton, TX, USA

Handling Associate Editor: Julie Robillard

Accepted 9 June 2017

Abstract. Virtual reality-based neuropsychological assessments proffer the potential to address the limited ecological validity of pen-and-paper measures of memory. To investigate the construct validity of a newly developed virtual reality measure of memory, the Virtual Environment Grocery Store (VEGS), traditional neuropsychological measures of memory and executive functioning were administered to 48 older adults and 55 young adults. Performances on the VEGS memory tasks and the traditional neuropsychological assessments of memory were positively correlated, indicating that memory for VEGS content was similar to memory for traditional paper-and-pencil measures. The older adults performed significantly worse than young adults on the VEGS and the California Verbal Learning Test, but the DKEFS Color-Word Interference failed to differentiate the groups. Furthermore, significant differences were found between groups for the VEGS memory and multitasking measures. The VEGS has the advantage over traditional measures of providing objective measurement of individual components of memory in simulations of everyday activities.

Keywords: Aging, episodic memory, neuropsychological tests, prospective memory, validation studies, virtual reality

INTRODUCTION

Declarative and prospective memory abilities are important neurocognitive processes that undergo a protracted course throughout the lifespan and exhibit decline even in healthy aging. The neural underpinnings of episodic memory involve extensive cortical networks (prefrontal cortex, lateral temporal neocortex, medial temporal lobe, and posterior parietal regions; see [1]). The prefrontal cortex (especially rostral) and medial temporal lobe memory system (including the hippocampus) subvene episodic and prospective memory [2]. Hippocampal volume reductions and associated reductions in memory performance have been found in healthy aging [3].

While semantic memory is specific to storage of data not tied to specific times, places, or events, episodic memory involves storage of distinctive experiences or events associated with explicit places and times. Moreover, episodic memory is susceptible to declines in performance with normal aging. Another area of memory decline that may have significant consequences for older adults is their ability to plan and remember to perform previously planned activities at a specific point in time [4]. Examples of such prospective memory activities include remembering to listen for one's prescription while shopping or remembering to go to a coupon machine to check for potential savings. Hence, prospective memory can be understood as a central component for performing activities of daily living [5].

*Correspondence to: Thomas D. Parsons, PhD, Associate Professor of Psychology; Director, Computational Neuropsychology and Simulation (CNS) Lab (http://psychology.unt.edu/cns-lab-parsons); Fellow, National Academy of Neuropsychology; Department of Psychology, University of North Texas, 1155 Union Circle #311280, Denton, TX 76203, USA. Tel.: +1 940 565 4329; Fax: +1 940 369 8672; E-mail: Thomas.Parsons@unt.edu.

ISSN 1387-2877/17/$35.00

Although neuropsychologists are increasingly being asked to assess an individual's everyday memory abilities, research suggests that people expected to perform poorly on neuropsychological tests may perform within normal limits [6]. These findings may reflect the limited generalizability of traditional paper-and-pencil neuropsychological assessments findings to task demands found in everyday activities [7, 8]. Burgess and colleagues (2006) have pointed out the need for neuropsychological assessments that incorporate more complex and lifelike scenarios that can tax multiple cognitive processes simultaneously [9]. For example, the Multiple Errands Test [6, 9] and the Instrumental Activity of Daily Living Profile [10] require participants to perform errands in a real shopping center. A limitation of these studies is that they can be expensive, time consuming, and difficult to replicate across testing sessions [11, 12].

Virtual reality (VR) is a relatively new technology that aims to develop and implement ecologically valid and interactive three-dimensional simulations of real world scenarios [13, 14]. VR-based neuropsychological assessments offer the potential to address the ecological validity limitations found in traditional neuropsychological measures [15]. These virtual environments are increasingly being used for assessment of memory in experimentally controlled simulations of everyday activities [16, 17]. Moreover, virtual environments can offer a balance of experimental control and simulation of everyday memory [18, 19].

Various virtual environment-based neuropsychological assessments have been used to assess everyday prospective and episodic memory [19, 20], including virtual kitchens [21, 22], virtual cities [23, 24], virtual shopping tasks [25, 26], and virtual offices [27]. It is important to note that research on VR-based neuropsychological assessments has been largely exploratory and often lacks psychometric rigor [28]. While virtual shopping tasks have been found to have some efficacy for assessing memory [25, 26], they have limited interactivity and simply expose participants to scripted events. There is a need for immersive and interactive virtual environments for assessment of everyday memory.

The Virtual Environment Grocery Store (VEGS) [29] was developed as a virtual multiple errands platform, in which participants perform various shopping tasks. Recently, the VR platform was extended to include assessments of episodic and prospective memory [30]. A number of neurocognitive measures are found in the VEGS that assess multitasking and memory (including prospective memory). Participants are immersed in the environment and instructed to 1) navigate to the back of the store and drop off a prescription to a virtual pharmacist, who gives the participant a number; 2) listen for the number while shopping and when heard return to the pharmacist for prescription pickup (event-based prospective memory); and 3) go to the coupon machine after two minutes (time-based prospective memory). A clock was superimposed onto the scene—next to the shopping cart. As a result, participants always had a way to know how much time had elapsed. Additional tasks include: navigating through VEGS by following specified routes through the aisles; finding and selecting items from the shopping list (List Items learned before entering the virtual environment) needed to prepare simple meals (e.g., peanut butter and jelly sandwich); ignoring items that were not on the shopping list learned before entering the virtual environment (Intrusions); and selecting other items so that no more than a budgeted amount is spent (Budget Score). At the completion of the VEGS, the participant performs delayed free (VEGS LDFR) and cued recall (VEGS LDCR) of the VEGS shopping items.

The present study aimed to compare the performance of older adults and young adults on memory reassures in the virtual shopping environment. Specifically, this study aimed to:

(1) Assess the construct validity of the VEGS memory scores via comparison to traditional memory measures (convergent validity) and executive function tasks (divergent validity). It was expected that the VEGS memory scores would be at least moderately correlated with traditional memory measures.
(2) Compare performance on the VEGS episodic memory and other neuropsychological measures to assess their capacity for discriminating older adults from young adults. It was expected that the older adults would perform more poorly than the young adults on the VEGS and traditional neuropsychological measures, but that group differences would be greater on the VEGS task than on the traditional neuropsychological measures.
(3) Evaluate group differences on the VEGS prospective memory and multitasking tasks. It was expected that the older adults would perform more poorly than the young adults on

the VEGS prospective memory and shopping measures.

MATERIALS AND METHODS

The study was approved by the university's Human Research Ethics committees.

Participants

Two groups (total N = 103) took part in this study. Forty-eight community-dwelling older adults (77% female; mean age = 75.21; SD 8.31) participated in the present study. Participants in the older adult group were healthy community dwelling volunteers without history of dementia or mild cognitive impairment. Estimated full scale IQ scores from the Wechsler Test of Adult Reading (WTAR) were in the average range (Mean = 106.34; SD = 9.25). On a measure of computer experience and usage activities, they reported comfort with computers and rated their competency as somewhat experienced. The young adult group included 55 healthy participants (73% female; mean age = 20.96; SD 2.85). Participants in the younger aged group were recruited from the undergraduate population at a large university in the southwestern United States. Participants in the younger aged group had estimated full scale IQ scores in the average range (Mean = 106.38; SD = 8.43) on the WTAR. On a measure of computer experience and usage activities, they reported comfort with computers and rated their competency as experienced. Participants within and across groups were comparable and no significant differences for age, sex, estimated full scale IQ, or computer comfort were noted.

Inclusion and exclusion criteria

Participant selection followed strict exclusion criteria to curtail potential confounding effects of comorbid factors that may adversely impact cognition (e.g., mental retardation, psychotic disorders, diagnosed learning disabilities, attention deficit/hyperactivity disorder, or bipolar disorder). Furthermore, participants were excluded if they had experienced substance-related disorders within 2 years of evaluation). Moreover, participants were excluded if they have known neurologic conditions (e.g., seizure disorders, closed head injuries with loss of consciousness greater than 15 min, and neoplastic diseases). No participant had a reported history of dementia or mild cognitive impairment.

Procedure and instruments

Experimental sessions took place over a 90-min period. Following informed consent, basic demographic information was gathered and participants responded to questions designed to measure computer experience and usage activities: how frequently participants use a computer (e.g., "How many hours per week do you spend on the computer?"); their perceived level of computer skill on a Likert scale (1—not at all to 5—very skilled); e.g., "How many hours per week do you spend playing video games?"; and what type of games they play (e.g., role-playing, strategy, sports, etc.). Participants completed the paper-and-pencil neuropsychological tests that were administered under standard conditions. Participants were also administered the VEGS. Measures were counterbalanced to control for order effects. Half the participants received the VEGS first and the other half received the paper-and-pencil neuropsychological tests first. No order effects were found.

The VEGS was run on a standard laptop and participants took part in a familiarization phase to make sure that they could navigate and make selections in the virtual environment before its administration. The amount of time needed for familiarization ranged from three and ten min—largely determined by the participant's comfort and prior experience with VR platforms. After the familiarization phase, the examiner read a shopping list to the participant. Following the reading of the 16 items on the shopping list, the examiner requested that the participant repeat the items on the shopping list (immediately) in any order. The interstimulus interval was 2 s. Immediate recall performance was recorded verbatim by a microphone and is logged for each of the immediate recall trials (Trials 1–3). Next the participant was informed that they were going to need to drop off a prescription once the VEGS protocol starts. They were also told that they needed to remember to go to the coupon machine after two minutes of shopping (VEGS Time-Based Prospective Memory). Before being reimmersed into the virtual environment, participants were provided instructions and the examiner answered any questions before starting the task. To make sure that the participants understood the instructions, they were asked to recall the prospective memory instructions; all participants could do so.

Virtual environment grocery store

A number of cognitive tasks are found in the VEGS that are involved in multitasking and memory (episodic and prospective memory). The participant was immersed in the VE and instructed on how to move about and interact with the environment (5 min). At this point the VEGS protocol began and the participant had to drop off a prescription with the virtual pharmacist. The virtual pharmacist gave the participant a number and instructed the participant to listen for the number while shopping. The participant shopped for items on the shopping list (VEGS List Items score = number of items that they had in their cart from the previously learned list of shopping items) and kept track of the cost (Budget = total amount of money spent on items in the cart). While shopping, they had to ignore items not on the shopping list (Intrusions Score = number of items in their cart that were not on the original shopping list). They also had to ignore announcements of other prescription numbers (at each 1-min interval of the protocol) from the virtual pharmacist. At 2 min, they were to go to a coupon machine to check for potential savings (Time-based score = 2 if they needed a prompt; and 1 if they self-reminded). After 10 min, the virtual pharmacist announced the participant's prescription number. At that time, the participant needed to return to the virtual pharmacist and clicked on her to end the simulation (VEGS Event-based Prospective Memory score = amount of time it took for them to return to the pharmacist after their number was called). At the completion of the VEGS, the participant was asked to perform free (VEGS Long Delay Free Recall score = total number of items that they recalled freely) and cued delayed recall (Long Delay Cued Recall = total number of items that they recalled in the cued recall condition). Participants were not warned that delayed recall would be tested later.

Testing environment

Participants sat in a rotating chair and interacted with the virtual environment using a joystick and a haptic glove. While participants navigated the VR shopping environment, their head movements were tracked by a sensor (3-Degrees of Freedom Position Sensor). When they turned their head, participants' view within the environment changed in a corresponding manner, allowing them to fully visualize the store. The haptic glove that participants wore on their dominant hand, allowing the head and hand to each be tracked in real time. The hand-tracking was used to enable participants to interact with objects in the virtual environment by reaching out and touching them. Participants controlled their viewpoint in the virtual environment using a Logitech Attack 3 joystick that was firmly mounted to a short wooden bar that extended forward beneath the armrest of an Aeron chair. We prepared left- and right-handed versions of the chairs, as well as of the gloves and hand models, so that each participant could both drive and select objects with their dominant hand. Participants were able to use the joystick to translate their viewpoint forward and backwards in the virtual environment, but they had to swivel in the chair to turn.

Traditional neuropsychological battery

Convergent validity was measured using the California Verbal Learning Test—Second Edition (CVLT–II) [31] as it is considered to have an important memory component and has been used clinically to estimate memory abilities. Divergent validity was assessed using the Color-Word Interference Test from the Delis–Kaplan Executive Function System (DKEFS Inhibition/Switching) [32]. To make sure that the groups (older adult and younger adult) had comparable estimated full scale intelligence scores, the WTAR [33] was used. The WTAR offers estimation of pre-morbid intellectual and memory abilities understood to be unaltered by neurocognitive decline accompanying aging.

Data analytics

All data were analyzed using SAS version 9.4 (Statistical Analysis System). Descriptive statistics were calculated for participant demographics and for results of the VEGS and the criterion neuropsychological tests. An inspection of the pattern of missing data revealed that 2 (1 older adult and 1 younger adult) cases could be recovered by imputing only 1 missing value, and data from another older aged participant could be used after imputing 2 values. Missing data were imputed by either mean substitution or last case carried forward. For all analyses, the results with and without imputed data were comparable. Three types of analyses were performed. The first involved computing basic correlations between VEGS measures and traditional neuropsychological measures. To make the results for continuous tests comparable, the test scores (separately for each group) were converted to standard scores with a mean of zero and standard deviation of one (z-score transformation). The second series of analyses involved computing basic correlations between VEGS measures and traditional neuropsychological tests assessing domains

Table 1
Construct (convergent and divergent) validity results for the older adult sample

Variable	VEGS LDFR	VEGS LDCR	VEGS Time-based	List Items	Intrusions	Budget	VEGS Event-Based
Convergent Validity							
CVLT LDFR	0.56*	0.46*	0.19	0.22	0.03	0.19	–0.01
CVLT LDCR	0.52*	0.42*	0.21	0.14	0.05	0.14	0.03
Divergent Validity							
DKEFS							
Color Word Score	0.26	0.23	–0.09	–0.01	0.00	–0.06	0.25
Inhibition	0.28	0.18	–0.09	0.07	0.05	0.12	–0.02
Inhibition/switching	0.08	0.07	–0.01	0.08	0.07	0.12	–0.06
Interference Total	0.20	0.14	–0.05	0.09	0.07	0.13	–0.05

*significant at the <0.05 level; CVLT-II, California Verbal Learning Test-Second Edition; DKEFS, Delis Kaplan Executive Functioning System; VEGS LDFR, Long Delay Free Recall of items from the VEGS shopping list; LDCR, Long Delay Cued Recall; VEGS, Virtual Environment Grocery Store; VEGS Event-based, VEGS Event-based Prospective Memory score; VEGS Event-based, VEGS Time-based Prospective Memory score.

traditionally understood to be other than memory (e.g., Color-Word Interference). Again, to make the results for continuous tests comparable, the test scores were converted to standard scores with a mean of zero and standard deviation of one (z-score transformation). The third series of analyses assessed differences between groups for VEGS and traditional neuropsychological test performance. Analyses of variance were used to assess group differences. A specific focus of these analyses was to assess discrimination of the VEGS memory and multitasking between groups.

RESULTS

Given the similarity of within-group participants (in terms of performances for both the VEGS and the standard neuropsychological measures) on age, sex, education, and ethnicity no correction for these variables was employed.

To provide preliminary data to support the validity of the VEGS as a measure of memory, recall indices from the VEGS and traditional neuropsychological tests (e.g., CVLT-II) were correlated. Indices were developed from linear composites derived from z-score transformations. Specifically, Pearson correlation analyses were used to compare recall from the VEGS with linear composites derived from traditional neuropsychological measures.

Convergent and discriminant validity tests

While VEGS memory scores were significantly correlated (as expected) with composites derived from established measures of learning and memory, they did not correlate (again, as expected) with traditional neuropsychological domain composites. Construct (convergent and divergent) validity results for the older adult sample can be found in Table 1. The results from the older aged group indicated that the VEGS correlated significantly with the traditional neuropsychological memory scores: CVLT-II long delay free recall $r = 0.56$, $p < 0.001$, with approximately 31% of the variance shared between the two indices; and CVLT-II long delay cued recall $r = 0.42$, $p < 0.001$, with approximately 18% of the variance shared between the two indices. There were no significant correlations between VEGS measures and the traditional neuropsychology measures of executive functioning. Hence, the VEGS tests did not correlate with theoretically unrelated abilities.

The results from the young adult group indicated that the VEGS correlated significantly with the traditional neuropsychological memory scores: CVLT-II long delay free recall $r = 0.30$, $p < 0.001$, with approximately 9% of the variance shared between the two indices; and CVLT-II long delay cued recall $r = 0.36$, $p < 0.001$, with approximately 13% of the variance shared between the two indices. There were no significant correlations between VEGS measures and the traditional neuropsychology measures of executive functioning. Hence, the VEGS tests did not correlate with theoretically unrelated abilities. Construct (convergent and divergent) validity results for the young adult sample can be found in Table 2.

For correlations between the VEGS and traditional psychometric measures, only those correlations that met the criterion of $p < 0.05$ were considered meaningful. Given the relatively small sample size, alpha was kept at this level.

Table 2
Construct (convergent and divergent) validity results for the young adult sample

Variable	VEGS LDFR	VEGS LDCR	VEGS Time-based	List Items	Intrusions	Budget	VEGS Event-Based
Convergent Validity							
CVLT LDFR	0.30*	0.26	–0.07	0.13	0.16	0.06	0.02
CVLT LDCR	0.32*	0.36*	–0.09	0.07	0.09	0.03	–0.02
Divergent Validity							
DKEFS							
Color Word Score	–0.07	–0.22	–0.15	–0.06	–0.03	0.09	–0.04
Inhibition	–0.02	0.00	–0.07	0.04	0.06	–0.05	–0.03
Inhibition/switching	0.23	0.24	0.11	0.15	0.08	0.13	0.08
Interference Total	0.14	0.16	0.04	0.12	0.08	0.06	0.04

*significant at the <0.05 level; CVLT-II, California Verbal Learning Test-Second Edition; DKEFS, Delis Kaplan Executive Functioning System; VEGS LDFR, Long Delay Free Recall of items from the VEGS shopping list; LDCR, Long Delay Cued Recall; VEGS = Virtual Environment Grocery Store; VEGS Event-based, VEGS Event-based Prospective Memory score; VEGS Event-based, VEGS Time-based Prospective Memory score.

Table 3
Comparison of group performance on neuropsychological measures

	Young Adult Older Adult							
Variable	*M*	*SD*	*CI*	*M*	*SD*	*CI*	*F*	η^2
CVLT								
LDFR	7.4	1.4	7.03–7.77	6.4	2.2	5.71–6.99	8.63*	0.08
LDCR	7.6	1.5	7.17–7.95	6.7	2.0	6.11–7.26	6.62*	0.06
DKEFS								
Inhibition	0.5	0.04	0.46–0.48	0.5	0.08	0.43–0.47	1.63	0.02
Inhibition/switching	0.5	0.05	0.44–0.46	0.4	0.08	0.41–0.35	2.47	0.02
Interference Total	0.5	0.04	0.45–0.47	0.4	0.07	0.42–0.46	2.64	0.03
VEGS								
LDFR	9.67	1.94	9.14–10.19	6.97	3.30	6.00–7.91	26.64*	0.21
LDCR	9.42	1.91	8.90–9.93	7.07	3.32	6.06–9.93	20.83*	0.17
Time-based	1.18	0.38	1.07–1.28	1.79	0.41	1.67–1.91	59.81*	0.37
List Items	7.27	2.71	6.92–8.45	5.41	4.49	4.44–7.14	6.62*	0.06
Intrusions	0.41	0.85	0.19–0.65	0.37	1.19	0.28–0.64	0.05	0.01
Budget	36.36	18.88	31.25–41.46	26.80	20.38	20.89–32.72	6.09*	0.06
Event-based	758.4	60.48	742–774	867.9	138.6	827–908	26.22*	0.20

*significant at the <0.05 level; CVLT-II, California Verbal Learning Test-Second Edition; DKEFS, Delis Kaplan Executive Functioning System; VEGS LDFR, Long Delay Free Recall of items from the VEGS shopping list; LDCR, Long Delay Cued Recall; VEGS, Virtual Environment Grocery Store; VEGS Event-based, VEGS Event-based Prospective Memory score; VEGS Event-based, VEGS Time-based Prospective Memory score; AUC, area under the curve; CI, confidence interval; η^2, eta squared—interpreted as 0.02 = small; 0.13 = medium; 0.26 = large.

Comparison of older adult and young adult group performance

Significant differences were found between groups for both traditional neuropsychological assessment of memory (CVLT-II) and the VEGS memory measures. Comparison of group performance on neuropsychological measures can be found in Table 3. The performance of older adults and young adults did not differ significantly on the DKEFS Color Word Interference tests. VR-based multitasking and prospective memory performances differed between the older aged and college aged groups, with the older adult group performing worse on tasks. In addition to taking longer to navigate to the pharmacist and shopping, older aged participants were slow to get to the VEGS coupon machine for the event-related prospective memory task and often needed to be reminded. The older aged cohort also tended to get fewer shopping items during the task.

DISCUSSION

This study provides preliminary validation of the VEGS's memory module. Convergent and discriminant validity were evaluated using paper-and-pencil neuropsychological tests. The VEGS was significantly related to a memory measure, in accordance with prediction. Furthermore, following expectation, VEGS memory scores did not correlate with nonmemory measures drawn from the traditional neuropsychological test battery. Additionally, the differences between age cohorts found in this study were consistent with those in previous studies demonstrating that older adults perform worse than younger adults on measures of verbal memory [34, 35] and prospective memory [36–38] but not on interference tests [39–42]. The result that the VEGS performed similarly to traditional neuropsychological tests among age cohorts adds additional support for its construct validity. Furthermore, the disparity in performance between younger and older adults was greater in VEGS measures (compared to traditional neuropsychological test measures). Together, these findings suggest that the VEGS assesses a construct that is similar to those measured by the other memory tests in this study. Accordingly, it can be concluded that the VEGS had appropriate levels of convergent and divergent validity in that the degree to which convergent validity coefficients (assessing memory domain) derived from the VEGS memory score and the traditional neuropsychological measures of memory were larger than correlations of different measures assessing domains other than memory. Another noteworthy finding was the discriminant capabilities of the VEGS for episodic and prospective memory measures.

Findings revealed that the VEGS prospective memory measures did not correlate with traditional paper-and-pencil measures of episodic memory or executive functions. This deviates from theoretical formulations that emphasize prospective memory as related to either episodic memory or as a form of executive function [43]. The lack of correlation between the VEGS prospective memory tasks and traditional measures of episodic memory is consistent with the generally accepted belief that prospective memory is more vulnerable than retrospective memory to age related cognitive decline [44]. Our findings support this perspective in that we found significant difference between younger and older adult performances on VEGS prospective memory tasks. Furthermore, in this study, there were no differences between groups on their ability to recall the prospective memory instructions. Moreover, when asked to recall the prospective memory instructions, all participants could do so. This argues against the prospect of disadvantaged retrograde recall. Alternatively, prospective memory can be thought of as related to attention and executive processing. Again, while there were differences between groups, there was not a significant correlation between VEGS prospective memory tasks and traditional paper-and-pencil measures of executive functioning. Future studies may be conducted to better investigate the specific contributions of attention and executive functioning to performance on prospective memory tasks in the virtual environment. The Rivermead Behavioral Memory Test (RBMT) [45] was developed to assess memory for everyday activities. Future studies of the VEGS would be enhanced via straightforward comparisons to the RBMT.

It is important to note that the VEGS prospective memory tasks may not correlate with traditional paper-and-pencil measures of episodic memory and or executive functioning due to the two general pathways that support prospective memory: 1) a top–down attentional control process that maintains activation of the intention to monitor the environment for target cues for intention execution; and 2) a bottom–up retrieval process that is automatically triggered by a prospective memory target cue [46–47]. It may be the case that VR-based assessments of prospective memory assess both top-down and bottom-up processing, while paper-and-pencil measures emphasize top-down processes in their assessment. Future studies should investigate both automatic and controlled processes using VR simulations of real-world activities.

While the VEGS memory measure correlates significantly with scores from memory measures, findings should be understood in the context of some limitations. These findings are based on a fairly small sample size. As a necessary next step, the reliability and validity of the test needs to be established using a larger sample of participants to ensure that the current findings are not an anomaly due to sample size. Additionally, the diagnostic utility of this VEGS assessment tool must be determined. The ability of the VEGS to accurately classify participants into memory impaired and nonimpaired groups based on carefully established critical values must be evaluated. This will involve the generation of specific cut-off points for classifying a positive (memory impaired likely) or negative (memory impaired unlikely) finding. Prediction of memory impairment with the VEGS must be evaluated by the performance indices of sensitivity, specificity, predictive value of a positive test, and predictive value of a negative test. Even though reliability is considered to be a unique asset of testing in computer-generated VEs, issues of test–retest reliability must be addressed. A related issue is the need for more refined analyses comparing learning slopes from the VEGS shopping list to those found in CVLT list learning and recall.

Qualitative commentary about the participants' experience with and reactions to the VEGS revealed an overall positive experience. Participants tended to report that they preferred the VR tasks over the paper-and-pencil measures and they felt that they better represented the sorts of activities that they took part in when they went shopping. Furthermore, participants in both groups tended to report that instructions for the task were clear and easy to understand. While these qualitative results are encouraging, future studies should develop questionnaires to assess the participants' subjective experience of the environment. In addition to these subjective metrics, it would be beneficial to log their autonomic response.

In sum, a primary goal of this study was to conduct an initial pilot study to validate the VEGS through the use of a standard neuropsychological battery for comparison of young adults to older adult participants. Findings support this validation. We recognize, however, that the current findings are only a first step in the development of this tool. Many more steps are necessary to continue the process of test development and to fully establish the VEGS as a measure that contributes to existing assessment procedures for the diagnosis of memory decline. Although the VEGS as a measure must be fully validated, current findings provide preliminary data regarding the validity of the VE as a memory measure. The VEGS was correlated with widely used memory assessment tools. Nevertheless, the fairly small sample size requires that the reliability and validity of the VEGS be established using a larger sample of well-matched participants. This will ensure that current findings are not a sample size–related anomaly. Finally, the ability of the VEGS to accurately classify participants not involved in the initial validation study must be examined for cross-validation purposes.

DISCLOSURE STATEMENT

Authors' disclosures available online (http://j-alz.com/manuscript-disclosures/17-0295r2).

REFERENCES

[1] Dickerson BC, Eichenbaum H (2010) The episodic memory system: Neurocircuitry and disorders. *Neuropsychopharmacology* **35**, 86-104.

[2] Okuda J, Fujii T, Ohtake H, Tsukiura T, Tanji K, Suzuki K, Kawashima R, Fukuda H, Itoh M, Yamadori A (2003) Thinking of the future and past: The roles of the frontal pole and the medial temporal lobes. *Neuroimage* **19**, 1369-1380.

[3] Raz N, Lindenberger U, Rodrigue KM, Kennedy KM, Head D, Williamson A, Dahle C, Gerstorf D, Acker JD (2005) Regional brain changes in aging healthy adults: General trends, individual differences and modifiers. *Cereb Cortex* **15**, 1676-1689.

[4] Kliegel M, Martin M, McDaniel MA, Einstein GO (2002) Complex prospective memory and executive control of working memory: A process model. *Psychologische Beiträge* **44**, 303-318.

[5] Woods SP, Weinborn M, Velnoweth A, Rooney A, Bucks RS (2012) Memory for intentions is uniquely associated with instrumental activities of daily living in healthy older adults. *J Int Neuropsychol Soc* **18**, 134-138.

[6] Alderman N, Burgess PW, Knight C, Henman C (2003) Ecological validity of a simplified version of the multiple errands shopping test. *J Int Neuropsychol Soc* **9**, 31-44.

[7] Manchester D, Priestley N, Jackson H (2004) The assessment of executive functions: Coming out of the office. *Brain Inj* **18**, 1067-1081.

[8] Chaytor N, Schmitter-Edgecombe M (2003) The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychol Rev* **13**, 181-197.

[9] Burgess PW, Alderman N, Forbes C, Costello A, Coates LM, Dawson DR, Anderson ND, Gilbert SJ, Dumontheil I, Channon S (2006) The case for the development and use of "ecologically valid" measures of executive function in experimental and clinical neuropsychology. *J Int Neuropsychol Soc* **12**, 194-209.

[10] Bier N, Belchior Pda C, Paquette G, Beauchemin É, Lacasse-Champagne A, Messier C, Pellerin ML, Petit M, Mioshi E, Bottari C (2016) The instrumental activity of daily living profile in aging: A feasibility study. *J Alzheimers Dis* **52**, 1361-1371.

[11] Logie RH, Trawley S, Law A (2011) Multitasking: Multiple, domain-specific cognitive functions in a virtual environment. *Mem Cognit* **39**, 1561-1574.

[12] Rand D, Basha-Abu Rukan S, Weiss PL, Katz N (2009) Validation of the Virtual MET as an assessment tool for executive functions. *Neuropsychol Rehabil* **19**, 583-602.

[13] Bohil CJ, Alicea B, Biocca FA (2011) Virtual reality in neuroscience research and therapy. *Nat Rev Neurosci* **12**, 752-762.

[14] Parsons TD (2015) Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Front Hum Neurosci* **9**, 660.

[15] Parsons TD, Carlew AR, Magtoto J, Stonecipher K (2017) The potential of function-led virtual environments for ecologically valid measures of executive function in experimental and clinical neuropsychology. *Neuropsychol Rehabil* **27**, 777-807.

[16] Benoit M, Guerchouche R, Petit PD, Chapoulie E, Manera V, Chaurasia G, Drettakis G, Robert P (2015) Is it possible to use highly realistic virtual reality in the elderly? A feasibility study with image-based rendering. *Neuropsychiatr Dis Treat* **11**, 557-563.

[17] Mueller C, Luehrs M, Baecke S, Adolf D, Luetzkendorf R, Luchtmann M, Bernarding J (2012) Building virtual reality fMRI paradigms: A framework for presenting immersive virtual environments. *J Neurosci Methods* **209**, 290-298.

[18] Plancher G, Gyselinck V, Nicolas S, Piolino P (2010) Age effect on components of episodic memory and feature binding: A virtual reality study. *Neuropsychology* **24**, 379-390.

[19] Plancher G, Tirard A, Gyselinck V, Nicolas S, Piolino P (2012) Using virtual reality to characterize episodic memory profiles in amnestic mild cognitive impairment and Alzheimer's disease: Influence of active/passive encoding. *Neuropsychologia* **50**, 592-602.

[20] Sauzéon H, Arvind Pala P, Larrue F, Wallet G, Déjos M, Zheng X, Guitton P, N'Kaoua B (2011) The use of virtual reality for episodic memory assessment: Effects of active navigation. *Exp Psychol* **59**, 99-108.

[21] Allain P, Foloppe DA, Besnard J, Yamaguchi T, Etcharry-Bouyx F, Le Gall D, Nolin P, Richard P (2014) Detecting everyday action deficits in Alzheimer's disease using a non-immersive virtual reality kitchen. *J Int Neuropsychol Soc* **20**, 468-477.

[22] Besnard J, Richard P, Banville F, Nolin P, Aubin G, Le Gall D, Richard I, Allain P (2016) Virtual reality and neuropsychological assessment: The reliability of a virtual kitchen to assess daily-life activities in victims of traumatic brain injury. *Appl Neuropsychol Adult* **23**, 223-235.

[23] Jovanovski D, Zakzanis K, Campbell Z, Erb S, Nussbaum D (2012) Development of a novel, ecologically oriented virtual reality measure of executive function: The Multitasking in the City Test. *Appl Neuropsychol Adult* **19**, 171-182.

[24] Parsons TD, Rizzo AA (2008) Initial validation of a virtual environment for assessment of memory functioning: Virtual reality cognitive performance assessment test. *Cyberpsychol Behav* **11**, 17-25.

[25] Canty AL, Fleming J, Patterson F, Green HJ, Man D, Shum DH (2014) Evaluation of a virtual reality prospective memory task for use with individuals with severe traumatic brain injury. *Neuropsychol Rehabil* **24**, 238-265.

[26] Grewe P, Lahr D, Kohsik A, Dyck E, Markowitsch HJ, Bien CG, Botsch M, Piefke M (2014) Real-life memory and spatial navigation in patients with focal epilepsy: Ecological validity of a virtual reality supermarket task. *Epilepsy Behav* **31**, 57-66.

[27] Matheis RJ, Schultheis MT, Tiersky LA, DeLuca J, Millis SR, Rizzo A (2007) Is learning and memory different in a virtual environment? *Clin Neuropsychol* **21**, 146-161.

[28] Renison B, Ponsford J, Testa R, Richardson B, Brownfield K (2012) The ecological and construct validity of a newly developed measure of executive function: The Virtual Library Task. *J Int Neuropsychol Soc* **18**, 440-450.

[29] Parsons TD, Rizzo AA, Brennan J, Zelinski EM (2008) Assessment of executive functioning using virtual reality: Virtual Environment Grocery Store. *Gerontechnology* **7**, 187-191.

[30] Parsons TD, McPherson S, Interrante V (2013) Enhancing neurocognitive assessment using immersive virtual reality. *Proceedings of the 17th IEEE Virtual Reality Conference: Workshop on Virtual and Augmented Assistive Technology (VAAT)*. 1-7.

[31] Delis DC, Kramer JH, Kaplan E, Ober BA (2000) *CVLT-II: California Verbal Learning Test: Adult Version*. The Psychological Corporation, San Antonio, TX.

[32] Delis DC, Kaplan E, Kramer JH (2001) *Delis–Kaplan Executive Function System (D-KEFS): Examiner's Manual*. The Psychological Corporation, San Antonio, TX.

[33] Holdnack HA (2001) Wechsler Test of Adult Reading: WTAR. The Psychological Corporation, San Antonio, TX.

[34] Cargin JW, Maruff P, Collie A, Shafiq-Antonacci R, Masters C (2007) Decline in verbal memory in non-demented older adults. *J Clin Exp Neuropsychol* **29**, 706-718.

[35] Davis HP, Small SA, Stern Y, Mayeux R, Feldstein SN, Keller FR (2003) Acquisition, recall, and forgetting of verbal information in long-term memory by young, middle-aged, and elderly individuals. *Cortex* **39**, 1063-1091.

[36] Henry JD, Macleod MS, Phillips LH, Crawford JR (2004) A meta-analytic review of prospective memory and aging. *Psychol Aging* **19**, 27-39.

[37] Rose NS, Rendell PG, McDaniel MA, Aberle I, Kliegel M (2010) Age and individual differences in prospective memory during a "Virtual Week": The roles of working memory, vigilance, task regularity, and cue focality. *Psychol Aging* **25**, 595-605.

[38] Schnitzpahn KM, Stahl C, Zeintl M, Kaller CP, Kliegel M (2013) The role of shifting, updating, and inhibition in prospective memory performance in young and older adults. *Dev Psychol* **49**, 1544-1553.

[39] Adólfsdóttir S, Wollschlaeger D, Wehling E, Lundervold AJ (2017) Inhibition and switching in healthy aging: A longitudinal study. *J Int Neuropsychol Soc* **23**, 90-97.

[40] Graf P, Uttl B, Tuokko H (1995) Color- and picture-word Stroop tests: Perfomance changes in old age. *J Clin Exp Neuropsychol* **17**, 390-415.

[41] Cohen AL, Dixon RA, Lindsay DS (2005) The intention interference effect and aging: Similar magnitude of effects for young and old adults. *Appl Cognit Psychol* **19**, 1177-1197.

[42] Schilling V, Chetwynd A, Rabbit P (2002) Individual inconsistency across measures of inhibition: An investigation of the construct validity of inhibition in older adults. *Neuropsychologia* **40**, 605-619.

[43] Cole SN, Morrison CM, Conway MA (2013) Episodic future thinking: Linking neuropsychological performance with episodic detail in young and old adults. *Q J Exp Psychol (Hove)* **66**, 1687-1706.

[44] Huppert FA, Beardsall L (1993) Prospective memory impairment as an early indicator of dementia. *J Clin Exp Neuropsychol* **15**, 805-821.

[45] Wilson BA, Cockburn J, Baddeley AD (1989) The Rivermead Behavioral Memory Battery. *Titchfield, Fareham, Hants: Thames Valley Test Company*.

[46] McDaniel MA, Einstein GO (2000). Strategic and automatic processes in prospective memory retrieval: A multiprocess framework. *Appl Cognit Psychol* **14**, S127-S144.

[47] McDaniel MA, Umanath, S, Einstein GO, Waldum ER (2015). Dual pathways to prospective remembering. *Frontiers Hum Neurosci* **9**, 392.