A comparison of immersive virtual reality with traditional neuropsychological measures in the assessment of executive functions

Davison SMC, Deeprose C, Terbeck S. A comparison of immersive virtual reality with traditional neuropsychological measures in the assessment of executive functions.

**Sophie Melissa Clare Davison,
Catherine Deeprose,
Sylvia Terbeck**

Department of Psychology, Faculty of Health and Human Sciences, University of Plymouth, Plymouth, UK

Objective: This study investigated immersive virtual reality (IVR), as a novel technique to test executive function of healthy younger and older adults. We predicted IVR tasks to have greater predictive power than traditional measures when assessing age-related cognitive functioning due to the real-world validity of the tasks.
Methods: Participants ($n = 40$) completed the Stroop colour–word test and the trail-making test (TMT) as traditional and commonly used assessments of executive functioning. Participants then completed three IVR tasks; a seating arrangement task, an item location task (both set in a virtual chemistry lab), and a virtual parking simulator.
Results: Younger adults completed significantly more parking simulator levels ($p < 0.001$), placed significantly more objects ($p < 0.001$), and located significantly more items than older adults ($p < 0.01$), demonstrating higher levels of performance. Significant correlations were found between performance on traditional neuropsychological measures and IVR measures. For example, Stroop CW performance significantly correlated with the number of parking simulator levels completed ($\tau = 0.43$, $p < 0.01$). This suggests that IVR measures assess the same underlying cognitive constructs as traditional tasks. In addition, IVR measures contributed a significant percentage of the explained variance in age.
Conclusion: IVR measures (i.e. number of parking simulator levels completed and number of objects placed in the seating arrangement task) were found to be stronger contributors than existing traditional neuropsychological tasks in predicting age-related cognitive decline. Future research should investigate the implementation of these real-world-based tasks in clinical groups given this promising initial work.

Significant outcomes

- As predicted, significant differences were found between healthy younger adults and healthy older adults in the Stroop test, trail-making test (TMT) A, and also on the immersive virtual reality (IVR) measures indicating sensitivity to age-related cognitive decline.
- Significant correlations were found between traditional neuropsychological measures and IVR measures indicating convergent validity.
- Multiple regression analysis revealed that only the IVR measures contributed a significant percentage of variation in predicting age.

Introduction

Executive function (EF) is an umbrella term for cognitive processes such as inhibition, working memory, cognitive flexibility, reasoning, problem solving and planning abilities [1,2]. These systems operate together to facilitate goal-directed problem-solving usually when a novel or complex situation is encountered [3,4]. It is well established that EF declines are associated with both normal and pathological ageing [5,6]. Traditional neuropsychological measures that are sensitive to age differences in EF include the Stroop colour–word test [7] and TMT [8]. Older adults experience increased interference in the Stroop task when compared with younger adults, indicating a decrease in inhibitory control [9,10]. Therefore, as age increases, performance (response times) becomes slower on both TMT A and B [11,12].

In recent years, research has begun to focus on developments in IVR technology as an alternative (or potentially better) assessment of EF compared with traditional neuropsychological tests. Importantly, IVR has the potential to offer a more realistic, ecologically valid alternative to pencil and paper tasks [13]. Daily living skills are often assessed indirectly (i.e. via self-report or via an informant), however, IVR offers a more direct method of assessment [14]. Although commonly used for clinical and research purposes, pencil and paper tasks are thought to lack the stresses and distractions of real life settings [15]. IVR also offers the advantage of shielding participants from the dangers of real-life environments, making their use ideal for testing vulnerable adults, such as adults with neurological impairments [15]. Furthermore, IVR can be versatile, offering standardised applications as an assessment and rehabilitation tool [16].

Before reviewing the literature into EF and IVR, it is important to reflect on what is considered 'virtual reality' (VR) technology. Non-immersive technologies, for example, tasks performed on desktop computers [17] or on computer tablets, are often also labelled as VR [18]. However, in our study we refer to 'immersive virtual reality' because of our use of IVR technology, which includes the use of head-mounted displays that immerse the participant into a 360° virtual world [19] by utilising head tracking, allowing for the change of the virtual scene according to the head movements of the participants.

One VR tool which has been developed to test EF is the Jansari assessment of executive functions (JEF) [20] which is delivered via a laptop. The JEF VR assessment involves a participant taking on the role of an office assistant. The participant is required to complete different tasks which would typically be completed as part of a working day. The assessment relies on the participant being able to switch between tasks and requires the use of working memory. EF was measured through eight constructs, one of which was planning. The planning construct required participants to arrange tables and chairs in preparation for a meeting. The JEF VR assessment has been shown to be effective in differentiating between the EF of participants with an acquired brain injury and healthy matched controls [21]. However, evidence has shown its lack of ability to assess inhibitory control [22].

Another non-immersive VR task which tested EF is the virtual library task (VLT) [23]. Participants were required to complete tasks based on the running of a library. For example, participants were required to problem solve and find an alternative approach when an air conditioning system fails to work. However, again, the VLT was a non-immersive task that was operated via a desktop computer, which therefore might be less well suited to predict ageing then more realistic IVR measures tested in the current study.

A substantial amount of literature has focussed on VR and EF in clinical groups with a marked deficit in EF capability. This includes neurological impairment such as Alzheimer's disease [24] or mild cognitive impairment (MCI) [18]. Individuals with deficits in EF typically show difficulty in learning new tasks, attentional switching, planning, and have an increased distractibility [25]. For instance, a number of programmes targeted at testing the EF of people with cognitive impairments are situated in everyday settings such as supermarkets [17–19,25]. One such programme is the virtual multiple errands test (VMET) which is a measurement of EF in daily life [17]. The VMET was originally developed to aid post-stroke patients to engage in complex everyday activities, such as going to a shopping centre or supermarket and is a virtual version of the multiple errands test [26]. In a study by Raspelli et al. [27] the VMET was used to test differences between post-stroke patients, healthy older individuals (aged between 50 and 70 years old) and healthy younger individuals (aged between 20 and 30).

The VMET was, however, also not immersive. The VMET involves three tasks in which the participant has to follow certain rules. These tasks include buying items, finding information and checking a shopping cart. The authors found statistically significant differences between the three groups in the time taken to complete the tasks and in the number of errors made, with patients making the most errors, followed by older adults and then younger adults. Patients also took the longest in executing the tasks. Thus although sensitive to the established age-related deficits in EF, the novel task was not compared with traditional paper-and-pencil measures, therefore yet to be determined whether the virtual reality tasks were more sensitive than established measures, and to what extent the novel tasks were assessing the same cognitive construct.

Another virtual construct aimed at assessing activities of daily living (ADL) is RehabCity (28). Run via a desktop computer, RehabCity aims to provide cognitive rehabilitation training and assessment in a real life context, for those who have suffered a neurological impairment, such as in the case of a stroke. The virtual city consists of everyday environments, such as a supermarket, bank, pharmacy and post office, set within a city comprising of streets, buildings, parks and moving vehicles. RehabCity assesses EF by requiring the user to accomplish simple goals and more complex ones which require problem solving. Points are acquired when a goal is completed, and removed when a mistake is made, or the user asks for help. A higher score on the RehabCity tool was associated with a younger age, whilst an older age was associated with spending more time completing ADLs (28). Despite its strengths in assessing ADLs within everyday environments, RehabCity lacks ecological validity due to its implementation via a desktop computer. RehabCity would have improved ecological validity if implemented via IVR.

Zygouris et al. (18) investigated EF in the setting of a virtual supermarket (VSM). Healthy older participants and participants with a MCI were asked to navigate through the VSM and buy items from a shopping list. They were then required to pay for the items with virtual cash. The variables measured included the number of the items purchased, the number of incorrect items purchased and the time taken to complete the task. This VR study was again however non-immersive and completed via the use of a tablet. VSM was able to differentiate between healthy older adults and adults with a MCI in 87.30% of cases. Kang et al. (19) also tested EF through performance in a VR supermarket task. However, this study utilised a head-mounted display. Performance quality on this shopping task was found

to be significantly different between a control group and patients who had suffered a stroke, with stroke patients performing the task significantly worse than control participants.

In the present study, three new 3D IVR tasks, aimed at assessing EF were tested on a healthy population of younger and older adults. All three IVR tasks required the participants to neglect irrelevant information in the reality simulation and concentrate on achieving their task goal in a specified time. The first IVR task utilised an IVR parking simulator application and investigated if its outcome measures would correlate with traditional measures of EF, and indeed explain age variance effects better. A parking simulator was chosen in particular as this task was judged to be highly ecologically valid, as it is representing a common everyday activity. The first IVR chemistry lab task (the seating arrangement task) was based on the VR planning construct task which formed part of JEF (i.e. the seating arrangement task) (20). This task aimed to test participants planning abilities when creating a specified seating arrangement. The second chemistry lab task (the item location task) was based on previous VR tasks which required participants to locate different items or information (17–19). In this novel interactive immersive 3D chemistry lab environment, participants were asked to locate six items.

Our predictions were as follows:

(1) Performance on traditional neuropsychological tasks and IVR task would significantly correlate.
(2) All tasks would differentiate between younger adults and older adults.
(3) Regression analysis would show that IVR tasks are better predictors of age-related decline then traditional tasks.

Aims of the study

The aim of the current experiment was to test the utility of virtual reality to executive function of healthy younger adults and healthy older adults compared to traditional neuropsychological measures.

Materials and methods

Participants

A total of 40 individuals were recruited (M_{age} = 42.78, age range: 19–77 years, 19 men, 21 women). Participants were recruited via the University of Plymouth participation pool and via recruitment in the local community. The sample consisted of two groups of participants, those aged from 18 to 25 years old (n = 22, M_{age} = 20.55, age range = 19–24 years,

12 men, 10 women), and those aged 65 years and over ($n = 18$, $M_{age} = 69.94$, age range = 66–77 years, 7 men, 11 women). There were two inclusion criteria for the experiment. First, participants were required to be aged between 18 and 25 years old, or 65 years and over. Second, participants were required to be 'cognitively healthy', meaning they could not have been diagnosed with any neurological impairment that could affect their cognitive abilities, for example Dementia, Parkinson's disease.

Materials

Experimental questionnaires. Experimental questionnaires included a demographics questionnaire and a post-experiment questionnaire. The demographics questionnaire asked participants for details of educational attainment, if they had any significant medical conditions and if they themselves or anyone in their family had a neurological impairment. The post-experiment questionnaire asked participants how much they had enjoyed the VR tasks (this was measures on a scale of 1–10, with 1 representing low enjoyment and 10 representing high enjoyment), whether they had a task preference (i.e. traditional measures or the VR measures), as well as whether they felt they had any difficulties in day to day planning, difficulties in completing simple everyday tasks and any difficulties in carrying out two tasks simultaneously.

Traditional neuropsychological measures. To test participant's executive function, two traditional neuropsychological tests were utilised; the Stroop colour–word test (29) and the TMT (30). Stroop C forms were used, as well as, Stroop CW forms. The Stroop C task involved participants reading aloud 112 coloured words. In this task the colour word was congruent with the colour the word was written in. For example, the word red was written in red ink. Participants were timed 120 s to complete this task. The Stroop CW task was exactly the same, however, the colour word was written in a colour which was incongruent with the colour word itself. For example, the word blue would be written in a colour other than blue. Participants were required to read aloud the colour the word was written in. TMT A were used and TMT B. The TMT A involved participants drawing a line between numbers consecutively. They would start at the number one, draw a line to the number two, and so on, until they had joined all 25 numbers. TMT B involved the participant drawing a line between numbers and letters consecutively, but alternating between numbers and letters. For example, drawing a line from 1 to A, from A to 2, and so on. Participants were timed for both TMT A and B.

IVR tasks. The experience of IVR, compared with desktop VR creates the most realistic and closest to life illusion currently available (31) and so IVR was decided to be the most effective technique to use for the VR tasks. The IVR tasks were run using Oculus Rift IVR head-mounted equipment. An Oculus Rift Development Kit 2 (DK2: 2015, Oculus VR, LLC company) headset (32) was used to immerse participants into the 360° environment of the virtual tasks. An Oculus Rift DK2 positional tracker camera (32) was also used. Two IVR applications were used; the parking simulator task and the virtual chemistry lab environment. The parking simulator task consisted of five levels, the first two levels required the participant to drive into a parking space. The third and fourth levels required the participant to reverse into a parking space. The final level required the participant to parallel parking. The participants were required to navigate themselves into the parking space by using the four arrows on a computer keyboard. These arrows allowed the participant to drive forwards, turn left and right, and reverse. Participants were allocated 5 min to complete this task. Figure 1 shows an image of the first parking simulator level.

The second and third IVR tasks were set in the environment of a chemistry lab. Images of the IVR chemistry lab can be seen in Fig. 2. The first chemistry lab task, the seating arrangement task, required the participant to create a seating plan by picking up and placing down a chair and five stools in front of a whiteboard. Participants could pick up items by looking directly at them and then clicking the left mouse button. Once the participant had picked up their chosen item, they could then move the item to where they wanted to place it by using the four arrow keys on the keyboard. Participants would then place the item down by clicking the left mouse button again. Participants were allocated 3 min to complete this task.

The second chemistry lab task, the item location task, required the participant to locate various items from around the chemistry lab. The experimenter read out which items the participant was required to locate two at a time. This required one item to be held in working memory whilst the other item was being located. To locate the item, participants were asked to use the four arrow keys to navigate around the lab. Once the item had been located, participants were asked stand in front of the item and tell the experimenter they had found the item. Following this, the participant moved onto the next item. There were six items to locate in total. Participants were allocated three minutes to complete this task. Participants were asked to complete all three IVR tasks as quickly as possible. These IVR tasks were developed from
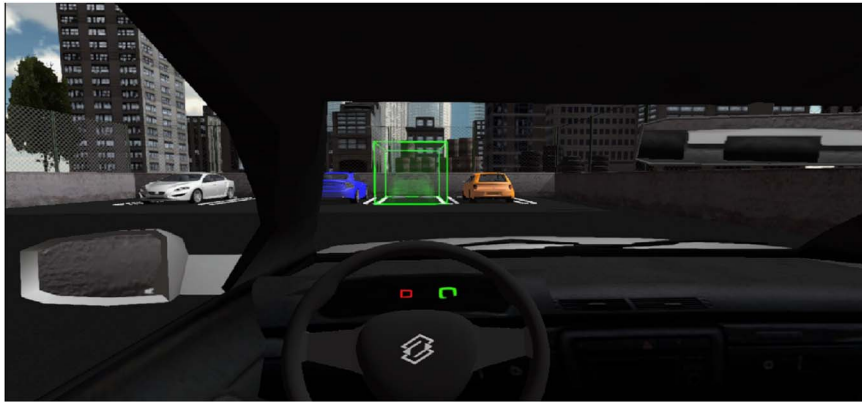
Fig. 1. Immersive Virtual Reality parking simulator.

Fig. 2. Immersive Virtual Reality chemistry laboratory task.

existing shared applications (the applications are available from the authors if needed).

Design and procedure. A mixed between and within subjects experimental design was used.

After informed consent, participants were required to complete the demographics questionnaire. Participants then completed the first neuropsychological test, the TMT, both parts A and B. The time taken to complete the task was recorded. Following this, participants completed the Stroop colour–word test, of which participants completed both Stroop C and CW.

The second part of the experiment required participants to undertake the three IVR tasks. All three tasks required the participant to wear an Oculus Rift DK2 headset (32), which immersed the participant into the task environment. Before undertaking the IVR tasks, participants were shown one of the virtual environments they would be immersed into (i.e. the chemistry lab environment). Whilst wearing the DK2 headset, they were instructed how to move around the environment and

how to pick up objects. The aim of this was to get participants familiar with the 360° immersive virtual environment and the navigation through it. This typically took around 5 min. Key outcome variables for each IVR task are shown in Table 1.

Following training in the IVR environment, participants undertook the three IVR tasks. If participants experienced any motion sickness during completion of the IVR tasks, this was noted by the experimenter. Following the IVR tasks, participants were asked to fill out a post-experiment questionnaire.

Results

Group characteristics

As shown in Table 2, older adults were more likely than younger adults to have corrected vision, to be able to drive, to have a significant medical condition, to have a family member with a neurological impairment and to be using medication which may have affected their cognitive abilities.

83

We found that 40.9% of younger participants expressed difficulty in planning in day-to-day life, this was compared with only 22.2% of older participants. A total of 45.5% of younger participants also expressed difficulty in carrying out two tasks simultaneously, this compared with only 11.1% of older participants. A total of 13.6% of younger participants expressed difficulty in completing simple everyday tasks, compared with 0% of older participants.

Both groups showed similar levels of preference for the pencil and paper tasks with 36.4% of younger participants preferring pencil and paper tasks compared with 38.9% of older participants. The IVR tasks were given a higher preference by younger participants (54.5%), compared with older participants (22.2%). Although, older participants were more likely to have no preference for either of the tasks (22.2%), enjoying them both equally, compared with younger participants (0%). Task preference did not affect how much participants enjoyed the IVR tasks. Younger participants ($Mdn = 8$) and older participants ($Mdn = 8$) did not significantly differ in how much enjoyed completing the virtual reality tasks, $U = 182$, $z = -0.44$, $p = 0.67$.

Table 1. The outcome measures of the three IVR tasks

	Variables measured
Immersive virtual reality task	
Parking simulator	Time taken to complete parking simulator level 1
	Time taken to complete parking simulator level 2
	Time taken to complete parking simulator level 3
	Time taken to complete parking simulator level 4
	Time taken to complete parking simulator level 5
	Number of parking simulator levels completed
	Number of times the virtual car is crashed
Seating arrangement task	Time taken to place the chair
	Time taken to place stool 1
	Time taken to place stool 2
	Time taken to place stool 3
	Time taken to place stool 4
	Time taken to place stool 5
	Number of items placed in total
Item location task	Time taken to locate object 1
	Time taken to locate object 2
	Time taken to locate object 3
	Time taken to locate object 4
	Time taken to locate object 5
	Time taken to locate object 6
	Number of items located in total

Shapiro–Wilk tests of normality and Levene's tests of homogeneity of variance were conducted on the data discussed in sections 2 and 3. If the data did not follow a normal distribution or meet the assumptions of homogeneity of variance, then transformations were conducted. In the event of these transformations being unsuccessful, non-parametric tests were conducted accordingly.

Traditional measures of executive function

Table 3 shows the mean and standard deviations for performance on Stroop task C, CW, TMT A and TMT B. No significant differences were found between the two age groups when completing the Stroop C task, $U = 196.50$, $z = -0.11$, $p = 0.70$. There was a significant difference between the two groups in the number of correct responses given in the Stroop CW task, $U = 57$, $z = -4.1$, $p < 0.001$.

Younger participants were significantly quicker to complete TMT A than older participants, $t(38) = -4.83$, $p < 0.001$. Younger participants were also quicker than older participants when completing TMT B (see Table 3). However, this difference was not statistically significant, $t(38) = -1.92$, $p = 0.62$.

Correlations

A Kendall's τ correlation test was conducted to analyse whether the traditional measures of executive function (i.e. Stroop test and TMT) had significant correlations with the virtual reality measures (i.e. parking simulator and chemistry lab tasks).

Performance on Stroop CW was significantly related to performance on TMT A ($\tau = -0.33$, $p = 0.01$), and TMT B ($\tau = -0.32$, $p = 0.01$). Performance on Stroop CW was also significantly related to performance on the second parking simulator task ($\tau = -0.32$, $p = 0.01$), the number of levels completed on the parking simulator task ($\tau = 0.43$, $p < 0.01$), the time taken to place the blue chair in the seating arrangement task ($\tau = -0.39$, $p = 0.01$), time taken to place the first stool in the seating arrangement task ($\tau = -0.33$, $p = 0.01$), the number of items placed in the seating arrangement task ($\tau = 0.33$, $p = 0.02$), and with the time taken to

Table 2. Characteristics of the young and older adult age groups

	Corrected vision	Driving	Significant medical condition	Family neurological impairment	Using medication	Undergraduate degree attained
Younger adults (%)	40.9**	68.2	9.1	9.1	0	45.5
Older adults (%)	88.9**	83.3	22.2	27.8	11.1	38.9

**$p < 0.01$.

Table 3. Mean number of correct responses in both Stroop tasks and the mean number of seconds taken to complete TMT A and B, by age group

Task completed	Age group					
	18–25			65+		
	M	SD	Range	M	SD	Range
Stroop C	111.82	0.85	108–112	111.89	0.47	110–112
Stroop CW	111.27***	1.32	108–112	92.28***	28.38	−1–112
TMT A (s)	22***	5.19	14–31	30.83***	6.07	21–45
TMT B (s)	52	20.43	28–104	71.61	41.14	29–188

Stroop CW, Stroop colour–word test; TMT, trail-making test.
The range and standard deviations of these scores are also displayed.
***$p < 0.001$.

locate the first item in the item location task ($\tau = -0.37$, $p = 0.01$).

Performance on TMT A was significantly related to performance on the first parking simulator level ($\tau = 0.34$, $p < 0.01$), and performance on the second parking simulator level ($\tau = 0.38$, $p < 0.01$). Performance on TMT A was also significantly related to how many parking simulator levels were completed ($\tau = -0.49$, $p < 0.01$). Significant correlations were also identified between performance on TMT A and the time taken to place the blue chair, in the seating arrangement task, ($\tau = 0.37$, $p = 0.01$), the first stool ($\tau = 0.25$, $p = 0.05$), and the second stool ($\tau = 0.3$, $p = 0.03$). Performance on TMT A also significantly related to the number of items placed in seating arrangement task ($\tau = -0.35$, $p = 0.01$), and the number of items located in the item location task ($\tau = -0.38$, $p < 0.01$). Performance on TMT B was significantly related to the number of levels completed of the parking simulator task ($\tau = -0.28$, $p = 0.03$).

Virtual reality measures

The speed at which participants completed each parking simulator level and how many parking simulator levels they completed was analysed. Younger participants were significantly quicker ($M = 17.50$, $SD = 9.43$) in completing the first parking simulator level than older participants ($M = 39.41$, $SD = 22.29$), $t(37) = -4.63$, $p < 0.001$. Younger participants ($M = 22.18$, $SD = 10.93$) were also significantly quicker in completing the second simulator level than older participants ($M = 84.18$, $SD = 43.39$), $t(37) = -8.57$, $p < 0.001$. It is worth noting that only two older participants were able to complete the third parking simulator level, and only one participant was able to complete the fourth parking simulator level. No older participants were able to complete the fifth level of the parking simulator and so comparison between the two groups was not possible.

The number of parking simulator levels completed by both age groups were investigated. Analysis found that younger participants ($Mdn = 4.5$) completed, on average, more levels of the parking simulator than older participants ($Mdn = 2$). A Mann–Whitney Test found this difference between groups to be significant, $U = 26.5$, $z = -4.97$, $p < 0.001$. No significant difference was found between groups regarding how many times they crashed the car during the parking simulator task, $t(38) = 0.09$, $p = 0.93$.

Analysis was also conducted on the first chemistry lab task, the seating arrangement task. Figure 3 shows that younger participants were quicker in placing a chair in the seating arrangement task than older participants. A Mann–Whitney U test showed this difference to be statistically significant, $U = 4$, $z = -4.38$, $p < 0.001$. The figure also shows that younger participants were quicker in placing the first stool, than older participants. An independent samples t-test showed this difference to be statistically significant, $t(32) = -4.10$, $p < 0.001$. Younger participants ($M = 18.15$, $SD = 14.48$) were also significantly quicker in placing the second stool, than older participants ($M = 33.78$, $SD = 11.88$), $t(27) = -3.34$, $p < 0.01$. With regards to the time taken to place the third stool, younger participants ($M = 14.89$, $SD = 9.26$) were significantly quicker than older participants ($M = 25.33$, $SD = 11.36$), $t(23) = -2.3$, $p = 0.03$. A Mann–Whitney U test indicated that younger participants ($Mdn = 6$) placed significantly more items in the seating arrangement task than older participants ($Mdn = 3$), $U = 23$, $z = -4.58$, $p < 0.001$.
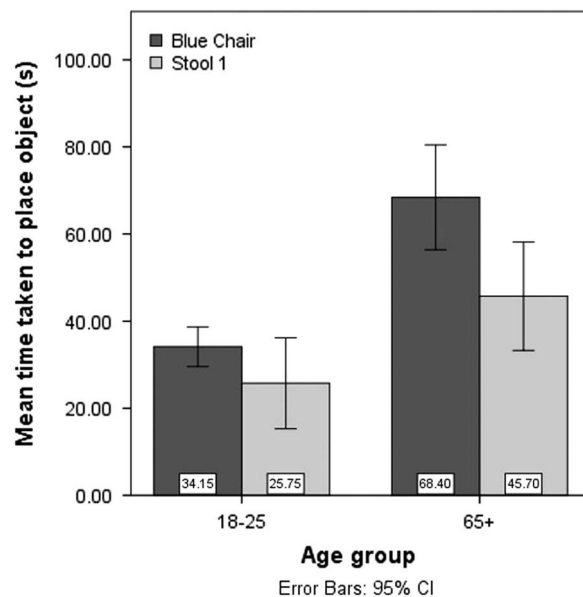

Fig. 3. Mean time to place objects in the chemistry laboratory task by age group.

Analysis was conducted on the speed at which participants located the specified items in the item location task and how many items they located in total. Younger participants ($M = 35.45$, $SD = 18.70$) were significantly quicker in locating the first item in the item location task than older participants ($M = 94.40$, $SD = 44.91$), $t(28) = -4.97$, $p < 0.001$. There was no significant difference found between groups with regards to how quickly item 2 was located, $t(16) = -0.35$, $p = 0.73$. A Mann–Whitney U test found that younger participants ($Mdn = 4$) located significantly more items than older participants ($Mdn = 1$), $U = 67.5$, $z = -2.98$, $p < 0.01$.

Altogether, 11 out of 40 participants experienced a mild degree of motion sickness from undertaking the virtual reality tasks (27.5%). For younger participants, seven out 22 experienced motion sickness (31.8%), and for older participants, four out of 18 experienced motion sickness (22.2%).

Multiple regression analysis

A multiple regression was performed using the backwards method to test whether the following variables were predictive of the dependent variable participant age; Stroop C, Stroop CW, TMT A, TMT B, number of levels completed of the parking simulator, number of objects placed in the seating arrangement task and number of items found in the item location task. A first model containing the aforementioned variables was found to produce a significant regression equation ($F(7,25) = 17.93$, $p < 0.001$, $R^2 = 0.83$). However, a final model containing only the variables of number of levels completed of the parking simulator and number of items placed in the seating arrangement task still produced a significant regression equation ($F(2,30) = 56.7$, $p < 0.001$, $R^2 = 0.79$). Table 4 below shows the results of the regression analysis.

Discussion

As predicted, we found significant correlations between performance on the traditional neuropsychological tasks and IVR tasks. We also found that some of measures in the IVR tasks differentiated between younger adults and older adults. Finally, regression analysis demonstrated that the IVR tasks were better predictors of age and age-related cognitive decline than traditional tasks.

Significant correlations of particular prominence in this study were between several traditional neuropsychological measures and IVR measures. Performance on TMT A, a task considered to

Table 4. Unstandardised and standardised coefficients of the six models of the multiple regression

Models		Unstandardised coefficients		Standardised coefficients
		B	SE	β
1	Constant	−55.02	318.57	
	Stroop C	1.61	2.78	0.05
	Stroop CW	−0.32	0.17	−0.30
	TMT A	−0.15	0.11	−0.21
	TMT B	0.05	0.42	0.12
	No. of levels completed (PS)	−9.79	2.07	−0.53***
	No. of objects placed (CL1)	−4.20	1.54	−0.31*
	No. of items located (CL2)	−0.40	1.10	−0.04
2	Constant	128.34	28.35	
	Stroop CW	−0.33	0.17	−0.31
	TMT A	−0.16	0.11	−0.22
	TMT B	0.04	0.40	0.11
	No. of levels completed (PS)	−9.91	2.03	−0.54***
	No. of objects placed (CL1)	−4.38	1.49	−0.32**
	No. of items located (CL2)	−0.29	1.07	−0.03
3	Constant	145.55	22.80	
	Stroop CW	−0.36	0.17	−0.34*
	TMT B	−0.16	0.11	−0.23
	No. of levels completed (PS)	−10.69	1.89	−0.58***
	No. of objects placed (CL1)	−4.43	1.49	−0.33**
	No. of items located (CL2)	−0.44	1.06	−0.04
4	Constant	113.86	9.45	
	Stroop CW	−0.16	0.10	−0.15
	No. of level completed (PS)	−10.23	1.90	−0.55***
	No. of objects placed (CL1)	−4.63	1.52	−0.34**
	No. of items located (CL2)	−0.55	1.08	−0.05
5	Constant	102.60	6.46	
	No. of levels completed (PS)	−11.08	1.88	−0.60***
	No. of objects placed (CL1)	−5.06	1.53	−0.37***
	No. of items located (CL2)	−0.57	1.11	−0.05
6	Constant	103.14	6.30	
	No. of levels completed (PS)	−11.23	1.83	−0.61***
	No. of objects placed (CL1)	−5.42	1.35	−0.40***

Stroop CW, Stroop colour–word test; TMT, trail-making test; PS, Parking Simulator; CL, Chemistry Laboratory.

$R^2 = 0.83$ for Model 1, $\Delta R^2 = -0.002$ for Model 2, $\Delta R^2 = -0.007$ for Model 3, $\Delta R^2 = -0.015$ for Model 4, $\Delta R^2 = -0.017$ for Model 5, $\Delta R^2 = -0.002$ for Model 6.

$*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

assess motor speed, was found to be significantly related to IVR measures which also assessed speed, such as the time taken to complete the first and second parking simulator levels and the time taken to place the blue chair, first stool and second stool in the seating arrangement task. Measures of EF, such as TMT B performance, was found to be significantly related to the number of parking simulator levels completed. Performance on Stroop CW was also

found to be significantly related to the number of parking simulator levels completed and the number of objects placed in the seating arrangement task. These correlations demonstrate that the IVR tasks used may measure the same construct as the traditional neuropsychological measures, that is EF and motor speed. It should be noted that the correlation between performance on TMT A and the time taken to place the first stool was marginally significant. A future experiment, with a larger sample, should therefore seek to replicate this result.

When comparing between groups, younger adults were significantly faster than older adults in completing TMT A. This result supports the finding of Tombaugh (12). Younger adults also committed significantly less errors than older adults in task Stroop CW, supporting the previous finding of Spieler et al. (19). In the IVR parking simulator task, younger adults were significantly faster than older adults in completing the first and second level. Younger adults completed significantly more parking simulator levels than older adults. In addition, younger adults were found to be significantly quicker in placing the chair, stool 1, stool 2 and stool 3 than older adults, in the seating arrangement task. Younger adults also placed significantly more objects than older adults. In the item location task, younger adults were significantly quicker in locating the first item than older adults. Younger adults, on average, located significantly more items then older adults. The IVR tasks in the current experiment therefore parallel the findings of previous experiments (27) who found that virtual reality tasks were able to differentiate between healthy younger and older adults. When interpreting the results of the item location task, we must acknowledge that the results could have been effected by working memory impairment. Future research should therefore study the effects of working memory impairment on task performance, unrelated to age.

The most important finding of this experiment was the contribution of IVR performance in predicting participant age. The multiple regression analysis showed that the first model which included the variables Stroop C, Stroop CW, TMT A, TMT B, number of parking simulator levels completed, number of objects placed in the seating arrangement task and number of items located in the item location task, accounted for the most variability in participant's age. However, the sixth model which only contained the IVR variables, number of levels completed of the parking simulator and the number of objects placed in the seating arrangement task, still accounted for a large percentage of the variability in age. This multiple

regression suggests the superiority of the IVR variables in predicting age and decline in cognitive function.

Given that this initial work has shown that novel IVR tasks may offer a more sensitive and ecologically valid assessment of cognitive function and daily living skills in normal ageing, future work should investigate the utility and sensitivity in assessing neurological impairment. A strength of the IVR approach is an increase in ecological validity compared with traditional neuropsychological measures which lack resemblance to real life tasks. A further advantage of this approach is the potential to identify additional indices of measurement, which may better inform neuropsychological assessment of key abilities such as the type and duration of errors made in route planning.

One of the potential limitations of this experiment may be that the superiority of younger adults over older adults on the IVR tasks could be due to their greater familiarity with interactive technology. We consider this unlikely as we implemented task training on the IVR tasks into the experimental protocol to attempt to overcome practice or familiarity effects. However, future research may enhance the above findings by investigating the correlation between age and IVR task performance in solely the older population. It is also encouraging that the two groups reported similar enjoyment in completing the VR tasks.

The methodology of the current experiment could be enhanced for future research, by the alteration of the following aspects. First, the IVR tasks were designed with a view of measuring executive function, however, they did not measure episodic memory which is viewed to be an integral part of cognitive decline. Future design may wish to add an element to the IVR tasks, which assesses episodic memory. Second, the cognitive health of participants was assessed via self-report. Self-report of cognitive health can be unreliable as an individual may not be aware of a cognitive deficit they may have. Future research may wish to employ a more valid method of assessing cognitive health. Third, the majority of IVR task outcome variables were based on response time. This was because impairments to speed of processing are thought fundamental in accounting for age-related effects on cognitive task performance (33). However, IVR technology does have the potential to investigate errors in real-world based tasks and the investigation of them in clinical groups would enhance the interpretation of task performance and the assessment of functional impairment. Finally, the reliability of this experiment could also be enhanced by recruitment of more participants as the sample size was relatively small. The addition of a third group

of participants with an age range between those aged 18–25 and 65 years and over, would also be advantageous.

It could be argued that the experiment had low ecological validity due to its implementation in a lab as oppose to a more a realistic environment. However, the methodology used, that is IVR, has previously been found to have very high ecological validity (31). It is also worth acknowledging experimental results could have been influenced by extraneous variables such as educational attainment or decreased cognitive function. Future research may wish to investigate the interaction of these variables.

A limitation in the use of VR technology in general is the potential side effect of motion sickness. Just over a quarter of participants experienced some form of motion sickness whilst completing the VR tasks. Recent technological developments suggest that motion sickness may be overcome in the near future (e.g. the Virtuix Omni Platform) (34) which would improve usability of VR in experimental and clinical research.

Concluding remark

This study suggests that IVR may provide a more sensitive and ecologically valid approach than existing traditional neuropsychological tasks in the assessment of cognitive function and daily living skills. Future research should explore the utility of IVR in larger healthy samples and in clinical groups.

Acknowledgements

Financial Support

Conflicts of Interest

None.

Ethical Standards

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and the Helsinki Declaration of 1975, as revised in 2008. Ethical approval was also received by the University of Plymouth.

References

1. Diamond A. Executive functions. Ann Rev Psychol 2013;**64**:135–168.
2. Huizinga M, Dolan CV, Van Der Molen MW. Age-related change in executive function: developmental trends and a latent variable analysis. Neuropsychologia. 2006;**44**:2017–2036.
3. Hongwanishkul D, Happaney KR, Lee WS, Zelazo PD. Assessment of hot and cool executive function in young children: age-related changes and individual differences. Dev Neuropsychol. 2005;**28**:617–644.
4. Happé F, Booth R, Charlton R, Hughes C. Executive function deficits in autism spectrum disorders and attention-deficit/hyperactivity disorder: examining profiles across domains and ages. Brain Cogn 2006;**61**:25–39.
5. Zelazo PD, Craik FI, Booth L. Executive function across the life span. Acta psychologica. 2004;**115**:167–183.
6. Baudic S, Dalla Barba G, Thibaudet MC, Smagghe A, Remy P, Traykov L. Executive function deficits in early Alzheimer's disease and their relations with episodic memory. Arch Clin Neuropsychol 2006;**21**:15–21.
7. Stroop JR. Studies of interference in serial verbal reactions. J Exp Psychol 1935;**18**:643–662.
8. Battery AI. Army individual test battery. Manual of directions and scoring. Washington, DC: War Department, Adjutant General's Office, 1944.
9. Spieler DH, Balota DA, Faust ME. Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. J Exp Psychol 1996;**22**:461–479.
10. West R, Alain C. Age-related decline in inhibitory control contributes to the increased Stroop effect observed in older adults. Psychophysiology. 2000;**37**:179–189.
11. Müller LD, Guhn A, Zeller JB et al. Neural correlates of a standardized version of the trail making test in young and elderly adults: a functional near-infrared spectroscopy study. Neuropsychologia. 2014;**56**:271–279.
12. Tombaugh TN. Trail making test A and B: normative data stratified by age and education. Arch Clin Neuropsychol 2004;**19**:203–214.
13. Parsons TD, Silva TM, Pair J, Rizzo AA. Virtual environment for assessment of neurocognitive functioning: virtual reality cognitive performance assessment test. Stud Health Technol Inform 2008;**132**:351–356.
14. Karagiozis H, Gray S, Sacco J, Shapiro M, Kawas C. The Direct Assessment of Functional Abilities (DAFA): a comparison to an indirect measure of instrumental activities of daily living. Gerontologist. 1998;**38**:113–121.
15. Elkind JS, Rubin E, Rosenthal S, Skoff B, Prather P. A simulated reality scenario compared with the computerized Wisconsin Card Sorting Test: an analysis of preliminary results. Cyberpsychol Behav 2001;**4**:489–496.
16. Rizzo AA, Schultheis M, Kerns KA, Mateer C. Analysis of assets for virtual reality applications in neuropsychology. Neuropsychol Rehabil. 2004;**14**:207–239.

17. RAND D, RUKAN SB, WEISS PL, KATZ N. Validation of the virtual MET as an assessment tool for executive functions. Neuropsychol Rehabil. 2009;**19**:583–602.

18. ZYGOURIS S, GIAKOUMIS D, VOTIS K et al. Can a virtual reality cognitive training application fulfill a dual role? Using the virtual supermarket cognitive training application as a screening tool for mild cognitive impairment. J Alzheimers Dis 2015;**44**:1333–1347.

19. KANG YJ, KU J, HAN K et al. Development and clinical trial of virtual reality-based cognitive assessment in people with stroke: preliminary study. Cyberpsychol Behav 2008;**11**:329–339.

20. JANSARI AS, FROGGATT D, EDGINTON T, DAWKINS L. Investigating the impact of nicotine on executive functions using a novel virtual reality assessment. Addiction. 2013;**108**:977–984.

21. JANSARI AS, DEVLIN A, AGNEW R, AKESSON K, MURPHY L, LEADBETTER T. Ecological assessment of executive functions: a new virtual reality paradigm. Brain Impairment 2014;**15**:71–87.

22. SOAR K, CHAPMAN E, LAVAN N, JANSARI AS, TURNER JJ. Investigating the effects of caffeine on executive functions using traditional Stroop and a new ecologically-valid virtual reality task, the Jansari assessment of Executive Functions (JEF©). Appetite. 2016;**105**:156–163.

23. RENISON B, PONSFORD J, TESTA R, RICHARDSON B, BROWNFIELD K. Virtual library task. Melbourne, Victoria, Australia: Monash University, 2008.

24. YAMAGUCHI T, FOLOPPE DA, RICHARD P, RICHARD E, ALLAIN P. A dual-modal virtual reality kitchen for (re) learning of everyday cooking activities in Alzheimer's disease. Presence 2012;**21**:43–57.

25. WERNER P, RABINOWITZ S, KLINGER E, KORCZYN AD, JOSMAN N. Use of the virtual action planning supermarket for the diagnosis of mild cognitive impairment. Dement Geriatr Cogn Disord 2009;**27**:301–309.

26. SHALLICE TI, BURGESS PW. Deficits in strategy application following frontal lobe damage in man. Brain. 1991;**114**:727–741.

27. RASPELLI S, PALLAVICINI F, CARELLI L et al. Validating the neuro VR-based virtual version of the multiple errands test: preliminary results. Presence 2012;**21**:31–42.

28. VOURVOPOULOS A, FARIA AL, PONNAM K, BERMUDEZ I, BADIA S. RehabCity: design and validation of a cognitive assessment and rehabilitation tool through gamified simulations of activities of daily living. Proceedings of the 11th Conference on Advances in Computer Entertainment Technology. 2014, vol. 11, p. 26.

29. TRENERRY MR, CROSSON B, DEBOE J, LEHER WR. Stroop neuropsychological screening test. Lutz, FL: Psychological Assessment Resources, 1989.

30. Alaska Department of Administration. Trail making test parts A & B. Available at http://doa.alaska.gov/dmv/akol/pdfs/uiowa_trailmaking.pdf. Accessed 2 June 2016.

31. LÉCUYER A, LOTTE F, REILLY RB, LEEB R, HIROSE M, SLATER M. Brain-computer interfaces, virtual reality, and videogames. IEEE Comp 2008;**41**:66–72.

32. Oculus Rift Development Kit 2 (DK2). Available at https://www3.oculus.com/en-us/dk2/ Accessed 18 August 2016.

33. SALTHOUSE TA. The processing-speed theory of adult age differences in cognition. Psychol Rev. 1996;**103**:403–428.

34. VIRTUIX. Explore virtual worlds without limits. Available at http://www.virtuix.com/ Accessed 31 May 2016.