Ecological Assessment of Executive Functions: A New Virtual Reality Paradigm

Ashok S. Jansari,[1] Alex Devlin,[2] Rob Agnew,[2] Katarina Akesson,[2] Lesley Murphy[2] and Tony Leadbetter[2]

[1] *Department of Psychology, Goldsmiths, University of London, Lewisham Way, New Cross, London SE14 6NW, United Kingdom*

[2] *School of Psychology, University of East London, Water Lane, London E15 4LZ, United Kingdom*

Acquired brain injury (ABI) can lead to a constellation of higher-order executive problems, which can impact significantly on everyday behaviour. While some neuropsychological assessments are able to measure these impairments objectively, increasingly, clinicians are finding that a subset of their patients passes these tests while still exhibiting difficulties in day-to-day living. Calls have therefore been made to develop assessments that are more sensitive and that are more ecologically valid. In this study, in Experiment 1, a multiple errands task (MET) based around a business office was created to assess concurrently nine aspects of executive functioning (planning, prioritisation, selective-thinking, creative-thinking, adaptive-thinking, multi-tasking, action-based prospective memory (PM), event-based PM and time-based PM). This new paradigm, the Jansari assessment of Executive Functions (JEF©) showed a significant difference between six individuals with ABI and matched healthy controls; further, it showed that across the nine constructs there was a range of performance. In Experiment 2, JEF© was recreated in a virtual environment resembling a computer game, and it was found that this version significantly differentiated between 17 individuals with ABI and 30 healthy controls. These results suggest that the virtual version of JEF© could be used as a new assessment of executive function. The profile of performance across the nine constructs for each individual provides a wealth of objective information that could potentially inform targeted rehabilitation.

Keywords: executive functions, ecological validity, virtual reality, assessment, multiple errand task

Introduction

Acquired brain injury (ABI) can result from traumatic or non-traumatic causes and impact upon many areas of an individual's life, including emotional regulation, affect, behaviour, cognition and social function. A common consequence of ABI is a pattern of cognitive impairment, characterised by problems in multiple areas of executive control, such as abstract thinking, planning, temporal sequencing, insight, perseveration, response inhibition, distractibility, decision making and disregard of social rules (Greenwood, Barnes, McMillan, & Ward, 2003). Typically, however, these problems can occur in the context of intact primary cognitive functions including memory, visual recognition and language. This class of higher cognitive dysfunction, previously known as the 'frontal lobe syndrome', is today referred to as 'dysexecutive syndrome' (Shallice & Burgess, 1991).

Executive functions can be defined as a set of mental processes or abilities that direct and orchestrate the activity of other cognitive systems involved in the execution of goal-directed behaviour such as memory and reasoning (Baddeley, Della Sala, Papagno, & Spinnler, 1997). Burgess and Simons (2005) and, more recently, Chan, Shum,

Address for correspondence: Ashok S. Jansari, Department of Psychology, Goldsmiths, University of London, Lewisham Way, New Cross, London SE 14 6NW, UK. E-mail: A.Jansari@gold.ac.uk

https://doi.org/10.1017/BrImp.2014.14 Published online by Cambridge University Press

Touloupoulou and Chen (2008) have provided reviews of the field, which suggest a lack of agreement on the functional architecture of the executive system. The range of theories includes those proposed by Luria (1966, 1973), Norman and Shallice (1986), Duncan (1986, for example) and Stuss and Benson (1984), with some emphasising its unitary nature while others propose multiple and more diverse systems. Equally, there seems little agreement on conceptual issues of executive functioning, such as defining which everyday tasks require activation of the purported cognitive domains and which can be completed without them. An important outcome from having this wide variety of frameworks is that it has resulted in the development of a number of differing tools to assess executive functions.

Two traditional tools are the Wisconsin Card Sorting Test (WCST), which is used to assess the ability to follow rules and respond to feedback, and the Controlled Oral Word Association Test (COWAT), which assesses verbal fluency. These have been used for decades and had been assumed to test the integrity of frontally based cognitive functions. However, there are cases in the literature where despite performing at normal levels on standard clinical tests of executive functions, individuals with ABI do, in fact, show signs of impaired executive functioning in their daily life (Damasio, 1996; Eslinger & Damasio, 1985). There are various ideas why this failure to detect impairment in some individuals may occur. One reason could be due to the supposed difference between so-called 'hot' components of executive function (e.g., those involving emotions, beliefs and regulation of social behaviour) and 'cold' components that involve mechanistic logical functions such as attention, resistance to interference, cognitive flexibility and dealing with novelty (Grafman & Litvan, 1999). It may be that the appropriate tests are not being used for a particular patient's specific dysfunction. This relates to the fact that most tests only assess isolated parts of the system and therefore any one single assessment will be insufficient to fully capture an individual's range of problems. In line with this, following a comparison of a range of different clinical assessments, Bennett, Ong and Ponsford (2005) have suggested that multiple tests are needed to detect executive dysfunction in clinical populations.

In addition to the need for more than one form of assessment, it has also been argued that perhaps the tests themselves are not robust enough. In a review of the utility of neuropsychological tests, Chaytor and Schmitter-Edgecombe (2003) point out that a very important issue is the ecological validity of assessments, given that a major purpose of the clinical referral is to make recommendations about everyday functioning. This concurs with earlier comments made by Shallice and Burgess (1991), who argued that the tests themselves do not fundamentally resemble the actual tasks that individuals might have difficulty with. Further issues included standard tests being too linear or simplistic by only involving one task at a time; task initiation is prompted rather than self-generated; there is little or no prioritising involved between competing tasks; planning behaviour required for task completion occurs over only seconds or minutes, whereas in life it more often occurs over much greater timescales ranging from seconds to months and successful completion is clearly characterised. Overall these tasks bear 'little resemblance to the ill-structured activities that characterize modern lives' (Bennett et al., 2005, p. 606).

One major theme highlighted above is the reductionist approach to the testing of executive functions. For example, the WCST and COWAT tests tap relatively separate aspects of executive behaviour and are, of course, administered quite independently of one another. However, if the executive system is conceived as the orchestrator and director of cognitive subsystems, is this approach valid? One might question a musical conductor who only heard each instrumental section rehearse separately in an attempt to assess how well the entire orchestra might execute the final symphony in concert. Similarly, deconstructing the executive system and assessing its functionality as singular, independent systems may not give an accurate representation of the integrity of the functioning of the executive system as a whole. In other words, the tests do not petition cognitive systems in the same way that they would be challenged in real life.

Efforts have been undertaken to construct new tests that can address these issues by more comprehensively taxing frontal functions. In Shallice and Burgess's (1991) seminal study, a real-world paradigm was devised as a departure from the traditional 'pen and paper' style neuropsychological test; the task involved performing multiple errands in a busy shopping centre and found that participants who had performed normally on the WCST were significantly impaired in the real-world scenario. The study had three main impacts: first, it lent support to the clinical observation of individuals being able to perform well on established cognitive tests despite experiencing significant problems in real-life settings, including activities of daily living and gross life functions, such as holding down a job. Second, it brought into sharp focus the necessity to devise realistic tests of higher

https://doi.org/10.1017/BrImp.2014.14 Published online by Cambridge University Press

cognitive functions. Finally, it introduced a new paradigm of neuropsychological and cognitive research in rehabilitation, the Multiple Errands Task (MET).

Since Shallice and Burgess's (1991) landmark study, a number of studies have created versions of the MET for clinical use. Knight, Alderman and Burgess (2002) developed a simplified MET for use in a hospital setting and found that this differentiated 20 people with ABI from an equal number of matched healthy controls. Alderman, Burgess, Knight and Herman (2003) then developed a simpler form of the original shopping task, known as the MET Simplified Version (MET-SV), and found that the pattern of errors made by individuals with ABI mirrored those that were observed in their everyday lives. While being successful in demonstrating that this new paradigm was important for accurate assessments that mapped to real-world behaviour, there are pragmatic reasons that can limit their use as general clinical tools. While affording more ecological validity, the real-life setting reduces the clinician's control, such that individuals can be exposed to potentially emotionally distressing situations (for example, if approached by strangers, as could happen in a shop) and the opportunity for accidents and misunderstanding to occur is elevated. The time involved in moving patients to suitable settings could also be considerable and this itself will incur costs on clinical time, something that assessment and rehabilitation centres may be struggling with anyway. As pointed out by Rand, Basha-Abu Rukan, Weiss and Katz (2009), patients with motor deficits could have particular difficulties being assessed in some settings unless these settings are set up for such individuals. In response to some of these issues, elements of the multiple errands paradigm were later incorporated into the Behavioural Assessment of Dysexecutive Syndrome (BADS; Wilson, Alderman, Burgess, Emslie, & Evans, 1996). The BADS was designed as an ecologically valid measure of executive impairment and its validity has been demonstrated many times (e.g., Wilson, Evans, Emslie, Alderman, & Burgess, 1998), and it is widely used, forming a standard instrument in the neuropsychological test library of many clinicians, services and organisations.

A different approach to addressing the challenges of conducting ecologically valid assessments while reducing some of the pragmatic difficulties has been the use of virtual reality (VR) technologies. VR is characterised by the creation of a three-dimensional computer environment that can simulate a real-world setting in terms of physical surroundings, objects and events. There are a number of reasons why VR has proven to be a useful tool in the effort to create reliable measures of executive function (Brooks & Rose, 2003; Rizzo & Kim, 2005). The facility to assess an individual while he or she is immersed in an environment offers the opportunity to observe behaviour in safe but taxing situations. Further, it is possible to enforce strict experimental control over pertinent variables in a way that a real-life setting does not allow. From a patient's perspective, such tasks can also be more engaging than traditional paper-and-pencil tasks, which seem divorced from their real-world problems. Finally, research has shown that VR assessments yield results that are replicable across laboratories and populations, and that they are *at least* as sensitive to cognitive functions as 'real-life' tests (Pugnetti et al., 1998).

Appendix A provides a review of many of the studies that have used VR to assess executive functions. As can be seen, these have varied in terms of populations sampled, the environment, the types of measures taken and the purported executive functions that are assessed. The studies have been successful and, for example, the Virtual Multiple Errands Test (VMET) created by Rand et al. (2009) was able to differentiate successfully between post-stroke patients and matched controls, as well as performance correlating well with the original Multiple Errands Task (MET; Knight et al., 2002).

An important limitation, however, pertains to the dependent measures collected in such studies. For example, Rand et al. (2009) and Klinger, Chemin, Lebreton and Marié (2006) used a VR task modelled on the original Shallice and Burgess (1991) shopping task and their analysis was limited to the number of errors, including those in completing tasks, rule breaks and what they referred to as 'non-efficiency' mistakes. While very useful in showing that their VMET task differentiated stroke patients from healthy controls, in terms of targeted rehabilitation it is difficult to see how these measures map on to the everyday problems that form the constellation of the dysexecutive syndrome. Similarly, Elkind et al. (2001) used a VR-based assessment that was simply an electronic form of the WCST and therefore, presumably, prone to many of the same difficulties encountered by the paper version of this neuropsychological test. Titov and Knight (2005) developed a VR assessment that uses a Virtual Street scene; however, the participant is only scored on the efficiency with which they remember to record different pieces of information as requested by the tester. It therefore offers a rather narrow set of dependent measures linked to prospective memory and not wider executive functions.

https://doi.org/10.1017/BrImp.2014.14 Published online by Cambridge University Press

Considering the strengths and weaknesses of the range of VR assessments available to date, a number of qualities are implied as a 'wish list' for a sensitive and informative assessment of executive functions. The environment needs to be realistic and the overall goals should not be linear. The assessment should have competing tasks that together assess a number of abilities concurrently. The tasks should not be overly structured and solutions should not always be obvious. Finally, to assess planning and behaviour, which are at least somewhat equivalent to real-world tasks, the assessment should not be too short.

The current study was an attempt to build on previous work using VR to create an ecologically valid assessment of executive functions. To address the limitation of dependent measures, the task was created such that participants were required to complete a number of tasks in a non-linear fashion and often with competing demands that required them to plan and prioritise their actions in order to achieve the stated goals. Problems and their corresponding solutions were chosen that reflected common tasks and then 'woven' into a realistic scenario.

A major improvement of this study was that the executive constructs investigated were clearly and operationally defined in order to promote further research based on this study, and to ensure that important functions were investigated in a way decipherable to clinicians, researchers and professionals working in the rehabilitation of individuals with dysexecutive syndrome. The tasks used in the assessment were based on these cognitive constructs and each construct was assessed at least once so as to create a profile of executive-type abilities for each individual. Performance on these tasks was rated on accuracy using a three-point scale (0, 1, 2) thereby reflecting the efficiency with which the participant fulfilled a task.

In summary, the aims of the current study were twofold. The first aim was to create a new real-world task that differentiates between individuals exhibiting dysexecutive syndrome and matched controls, and which yields measures that can be used to guide neuropsychological or vocational rehabilitation. The second aim was to evaluate whether such differences are maintained in a virtual version of this task. To address these aims, two studies were conducted; in Experiment 1, a real-life task was created and administered to a small group of individuals with ABI ($N = 6$) and matched controls; in Experiment 2, the task was recreated in a virtual environment and administered to a larger group ($N = 17$) of individuals with ABI and matched controls.

Experiment 1

Design

The study was an experimental design with group (ABI or healthy control (HC)) being the between-subjects independent variable and scores on the new assessment of executive functions (see below) as the dependent variables.

Participants

Experimental participants in the ABI group were recruited from a vocational rehabilitation centre in London, Rehab UK, and all had acquired brain injury of mixed aetiologies, showed difficulty in running their own affairs as reported by significant others and also exhibited classical signs of executive dysfunction; this was not formally defined but was decided by psychologists working with the clients. Due to the eventual aim of aiding clients' return to work, individuals were only admitted to the rehabilitation programme if they were medium- to high-functioning and reasonably independent (for example, were able to travel to the centre unaided). Participants were of varying age and educational levels, with different aetiologies of brain damage. Where possible from referral notes, performance on the BADS and Brixton and Hayling tests (Burgess & Shallice, 1997) was recorded (see Table 1 for demographic and neuropsychological characteristics). Consistent with the study of McGeorge et al. (2001) the main inclusion criterion for participating in the current study involved not falling in the 'impaired' range on the BADS and (for individuals who spoke English as a first language) having a Wechsler Test of Adult Reading (WTAR) IQ in the average range. No participants had specific language or reading problems.

Six healthy individuals (5 male, 1 female) with no known history of neurological or psychiatric illness served as a HC comparison group. There were no significant differences between the ABI and HC groups in terms of age (ABI mean 31.8 (*SD* 11.6), HC mean 30.8 (*SD* 9.6); $t(10) = .16$, $p > .05$) or WTAR IQ (ABI mean 108.3 (*SD* 10.8), HC mean 113.3 (*SD* 2.7); $t(7) = 1.22$, $p > .05$).

Measures

Task Development: Jansari assessment of Executive Functions (JEF©). While a number of other VR tasks have used a shopping scenario (Appendix A), given that as part of their rehabilitation most of the clients at Rehab UK attempted work placements in office environments, this was the scenario chosen. The advantage of this scenario is that, unlike shopping where generally there is a

https://doi.org/10.1017/BrImp.2014.14 Published online by Cambridge University Press

TABLE 1

Patient Demographics and Available Neuropsychological Information for ABI Participants in Experiment 1 (NB: The WTAR IQ was not conducted on HV, RW and BA since English was not their first language). Please note that information was extracted from existing files and in some cases limited information was available

Participant	Age	Gender	WTAR IQ	BADS	Brixton	Hayling	Probable area of brain damage
HV	26	M	–	Borderline	–	–	Right fronto-parietal
BH	21	M	96	Low average	–	Average	Right fronto-parietal
DG	24	M	111	Average	High average	Average	Frontal horn atrophy, occipital, temporal
RW	47	F	–	Average	Superior	Moderate average	Anterior left temporal
NS	27	M	117	Average	Very superior	Moderate average	Bilateral occipital, right frontal
BA	46	M	–	Borderline	Low average	Impaired	Unknown

WTAR IQ, Wechsler Test of Adult Reading IQ; BADS, Behavioural Assessment of Dysexecutive Syndrome.

finite start point and end point, in an office, it was possible to include a number of competing tasks, making the overall assessment less linear and therefore less likely to mask or mediate the difficulties participants may be likely to experience in the work place. It should be noted that no actual knowledge of how an office functions is needed to complete the assessment. Next, cognitive constructs that have been linked to the dysexecutive syndrome, either from a theoretical or behavioural perspective, were identified. This involved discussions between the researchers, vocational job coaches at the rehabilitation centre and a clinical neuropsychologist. As much as possible, an attempt was made to map these on to real-world behaviour, and to be ones that the vocational coaches or rehabilitation therapists could use. Deficits in planning ability and problem solving feature heavily in much dysexecutive syndrome literature (e.g., Burgess, Veitch, de Lacy Costello, & Shallice, 2000; Lezak, Howieson, & Loring,, 2004; Mateer, 1999; Shallice & Burgess, 1991) so constructs were identified that included both direct parallels to these and closely associated analogues, namely planning, prioritisation, selection, adaptive-thinking, creative-thinking and multi-tasking. In addition, since it has been suggested that the differences in various types of prospective memory are potentially significant (Ellis, 1996; Ellis & Kvavilashvili, 2000), three separate constructs were identified, namely action-based prospective memory, event-based prospective memory and time-based prospective memory. In total, there were nine constructs, each of which was given an operational definition. For each of these nine constructs, realistic tasks that could be found in an average office environment were chosen. For planning, prioritisation, creative-thinking, selective-thinking, creative-thinking and adaptive-thinking, two tasks were devised in order to tax these constructs as ecologically as possible; one task was devised for each of the remaining constructs (see Appendix B for operational definitions of each construct and an example of a task for each construct). Tasks were designed to have (or at least appear to have) ambiguous and multiple solutions, as is the case in real-life situations, so that solutions were not always immediately obvious.

Briefly, the task involved the participant being invited to play the role of an office-temp worker whose main goal was to organise a meeting later that day and to prepare an appropriate room for this meeting. Participants were informed that the manager had left a list of tasks for them to complete and that all information could be found in their office. The responsibility of planning for overall task

https://doi.org/10.1017/BrImp.2014.14 Published online by Cambridge University Press

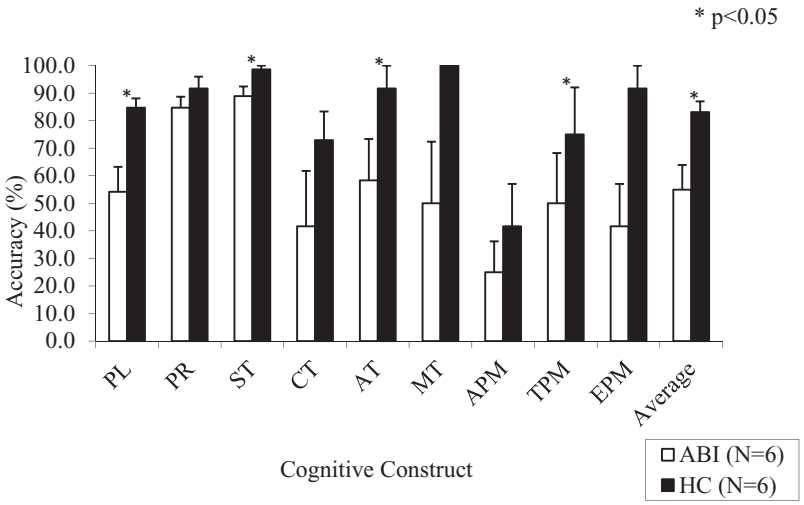


FIGURE 1

Performance on the real-life version of JEF© (Experiment 1) as a function of group and cognitive construct. PL, planning; PR, prioritisation; ST, selective-thinking; CT, creative-thinking; AT, adaptive-thinking; MT, multi-tasking; APM, action-based prospective memory; TPM, time-based prospective memory; EPM, event-based prospective memory, Average, average score across all constructs. Error bars represent one standard error.

completion was given to the participant with no clues as to possible solutions or courses of action.

A corridor at the University of East London was used for recreating the scenario necessary for the task. This consisted of a small office-like room linked by a corridor to a larger room that would be appropriate for holding a meeting for 20 people. The items required for the tasks defined the contents of these rooms, and additional relevant but non-used items (such as staplers and extra desks) were included for realism.

Procedure

After initial piloting with healthy individuals, each experimental participant was taken to the small office and informed of the role that they had to play to complete the task. They were informed that the experimenter would mostly read from a pre-prepared script and that questions could be asked at any point during this reading but that the experimenter would only be able to clarify instructions and would not be able to provide any additional information. As the participant completed each task, his or her performance was recorded and scored on a clearly delineated scoresheet that was broken down by cognitive construct. All tasks were scored on a three-point scale for success; 0 for failure, 1 for a partial or non-optimal completion and 2 for satisfactory completion. For example, for the prioritisation task that requires the ordering of five agenda items in order of importance, two of these are urgent; to score 2 points, these items need to be placed in the first and second positions, while if only one of them is placed in these positions, only a 1 can be awarded. If the participant failed to attempt a task or their performance did not meet the criteria for a 1 (e.g., if in the agenda task neither of the two urgent items was put in the first two places), then a score of 0 was given. In total, JEF© took approximately 40 min to complete.

Data Analysis

The scores for the separate tasks were summed and since some of the constructs involved one task while others included two, to allow comparisons, a percentage score was calculated for each individual construct. An overall JEF© Average performance score was calculated as the mean average of the construct percentage scores. Overall, ten scores were derived for each participant, nine for the individual constructs and one for the JEF© Average performance score. Due to the small numbers in each group, Mann–Whitney *U* tests were used for comparing the two groups.

https://doi.org/10.1017/BrImp.2014.14 Published online by Cambridge University Press

Results

The Mann–Whitney *U* test revealed a significant difference between the ABI and HC groups for JEF© Average score ($U = 31.5$, $N = 12$, $p = .03$). Analysis on the individual constructs revealed that there were differences between the two groups for planning ($U = 32.5$, $N = 12$, $p = .018$), selective-thinking ($U = 31$, $N = 12$, $p = .023$), adaptive-thinking ($U = 32$, $N = 12$, $p = .02$), and time-based prospective memory ($U = 31$, $N = 12$, $p = .023$). As can be seen in Figure 1, on all of these constructs, the ABI group performed more poorly than the matched controls.

Experiment 2

With Experiment 1 showing that the office task was able to differentiate individuals with ABI from matched controls, Experiment 2 aimed to recreate the task in VR. The goal was to see if assessment using a virtual environment could maintain the difference observed in real-life. A computer-based assessment would be more feasible than the real-world task, being more practical and accessible to clinicians and other researchers.

Research Design

As in Experiment 1, this study was an experimental design with group (ABI or HC) being the between-subjects independent variable and scores on the new virtual reality version of JEF© as the dependent variables.

Participants

The six brain-injured participants from Experiment 1 and 11 new participants from Rehab UK were recruited. Thirty new healthy matched controls also participated. See Table 2 for demographic and neuropsychological characteristics of new ABI individuals.

There were no significant differences between the ABI and HC groups in terms of age (ABI mean 33.7 (*SD* 9.1), HC mean 33.3 (*SD* 10.2); $t(45) = .14$, $p > .05$) or WTAR IQ (ABI mean 109.1 (*SD* 12.7), HC mean 110.7 (*SD* 9.2); $t(43) = .43$, $p > .05$).

Measures

An exact three-dimensional replica of the office and corridor used in Experiment 1 was created in a computer-generated non-immersive VR environment using Visual Basic with the addition of special 3D add-on software known as 3d State. See Figure 2 for a screenshot of the main meeting room.

TABLE 2
Patient Demographics and Available Neuropsychological Information for new ABI Participants who took part in Experiment 2

Participant	Age	Gender	WTAR IQ	BADS	Brixton	Hayling	Probable area of brain damage
EG	39	M	120	Average	Very superior	Average	Left focus (no other details)
EH	30	M	117	Low average	Good	Poor	Frontal and parietal haematoma
BM	34	M	120	Average	Average	Average	Left temporal-parietal
JD	24	M	85	Average	–	Moderate average	Anterior left temporal
PM	40	M	94	Borderline	–	Abnormal	Bilateral occipital, right frontal
JC	24	F	110	Average	–	–	Unknown
SP	42	M	113	–	–	–	Right fronto-temporal
KS	47	M	119	–	–	–	Unknown
GS	40	M	117	Average	Average	Average	Right frontal
JT	25	M	111	–	–	–	Unknown
DG	38	M	122	Borderline	–	–	Diffuse

WTAR IQ, Wechsler Test of Adult Reading IQ; BADS, Behavioural Assessment of Dysexecutive Syndrome.

https://doi.org/10.1017/BrImp.2014.14 Published online by Cambridge University Press



FIGURE 2

(Colour online) Screen shot of the main meeting room as it appears at the beginning of the assessment. The door in the far left-hand corner of the room leads to a corridor with one of the doors on the left opening into the small office where the participant begins the assessment.

Procedure

To ensure that performance on the new assessment was not affected by lack of experience with using computers and moving around a virtual environment, before being tested on the experimental paradigm, participants familiarised themselves with another virtual reality environment that had been created as a training tool for a commercial business. This involved moving around a virtual environment, becoming aware of objects within it and interacting with those objects. Navigation around the environment was by use of the arrow keys on a standard computer keypad and objects could be picked up by simply clicking them with the computer mouse. When the participant reported that he or she had become familiar with interaction in the practice virtual environment, formal testing began.

As in the real-life phase, the script was read to the participant; all rules and procedures were identical to those used in the real-life test. The only aspects of the test that required physical interaction outside the virtual environment were those that involved filling out of certain lists (e.g., the initial to-do list). A decision was made to perform these writing tasks on hard copy rather than in the virtual environment for ease of subsequent scoring of responses. Scoring was conducted in the same way as in the real-life test and questions were addressed similarly. This study was conducted 3 months after the real-world experiment, which meant that the chances of remembering performance in that version for the ABI participants who performed both tasks were minimal.

Data Analysis

As in Experiment 1, there were ten scores for each participant, comprising separate scores for the nine constructs and one average score. Due to the larger number of participants in this study, *t*-tests were used to examine differences between the two groups. One of the advantages of JEF© is the evaluation of a number of cognitive constructs concurrently, which allows a level of fine-grained analysis not afforded by other assessments. To demonstrate this, the performance for individual constructs and overall performance of the individual ABI participants was standardised to *z*-scores relative to the means and standard deviations derived from the HC group.

Results

An unrelated *t*-test on the overall Average JEF© score comparing the ABI and HC groups revealed a significant difference between groups ($t(45) =$

https://doi.org/10.1017/BrImp.2014.14 Published online by Cambridge University Press

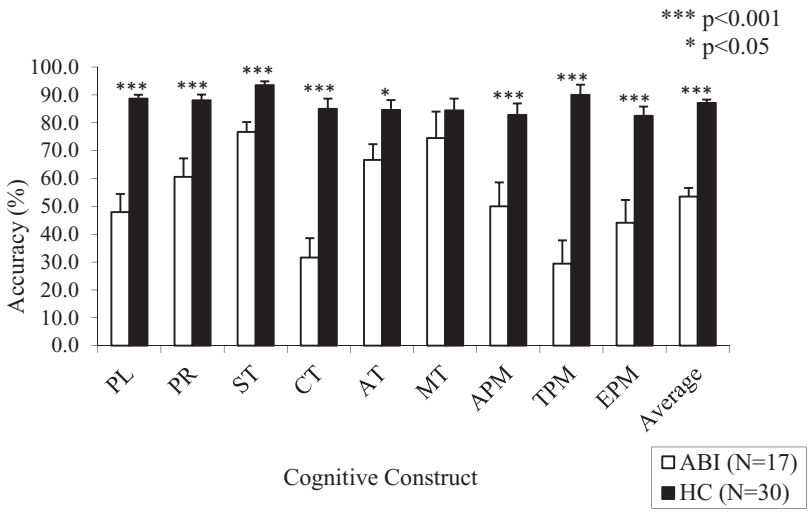


FIGURE 3

Performance on the virtual reality version of JEF© (Experiment 2) as a function of group and cognitive construct. PL, planning; PR, prioritisation; ST, selective-thinking; CT, creative-thinking; AT, adaptive-thinking; MT, multi-tasking; APM, action-based prospective memory; TPM, time-based prospective memory; EPM, event-based prospective memory, Average, average score across all constructs. Error bars represent one standard error.

11.34, $p < .001$) with the ABI group performing significantly worse than the control group (see Figure 3). For individual constructs, there were differences between the two groups for planning ($t(45) = 7.69$ $p < .001$), prioritisation ($t(45) = 4.66$, $p < .001$), selection ($t(45) = 4.96$, $p < .001$), creative-thinking ($t(45) = 6.98$, $p < .001$), adaptive-thinking ($t(45) = 2.58$, $p = .013$), action-based prospective memory ($t(45) = 3.62$, $p < .001$), event-based prospective memory ($t(45) = 7.18$, $p < .001$) and time-based prospective memory ($t(45) = 4.83$, $p < .001$). As can be seen in Figure 3, on all of these constructs, the ABI group performed worse than the matched controls.

Since six of the ABI participants had participated in Experiment 1, to check whether a practice effect of having done the real-life test improved their performance relative to those who did not, the overall performance on the virtual task was compared for the two groups. A Mann–Whitney test revealed no significant difference between the two groups ($U = 40.0$, $N = 17$, $p > .05$).

Table 3 presents the z-score analysis in three ranges, namely 2–3.9, 4–7.9 and 8 standard deviations from the mean of the HC group. It is clear that although a patient can be impaired overall on the task (the Average score), that does not equate to impairment across all domains. Furthermore, while two patients can perform similarly on the task overall (e.g., HV and BM who both had an Average z-score of 2.3), their underlying performance across the individual constructs can vary substantially; this variance between two individuals is the clinical pattern that is often observed. The fact that there is at least one patient who is impaired for each construct, justifies the inclusion of each construct; similarly, for no single construct are all patients impaired, meaning that none of the tasks in isolation provides a measure of general brain injury. Finally, it is possible for an individual (e.g., JC) to perform within normal limits on the task overall (i.e., a z-score of less than 2) and yet have underlying impairments on individual constructs.

Discussion

The current study involved the creation of a new MET based in an office environment that addressed the limitations previously highlighted in the assessment of executive functions. A real-life version of the JEF© task, where participants physically moved around an office environment, was able to differentiate the performance of a small group of individuals with ABI from that of matched healthy controls. Importantly, all of the individuals in the ABI group had performed at 'borderline' or above on the BADS assessment and

https://doi.org/10.1017/BrImp.2014.14 Published online by Cambridge University Press

TABLE 3

Z-Score Ranges for Individual Patient Performance on JEF© as a Function of Cognitive Construct (PL, planning; PR, prioritisation; ST, selective-thinking; CT, creative-thinking; AT, adaptive-thinking; MT, multi-tasking; APM, action-based prospective memory; TPM, time-based prospective memory; EPM, event-based prospective memory; Average, average score across all constructs; NCI, number of constructs impaired for each individual ABI patient; NPI, number of ABI patients impaired on each construct). – = <2 (within normal limits), * = 2–3.9, ** = 4–7.9, *** = >8

Patient	PL	PR	ST	CT	AT	MT	APM	TPM	EPM	Average	NCI
HV	**	–	–	–	**	–	–	–	–	*	2
BH	***	**	–	–	–	*	–	*	*	**	5
DG	*	**	**	–	**	–	–	*	–	**	5
RW	**	*	–	*	**	–	*	–	–	**	5
NS	*	**	–	–	*	–	–	*	–	*	4
BA	**	**	*	*	–	–	*	*	*	**	7
EG	–	–	–	**	**	*	*	*	*	**	6
EH	**	–	–	**	–	*	–	*	–	**	4
BM	***	–	–	–	–	–	–	–	–	*	1
JD	***	–	–	–	–	–	–	*	–	*	2
PM	***	*	–	**	**	–	–	*	–	**	5
JC	*	–	–	–	**	–	–	–	–	–	2
NPI:	11	5	2	4	6	3	3	8	3	11	

only one performed poorly on either the Brixton or Hayling tests. In addition to an overall assessment, analysis of individual cognitive constructs revealed that JEF© was able to detect differences in specific executive functions. A VR replication of the study with a larger number of participants showed that the difference between the ABI and HC groups was maintained and, possibly because of a larger pool of data, significant differences were found between more of the individual constructs. In light of the largely unimpaired performance of the participants on two well-known clinical tests, the preliminary suggestion from these findings is that JEF© is an ecologically valid assessment that identifies the real-world impairments described by the individuals participating in this study. Given the oft-cited difficulties in quantifying these real-world problems, JEF© could potentially contribute significantly to clinical evaluation of dysexecutive syndrome in individuals with preserved IQ.

The individual-level analysis presented in Table 3 is an advance on previous virtual METs in that it allows a fine-grained analysis of an individual's strengths and weaknesses that previous assessments can not. This may help to explain the clinical picture that is often seen, where real-world behaviour suggests profound problems that are not picked up by the standard assessments. One of the most famous examples of this is patient EVR who underwent removal of a tumour in the orbitofrontal region of the brain (Damasio, 1996; Eslinger & Damasio, 1985). Despite the invasive surgery, EVR's verbal IQ was found to be 'superior', his performance IQ was average and his memory performance was above average. Yet in his own affairs he could not manage his life effectively, became obsessed with the smallest mundane activities and, despite having been an accountant, managed to become bankrupt following surgery. However, in the context of such obvious brain damage and changes in everyday behaviour, it was found that the most robust tests of frontal lobe function did not seem to detect any difficulties; he performed well on them without the impairment expected from an individual with categorical frontal lobe damage. As an extension of this, it is possible to be in the unimpaired range on a task but have significant underlying impairments. JC, for example, performs normally on the BADS but is seen to have particular difficulties with the planning and adaptive-thinking sub-constructs on JEF© despite being within normal limits overall on the task.

An example of the sensitivity of JEF© is its applicability in other clinical areas outside the realm of ABI. For example, Mills (2009) used it to assess men with prostate cancer undergoing androgen-deprivation therapy (ADT), which is one of the main treatments in men when surgical removal of the tumour is not sufficient. Anecdotally, cognitive problems have been mentioned in the literature but never been demonstrated objectively, despite

https://doi.org/10.1017/BrImp.2014.14 Published online by Cambridge University Press

attempts to do so (Cherrier, Aubin, & Higano, 2009), but Mills was able to demonstrate a significant decrement in executive performance using JEF©.

An extension of the versatility of this new assessment comes from the fact that it has been created to assess executive functions rather than solely to detect *dysfunction*. As pointed out by Bottari, Swaine and Dutil (2007), even professionals such as occupational therapists have can have difficulty interpreting errors, and therefore subjective observation of errors may not be sufficient. They suggest that more performance-based measures need to be developed. JEF© provides such a tool by being based on objectively defined criteria of performance which aren't based solely on errors but on correct performance as well. The scoring criteria allow for perfect performance (2), incomplete performance or partial errors (1) or failure (0). The result of this is to make the assessment sensitive at higher levels of performance. As such it can be used as an experimental tool in fields outside standard neuropsychology of ABI. Further, because the virtual reality version of JEF© created in Experiment 2 is able to detect differences between individuals with ABI and healthy controls, this version can be used instead of the real-life setup, which involves considerably more physical resources. For example, given its 'portability', Montgomery, Hatton, Fisk, Ogden and Jansari[1] (2010) used this virtual version of JEF© to show the detrimental impact of recreational ecstasy in healthy young adults. Similarly, Montgomery, Ashmore and Jansari[1] (2011) have shown that JEF© is sensitive enough to detect the impact on executive functions of even a modest dose of alcohol that was within the legal limits to allow an individual to drive in the UK. Recently, Montgomery, Seddon, Fisk, Murphy and Jansari[1] (2012) have shown that JEF© is sensitive to executive-function deficits and, in particular, problems with prospective memory in cannabis users. Finally, Jansari, Froggatt, Edginton and Dawkins (2013) used the assessment in a single-blind administration of nicotine or placebo to smokers and never-smokers to demonstrate impairments in executive functions in abstinent smokers and that this deficit can be reversed with nicotine; these findings may have important implications for theoretical issues regarding the impact of nicotine in smokers. A common theme across these differing literatures is that tools that have already been found to be insensitive to the impact of brain damage have been used in populations where the effects may be even more subtle. It is possible that the greater sensitivity of JEF© has allowed these effects to be revealed.

Given the applicability of JEF© to address issues of importance outside the clinical realm, at a theoretical level, it will be useful to see how performance across the constructs of JEF© maps on to models of executive functioning. For example, whether the aspects of dysexecutive syndrome measured by JEF© cluster around specific theoretical concepts as suggested by Stuss and Alexander (2000). This may then help to predict the possible impact of different aetiologies of brain damage on executive performance. Further, given the purported involvement of executive functions in a number of disorders, such as schizophrenia and attention deficit hyperactivity disorder (ADHD), it would be predicted that there would be different JEF© profiles of performance as a function of population, as found by Josman, Klinger and Kizony (2008) using their virtual assessment.

Limitations

To consider use of JEF© more widely, it is necessary to address a number of limitations. As can be seen in Figures 1 and 3, in both the real life and VR versions of the task, on all cognitive constructs as well as on the total score, the ABI group performs worse than the HC group. However, for the real-life task, there were significant differences on only four out of the nine individual constructs while in the VR version, this was the case for eight. This difference may be due to the larger sample size, or it may be that the current virtual reality JEF© assessment is more sensitive to some executive functions than others; larger-scale replications are necessary to address this possibility.

Further, it will be important to establish the psychometric properties of the assessment, as the current studies have generated a number of questions. What is the concurrent validity of JEF© against subscales of standard batteries including subscales of the BADS, the Delis–Kaplan Executive Functions System (D-KEFS) and other specific measures of executive function, such as the Brixton–Hayling and the Cambridge Prospective Memory Test (CAMPROMPT)? What 'added value' might JEF© give over and above these standard tests? Chaytor, Schmitter-Edgecombe and Burr (2006) have shown that a combination of the Trail-Making Test, the WCST, Stroop and COWAT was only able to account for 18–20% of the variance in everyday executive ability as measured by two rating scales, the Dysexecutive Questionnaire (DEX) and Brock Adaptive Functioning

[1] In previous papers, the JEF© was known by a different acronym, JAAM; however, the assessment is unchanged between the different studies.

https://doi.org/10.1017/BrImp.2014.14 Published online by Cambridge University Press

Questionnaire. Bennett et al. (2005) compared performance on a number of different tests to ratings on the DEX, and suggested that because no single test made a unique contribution to DEX ratings, multiple tests taken from the BADS and other sources should be used with clinical populations. In a preliminary study, Renison, Ponsford, Testa and Jansari (2008) compared the performance of JEF© to the Brixton Spatial Anticipation Test, Verbal Fluency, the Modified Six Elements Test and the Zoo Map. They found that after controlling for IQ, verbal memory, education and performance on all the neuropsychological measures of executive function, performance on JEF© still significantly differentiated their ABI patients from controls, suggesting that it may have captured more complex aspects of executive dysfunction than the standard neuropsychological measures.[2] Indeed, Renison, Ponsford, Testa, Richardson and Brownfield (2012) have gone on to use the JEF© paradigm to create a virtual library task for further clinical research with their patients.

To assess the inter-rater reliability, in unpublished data, two assessors rated the performance of five individuals with ABI, and a Pearson *r* correlation coefficient was conducted between the construct scores that were awarded independently by each rater. This found highly significant correlations between the two raters for the assessment of each individual, ranging from $r = 0.956$ to $r = 1.0$. This small-scale finding needs to be extended but suggests a robust scoring system for JEF©. Finally, test–retest reliability needs to be established to allow researchers or clinicians to evaluate performance at two different time points, possibly to look at the impact of an intervening rehabilitation regime.

A final limitation is that as the ABI individuals who participated in the current study were of medium to high functioning, the current form of JEF© is unlikely to be suitable for lower-functioning patients. Masson, Dagnan and Evans (2010) showed that since the standard Tower of London test of planning and problem-solving was not appropriate for individuals with intellectual disabilities, it was possible to adapt the task for this population, but the same rationale could be used to create less taxing assessments. Along similar lines, using the rationale for development of JEF©, a less taxing version could be developed for the assessment of lower-functioning individuals.

[2] It should be noted that in this study, the ABI group was not part of a vocational rehabilitation programme and therefore use of computers was not expected. Despite this lack of extensive experience in using computers, it was still possible to complete the task. This demonstrates that while JEF© has been created to be used on a standard laptop, a participant does not need to be familiar with computers to allow an assessment to be conducted.

Clinical Implications

JEF© has the potential to be used as a more ecologically valid assessment by clinicians, especially for patients who pass the current tests. The rich performance profile across the constructs for individual participants could be used for targeting neuropsychological or vocational rehabilitation. For example, if mainly deficits in prospective memory are found, then resources could be aimed specifically at improving this skill, but if this is not possible, then attempts can be made to avoid work or everyday scenarios where this impairment will be exposed. Also, parallel versions of the virtual assessment could be used to examine the impact of interventions in a controlled environment before an individual is exposed to real-world scenarios. Intriguingly, Gega, White, Clarke, Turner and Fowler (2013) have used VR in conjunction with Cognitive Behavioural Therapy (CBT) in socially anxious individuals, to help them see the impact of avoidance behaviours in the maintenance of their anxieties. Using such a rationale, it may be possible to use JEF© to explore the interaction between emotion and executive functions. For example, the role of anxiety on performance could be assessed and, if there is an impact, anxiety-reduction strategies could be employed in a safe environment before the individual has to cope in the outside world.

Future Directions

In addition to addressing the limitations outlined above with regards to replication on larger groups and evaluating concurrent validity, a number of other directions can be taken, expanding on this work. For clinicians who want to assess the impact of a rehabilitation regime, to avoid practice effects a parallel version that could be used post-rehabilitation needs to be developed. In addition, if clinicians feel that the general paradigm can be of use to their clients who are of lower functioning than the ones included in the current studies, then an important direction is the development of simpler versions of JEF©. To allow clinicians to use JEF© it is necessary to translate it into other languages and assess validity in the new cultures. Preliminary studies have found that both Swedish (Jansari, Debreceni, Bartfai, & Eriksson, 2008) and French (Jansari, Sasson, & Samson, 2013) versions of the assessment discriminate between brain damaged and healthy individuals in those cultures. Similar replications are underway using

https://doi.org/10.1017/BrImp.2014.14 Published online by Cambridge University Press

Italian, Dutch, Finnish and (Brazilian) Portuguese versions of JEF©.

Finally, given the fact that executive functions are amongst the last to fully mature in the developing brain, a children's version of JEF©, known as JEF-C©, has recently been developed. This assessment involves a child's birthday party as the environment and preliminary results show that it is able to track the developmental trajectory of executive functions in children ranging in age from 7 to 18. Furthermore, the findings from healthy children are being used to assess the impact of traumatic brain injury in children using a translated version of JEF-C© in French (Jansari et al., in press). This work will be used to inform potential rehabilitation for such children, and to track the long-term impact of paediatric brain damage.

Conclusions

In summary, JEF© is a new paradigm for the assessment of executive functions. Its sensitivity has been demonstrated on a group of individuals with ABI who perform in the unimpaired range on a number of standard clinical tests. Further research has shown that it can be used in a number of different areas other than simply brain damage. The assessment provides a wealth of data across a range of different cognitive constructs. Further research is needed to validate JEF© against currently used tests. If this concurrent validity can be firmly established, and JEF© can make a unique contribution to capturing the variance in everyday executive abilities, then it should be considered as part of the arsenal of neuropsychological tests as well as a potential research tool for the future.

Acknowledgements

The authors would like to thank Rehab UK and their clients who participated in this study. We would also like to thank Jonathan Evans, Anna Adlam, Jenny Limond and Trudi Edginton for their comments on earlier versions of this manuscript.

References

Alderman, N., Burgess, P.W., Knight, C., & Herman, S. (2003). Ecological validity of a simplified version of the multiple errands shopping test. *Journal of the International Neuropsychological Society*, *9*(1), 31–44.

Baddeley, A., Della Sala, S., Papagno, C., & Spinnler, H. (1997). Dual-task performance in dysexecutive and nondysexecutive patients with a frontal lesion. *Neuropsychology*, *11*(2), 187–194.

Bennett, P.C., Ong, B., & Ponsford, J. (2005). Assessment of executive dysfunction following traumatic brain injury: Comparison of the BADS with other clinical neuropsychological measures. *Journal of the International Neuropsychological Society*, *11*, 606–613.

Bottari, C., Swaine, B., & Dutil, E. (2007). Interpreting activity of daily living errors for treatment and discharge planning: The perception of occupational therapists. *Journal of Head Trauma Rehabilitation*, *1*, 26–30.

Brooks, B.M., & Rose, F.D. (2003). The use of virtual reality in memory rehabilitation: current findings and future directions. *Neurorehabilitation*, *18*, 147–157.

Burgess, P., & Shallice, T. (1997). *The Hayling and Brixton tests. Test manual*. Bury St Edmunds, UK: Thames Valley Test Company.

Burgess, P.W., & Simons, J.S. (2005). Theories of frontal lobe executive function: clinical applications. In P.W. Halligan & D.T. Wade (Eds.), *Effectiveness of rehabilitation for cognitive deficits* (pp. 211–232). Oxford: Oxford University Press.

Burgess, P.W., Veitch, E., de Lacy Costello, A., & Shallice, T. (2000). The cognitive and neuroanatomical correlates of multitasking. *Neuropsychologia*, *38*, 848–863.

Carelli, L., Morganti, F., Weiss, P.L., Kizony, R., & Riva, G. (2008, August). A virtual reality paradigm for the assessment and rehabilitation of executive function deficits post stroke: Feasibility study. In *Virtual Rehabilitation*, 2008 (pp. 99–104). Piscataway, New Jersey: IEEE.

Chan, R.C.K., Shum, D., Touloupoulou, T., & Chen, E.Y.H. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology*, *23*, 201–216.

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review*, *13*, 181–196.

Chaytor, N., Schmitter-Edgecombe, M., & Burr, M. (2006). Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology*, *21*, 217–227.

Cherrier, M.M., Aubin, S., & Higan, C.S. (2009). Cognitive and mood changes in men undergoing intermittent combined androgen blockade for non-metastatic prostate cancer. *Psycho-Oncology*, *18*(30), 237–247.

Damasio, A.R. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *351*, 1413–1420.

Duncan, J. (1986). Disorganization of behavior after frontal lobe damage. *Cognitive Neuropsychology*, *3*, 271–290.

Elkind, J.S., Rubin, E., Rosenthal, S., Skoff, B., & Prather, P. (2001). A simulated reality scenario compared with the computerized Wisconsin Card Sorting

https://doi.org/10.1017/BrImp.2014.14 Published online by Cambridge University Press

Test: An analysis of preliminary results. *CyberPsychology & Behavior*, *4*, 489–496.

Ellis, J. (1996). Prospective memory or the realization of delayed intentions: A conceptual framework for research. In M. Brandimonte, G.O. Einstein & M.A. McDaniel (Eds.), *Prospective memory: Theory and applications* (pp. 1–22). Mahwah, NJ, US: Lawrence Erlbaum Associates.

Ellis, J., & Kvavilashvili, L. (2000). Prospective memory in 2000: Past, present and future directions. *Applied Cognitive Psychology Special Issue: New Perspectives in Prospective Memory*, *14*(7), 1–9.

Eslinger, P.J., & Damasio, A.R. (1985). Severe disturbance of higher cognition after bilateral frontal ablation: Patient EVR. *Neurology*, *35*(3), 1731–1741.

Gega, L., White, R., Clarke, T., Turner, R., & Fowler, D. (2013). Virtual environments using video capture for social phobia with psychosis. *Cyberpsychology, Behavior, and Social Networking*, *16*(6), 473–479.

Grafman, J., & Litvan, I. (1999). Importance of deficits in executive functions. *Lancet*, *354*, 1921–1923.

Greenwood, J.R., Barnes, B.P., McMillan, T.M., & Ward, C.D. (2003). *Handbook of neurological rehabilitation* (2nd ed.). East Sussex, England, Psychology Press.

Jansari, A., Debreceni, K., Bartfai, A., & Eriksson, M. (2008). Swedish JAAM? Adaptation of a virtual reality assessment of dysexecutive syndrome to a new culture and language. *Brain Impairment*, *9*(2), 220.

Jansari, A., Devlin, A., Kerrouche, B., Gilboa, Y., Davies, M., Fisher, H., . . . Chevignard, M. (in press). C'est ma fête! Can a French adaptation of a virtual reality assessment of children's executive functions (JEF-C©) work and be used with paediatric patients with acquired brain injury? *Journal of the International Neuropsychological Society*.

Jansari, A., Froggatt, D., Edginton, T., & Dawkins, L. (2013). Investigating the impact of nicotine on executive functions using a novel virtual reality assessment. *Addiction*, *108*, 977–984.

Jansari, A., Sosson, C., & Samson, D. (2013). Can a French adaptation of a virtual reality assessment of executive functions (JEF©) work? *Brain Impairment*, *14*(2).

Josman, N., Klinger, E., & Kizony, R. (2008). Performance within the virtual action planning supermarket (VAP-S): an executive function profile of three different populations suffering from deficits in the central nervous system. *Proceedings of the 7th International Conference on Disability, Virtual Reality and Associated Technologies*, Maia & Porto, Portugal (pp. 33–38).

Klinger, E., Chemin, I., Lebreton, S., & Marié, R.M. (2006). Virtual action planning in Parkinson's disease: A control study. *CyberPsychology & Behavior*, *9*(3), 342–347.

Knight, C., Alderman, N., & Burgess, P.W. (2002). Development of a simplified version of the multiple errands test for use in hospital settings. *Neuropsychological Rehabilitation*, *12*(3), 231–255.

Lezak, M.D., Howieson, D.B., & Loring, D.W. (Eds.) (2004). *Neuropsychological assessment* (4th ed.). New York: Oxford University Press.

Luria, A. (1973). *The working brain: An introduction to neuropsychology*. New York: Basic Books.

Luria, A.R. (1966). *Higher cortical functions in man*. New York: Basic Books.

Masson, J.D., Dagnan, D., & Evans, J. (2010). Adaptation and validation of the Tower of London test of planning and problem solving in people with intellectual disabilities. *Journal of Intellectual Disability Research*, *54*, 457–467.

Mateer, C.A. (1999). Executive function disorders: Rehabilitation challenges and strategies. *Seminars in Clinical Neuropsychiatry*, *4*, 50–59.

McGeorge, P., Phillips, L.H., Crawford, J.R., Garden, S.E., Della Sala, S., & Milne, A.B. (2001). Using virtual environments in the assessment of executive dysfunction. *Presence: Teleoperators & Virtual Environments*, *10*, 375–383.

Mills, R. (2009). Investigating higher level cognitive impairments as a consequence of Androgen Deprivation Therapy (ADT) in men with prostate cancer using a new virtual reality assessment. Unpublished Graduate Diploma dissertation, University of East London.

Montgomery, C., Ashmore, K., & Jansari, A. (2011). The effects of a modest dose of alcohol on executive functioning and prospective memory. *Human Psychopharmacology: Clinical & Experimental*, *26*, 208–215.

Montgomery, C., Hatton, N., Fisk, J., Ogden, R., & Jansari, A. (2010). Assessing the functional significance of ecstasy-related memory deficits using a virtual paradigm. *Human Psychopharmacology: Clinical and Experimental*, *25*, 318–325.

Montgomery, C., Seddon, A.L., Fisk, J.E., Murphy, P.N., & Jansari, A. (2012). Cannabis-related deficits in real-world memory. *Human Psychopharmacology: Clinical and Experimental*, *27*, 217–225.

Norman, D.A. & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. *Consciousness and Self-Regulation*, *4*, 1–18.

Pugnetti, L., Mendozzi, L., Attree, E.A., Barbieri, E., Brooks, B.M., Cazzullo, C.L., Motta, A., & Rose, F.D. (1998). Probing memory and executive functions with virtual reality: Past and present studies. *CyberPsychology & Behavior*, *1*, 151–161.

Rand, D., Basha-Abu Rukan, S., Weiss, P.L., & Katz, N. (2009). Validation of the Virtual MET as an assessment tool for executive functions. *Neuropsychological Rehabilitation*, *19*(4), 583–602.

Reger, G.M., Parsons, T.D., Gahm, G.A., & Rizzo, A.A. (2010). Virtual reality assessment of cognitive functions: A promising tool to improve ecological validity. *Brain Injury*, *7*, 24–26.

https://doi.org/10.1017/BrImp.2014.14 Published online by Cambridge University Press

Renison, B., Ponsford, J., Testa, R., & Jansari, A. (2008). The use of virtual reality in assessment of executive function following traumatic brain injury. *Brain Impairment*, *9*(1), 47.

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. *Journal of the International Neuropsychological Society*, *18*, 440–450.

Rizzo, A., & Kim, G.J. (2005). A SWOT analysis of the field of virtual reality rehabilitation and therapy. *Presence: Teleoperators & Virtual Environments*, *14*(2), 119–146.

Shallice, T., & Burgess, P. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, *114*, 727–741.

Stuss, D.T., & Alexander, M.P. (2000). Executive functions and the frontal lobes: a conceptual view. *Psychological Research*, *63*, 289–298.

Stuss, D.T., & Benson, D.F. (1984). Neuropsychological studies of the frontal lobes. *Psychological Bulletin*, *95*, 3–28.

Titov, N., & Knight, R.G. (2005). A computer-based procedure for assessing functional cognitive skills in patients with neurological injuries: The virtual street. *Brain Injury*, *19*(5), 315–322.

Wilson, B.A., Alderman, N., Burgess, P.W., Emslie, H., & Evans, J.J. (1996). *Behavioural Assessment of the Dysexecutive Syndrome (BADS)*. Bury St. Edmunds, UK, Thames Valley Test Company.

Wilson, B.A., Evans, J.J., Emslie, H., Alderman, N., & Burgess, P. (1998). The development of an ecologically-valid test for assessing patients with a dysexecutive syndrome. *Neuropsychological Rehabilitation*, *8*, 213–228.

Zalla, T., Plassiart, C., Pillion, B., Grafman, J., & Sirigu, A. (2001). Action planning in a virtual context after prefrontal cortex damage. *Neuropsychologia*, *39*(8), 759–770.

https://doi.org/10.1017/BrImp.2014.14 Published online by Cambridge University Press

https://doi.org/10.1017/BrImp.2014.14 Published online by Cambridge University Press

APPENDIX A

Review of studies using virtual reality for assessing executive functions (in chronological order of publication)

Authors	Participants	Environment/task	Measures	Executive functions
Zalla, Plassiart, Pillon, Grafman and Sirigu (2001)	ABI	Home/morning routine	Action sequences and mistakes	Action planning
Elkind, Rubin, Rosenthal, Skoff and Prather (2001)	Healthy normal controls (age range 18–74)	Beach/sorting beach apparel and refreshments	WCST-type categorisation and errors	Unclear
McGeorge et al. (2001)	ABI	Office	Time taken, quality of action plan	Planning
Titov and Knight (2005)	ABI	Street	Ability to complete delayed intentions	Prospective remembering
Klinger, Chemin, Lebreton and Marié (2006)	Parkinson's disease	Supermarket	Number of errors, time selecting items	Shifting attention, planning
Carelli, Morganti, Weiss, Kizony and Riva (2008)	Healthy elderly	Supermarket	Number of errors, time selecting items	Shifting attention, planning
Josman, Klinger and Kizony (2008)	Stroke, mild cognitive impairment, schizophrenia	Supermarket	Time taken, distance travelled, number of actions	Strategy
Rand, Bacha-Abu Rukan, Weiss and Katz (2009)	Traumatic brain injury (stroke)	Supermarket	Number of mistakes, rule breaks, non-efficiency mistakes	Multi-tasking, strategy-formation, planning, selection
Reger, Parsons, Gahm and Rizzo (2010)	Military personnel	Middle Eastern city	Serial Attention Test, Stroop time and errors	Attention

https://doi.org/10.1017/BrImp.2014.14 Published online by Cambridge University Press

APPENDIX B

List of cognitive constructs, their operational definitions and an example of a task for each construct

Construct	Definition	Example of task
Planning (PL)	Ordering events/objects due to logic (not importance)	Rearrange manager's tasks-for-completion into logical order
Prioritisation (PR)	Ordering events due to perceived importance	Arrange five agenda topics for discussion at the meeting into order of importance
Selective-thinking (ST)	Choosing between two or more alternatives by drawing on acquired knowledge	Choose which mail company should send each item of post based on each company's speciality
Creative-thinking (CT)	Looking for solutions to problems using unobvious and/or unspecified methods	Finding a method to cover graffiti that is written on a whiteboard in the meeting room and which cannot be erased because it has been written in indelible ink
Adaptive-thinking (AT)	Re-achieving goals in the face of changing conditions of success	The overhead projector (OHP) needed for use in the meeting is broken and needs to be replaced
Multi-tasking (MT)	Maintaining progress on two or more tasks at the same time	While engaged in one task, an urgent memo that requires action arrives and the participant has to find a way to complete both tasks successfully
Time-based prospective memory (TPM)	Remembering to execute a task at a pre-determined future point in time	Turn on the OHP 10 min before the scheduled start of the meeting
Event-based prospective memory (EPM)	Remembering to execute a task cued by an external stimulus/event	Note down the times of fire alarms that are being tested before the meeting starts
Action-based prospective memory (APM)	Remembering to execute a task cued by a stimulus related to an action the individual is already engaged in	Make a note of any equipment that breaks or malfunctions during the day