Journal of the International Neuropsychological Society (2012), **18**, 440–450.
Copyright © INS. Published by Cambridge University Press, 2012.
doi:10.1017/S1355617711001883

The Ecological and Construct Validity of a Newly Developed Measure of Executive Function: The Virtual Library Task

Belinda Renison,[1] Jennie Ponsford,[1,2] Renee Testa,[1,3] Barry Richardson,[4] AND Kylie Brownfield[1]

[1]Monash University, School of Psychology and Psychiatry, Melbourne, Australia
[2]Monash-Epworth Rehabilitation Research Centre (MERRC), Melbourne, Australia
[3]Melbourne Neuropsychiatry Centre, The University of Melbourne, Melbourne, Australia
[4]Monash University, School of Humanities, Communications and Social Sciences, Melbourne, Australia

(RECEIVED January 25, 2011; FINAL REVISION December 18, 2011; ACCEPTED December 19, 2011)

Abstract

Virtual reality (VR) assessment paradigms have the potential to address the limited ecological validity of pen and paper measures of executive function (EF) and the pragmatic and reliability issues associated with functional measures. To investigate the ecological validity and construct validity of a newly developed VR measure of EF, the Virtual Library Task (VLT); a real life analogous task—the Real Library Task (RLT); and five neuropsychological measures of EF were administered to 30 patients with traumatic brain injury (TBI) and 30 healthy Controls. Significant others for each participant also completed the Dysexecutive Questionnaire (DEX), which is a behavioral rating scale of everyday EF. Performances on the VLT and the RLT were significantly positively correlated indicating that VR performance is similar to real world performance. The TBI group performed significantly worse than the Control group on the VLT and the Modified Six Elements Test (MSET) but the other four neuropsychological measures of EF failed to differentiate the groups. Both the MSET and the VLT significantly predicted everyday EF suggesting that they are both ecologically valid tools for the assessment of EF. The VLT has the advantage over the MSET of providing objective measurement of individual components of EF. (*JINS*, 2012, *18*, 440–450)

Keywords: Executive function, Brain injuries, Neuropsychological tests, Validation studies, Cognitive function, Virtual systems

INTRODUCTION

Executive functions (EFs) enable individuals to plan, initiate, execute and monitor goal directed behavior in novel situations. Theoretical and factor analytic studies have identified several component skills which underpin goal-directed or executive behavior. These include planning, which involves analyzing a task and generating strategies to implement it, monitoring or regulating its execution and holding relevant information in mind to guide future behavioral responses, dealing with interference and multiple demands by switching flexibly between aspects of the task, inhibiting responses to inappropriate stimuli and checking that the task is executed as planned (Bennett, Ong, & Ponsford, 2005a; Bennett, Ong, & Ponsford, 2005b; Busch, McBride, Curtiss, & Vanderploeg, 2005; Chan, Shum, Toulopoulou, & Chen, 2008; Crawford & Henry, 2005; Cushman & Duffy, 2008; Damasio & Anderson, 2003; Duncan, Johnson, Swales, & Freer, 1997; Lezak, Howieson, & Loring, 2004; Stuss, 2007; Testa & Bennett, 2011). EF processes are commonly impaired in people with traumatic brain injury (TBI; Busch et al., 2005) resulting in difficulty performing daily activities and roles.

Neuropsychologists are frequently asked to predict patients' everyday functional abilities, necessitating that neuropsychological assessments be sensitive and ecologically valid. However, numerous studies have demonstrated that people expected to perform poorly on executive tests may perform within normal limits (Alderman, Burgess, Knight, & Henman, 2003; Norris & Tate, 2000; Ord, Greve, Bianchini, & Aguerrevere, 2009), and several studies have reported either no relationship or weak to moderate relationships between performance on EF tests and measures of everyday functioning (Chaytor et al., 2006; Manchester, Priestley, & Jackson, 2004). Such findings may reflect the low verisimilitude of neuropsychological tests;

Correspondence and reprint requests to: Jennie Ponsford, School of Psychology and Psychiatry, Monash University, Clayton, 3800, Victoria, Australia. E-mail: jennie.ponsford@monash.edu

https://doi.org/10.1017/S1355617711001883 Published online by Cambridge University Press

that is, the low parity between the task demands and the demands of the everyday environment that are complex and multi-factorial (Manchester et al., 2004; Morganti, 2004; Shallice & Burgess, 1991).

There is increasing recognition that neuropsychological assessment tools need to incorporate more complex and life-like scenarios, capable of taxing multiple executive processes simultaneously to be more predictive of real-world performances (Burgess et al., 2006; Chan, Shum, Toulopoulou, & Chen, 2008; Manchester et al., 2004). The Multiple Errands Test (MET; Alderman et al., 2003; Burgess et al., 2006; Shallice & Burgess, 1991) addresses this need by requiring participants to perform errands in a real shopping centre according to certain rules. However, functional assessments such as the MET can be time consuming, costly, and difficult to replicate or standardize across settings (Rand, Rukan, Weiss, & Katz, 2009). It is also not always feasible for people with significant mobility, behavioral, or psychological difficulties to access the community (Knight & Alderman, 2002).

Virtual reality (VR) technology incorporates the principles of verisimilitude by creating three-dimensional, interactive computer-generated environments that simulate real world scenarios (Matheis, 2004). Virtual reality assessments offer the potential to address the ecological validity limitations inherent in traditional neuropsychological measures, as well as the reliability and utility issues associated with functional neuropsychological assessments such as the MET (Shallice & Burgess, 1991). A range of VR environments have been used to assess executive functioning, including a 3D virtual apartment (Zalla, Plassiart, Pillon, Grafman, & Sirigu, 2001), virtual supermarket (Klinger, Chemin, Lebreton, & Marie, 2006), virtual university (McGeorge et al., 2001), virtual office (Jansari, Agnew, Akesson, & Murphy, 2004), virtual beach (Elkind, Rubin, Rosenthal, Skoff, & Prather, 2001), and virtual street (Titov & Knight, 2005). Recently a VR version of the MET has been developed, but this has only been studied in nine post-stroke patients (Rand et al., 2009). Research on these tools has been largely exploratory and has lacked psychometric rigor (Knight & Titov, 2009). Validation of the relationship between VR and real life environments has also yet to be established (Lamberts, Evans, & Spikman, 2010).

The Virtual Library Task (VLT; Renison, Ponsford, Testa, Richardson, & Brownfield, 2008) is a VR measure of EF designed within a function-based framework. Assessment involves observing participants perform several library-based tasks in a virtual library (Spooner & Pachana, 2006). Participants are required to prioritize and complete multiple tasks while managing interruptions, in addition to the presentation of new information that necessitates a shift in their behavioral approach. The present study aimed to examine the construct validity and ecological validity of the VLT. Specifically it aimed to:

(1) Examine the inter-rater and intra-rater reliability of the VLT and the same task performed in the real world environment; the Real Library Task (RLT).
(2) Compare performance on the VLT with performance on the RLT. It was hypothesized that scores on the VLT and RLT would be highly and significantly correlated.
(3) Examine whether the VLT and five other neuropsychological measures of EF were able to discriminate individuals with TBI with reported executive difficulties from healthy Controls. It was expected that the TBI participants would perform more poorly than Controls on the VLT and some measures of EF, but that group differences would be greater on the VLT task than on the neuropsychological measures.
(4) Examine the convergent validity of the VLT by examining correlations between the VLT and hypothetically related constructs. It was expected that the VLT would be moderately correlated with EF measures.
(5) Investigate the discriminant validity of the VLT by comparing correlations between the VLT and hypothetically related constructs (EF measures) with correlations between the VLT and a hypothetically unrelated construct, namely a measure of immediate attention, It was hypothesized that the VLT would be moderately correlated with EF measures, and not with immediate attention, and that the correlations of the VLT with EF measures would differ significantly from its correlation with immediate attention.
(6) Compare the ability of the VLT and neuropsychological measures of EF to predict executive functioning in everyday life. It was hypothesized that the VLT would be the strongest predictor of everyday EF.

METHOD

The study was approved by the Human Research Ethics committees of Epworth Hospital and Monash University.

Participants

Thirty participants with TBI and executive difficulties were recruited from Epworth Hospital ($n = 25$) and Osborn Sloan and Associates, a community based brain injury rehabilitation service ($n = 5$). 86.67% of TBI participants sustained injuries in motor vehicle accidents, 10% in falls and 3.33% in assaults. Duration of post traumatic amnesia (PTA), measured prospectively using the Westmead PTA Scale (Shores, Marosszeky, Sandanam, & Batchelor, 1986), ranged from 1 to 120 days (Median = 34.00; 6.67% PTA 0–7 days, 33.33% PTA 7–28 days; 60% PTA >28 days) and Glasgow Coma Scale (GCS) scores ranged from 3 to 15 (Median = 7.00; 63.33% GCS 3–8, 26.67% GCS 9–12, 10% GCS 13–15). PTA durations for the three TBI participants with GCS score of 13–15 were 14, 14 and 35 days, respectively. Time since injury ranged from 1 to 28 years (median = 5.00).

Participants were recruited if they, their families, or referring clinician considered they were experiencing EF problems, as per the definition of EF provided previously. All participants spoke English and were able to give informed consent. Exclusion criteria included a known psychiatric diagnosis, substance

https://doi.org/10.1017/S1355617711001883 Published online by Cambridge University Press

abuse, dysphasia, and physical or cognitive inability to use a computer.

A comparison group of 30 healthy adults with no history of neurological impairment was recruited from the general community.

Measures

Virtual Library Task (Renison et al., 2008)

The Virtual Library Task (VLT) is a non-immersive VR role-play task that can operate on any modern computer. It was built using a specially adapted version of the Genesis3D software program. It aims to measure EF in a reliable, functional, economical, and ecologically valid manner. The virtual environment (VE) models the exact dimensions and associated contents of two rooms in the Library at Epworth Hospital. This VE is navigated using an X-box and Playstation compatible handset. Participants were required to perform several specified tasks associated with the day to day running of the library, while adhering to predetermined rules. For example, to cool the library participants must "walk" to the air conditioner by manipulating the right hand joystick button. They turn the air conditioner on by moving the cursor over the air conditioner, *via* the left hand joystick button, and pressing the "select" button on the handset. They are informed *via* a visual prompt that the "air conditioner is out of order." To problem solve an alternative method to cool the room participants pick up the fan by moving the cursor over it and pressing the "select" button. They must then carry the fan to the meeting room and then move the cursor to the power point and press "select," which results in the fan being plugged in and automatically turned on. Another example is that of checking items that appear in the Intray, which participants do by using the right hand joystick control to "walk" within reach of the Intray and using the left hand joystick control to move the cursor over the Intray. The object is picked up by pressing the "select" button on the handset. To put the object down participants must "walk" to where they want to put the object and move the cursor to the desired location, that is, table in meeting room. Once they press the "select" button the object appears in the desired location. The VLT comprises functional tasks (Table 1) designed to reflect seven components of EF. These components were drawn from a survey of theoretical models and factor analyses of EF (Busch et al., 2005; Crawford & Henry, 2005; Damasio & Anderson, 2003; Duncan et al., 1997; Stuss, 2007; Testa & Bennett, 2011). Two prospective memory components were included as they were considered crucial in the coordination of complex behavior (Burgess, Alderman, Volle, Benoit, & Gilbert, 2009). Administration time typically ranges between nine and twenty minutes. Factors influencing the time taken include level of impulsivity and planning and problem solving abilities.

Trained raters used operationalized scoring criteria to rate how accurately the functional tasks were completed on a three-point scale (0, 1, 2). The functional tasks mapped on to seven components of EF, weighted proportionately according to the number of functional tasks that mapped onto them. Outcome measures included seven subtask scores ranging between 0 and 8 when weighted, summed to provide a Total Score from 0 to 56. Low scores reflected poor EF.

Real Library Task (Renison et al., 2008)

The Real Library Task (RLT) and scoring system was identical to the VLT, but was performed in the real library at the hospital and participants interacted with real objects.

Neuropsychological tests of intelligence, immediate memory, working memory, and verbal memory

Intelligence was estimated using the Wechsler Test of Adult Reading (WTAR). Verbal memory was assessed on the Logical Memory II subtest (scaled score) from the Wechsler Memory Scale – Third Edition (WMS-III). The Digit Span subtest from the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III) was used to examine immediate memory and working memory. *Z*-scores were used for the longest digits forwards (immediate memory) and longest digits backwards (working memory).

Neuropsychological tests of EF

Traditional neuropsychological measures included: Verbal Fluency total number of correct words generated over three trials (Benton, 1968), Wisconsin Card Sorting Test 64: Computer version 2 percentage of perseverative errors (Heaton, 2005), the Brixton Spatial Anticipation Test raw score (Burgess & Shallice, 1996), and two tasks proposed to have greater ecological validity, the Zoo Map and Modified Six Elements Test (MSET) from the Behavioural Assessment of the Dysexecutive Syndrome (Wilson, Alderman, Burgess, Emslie, & Evans, 1996). Raw scores were also used due to limited variability of generated scaled scores. These commonly used measures are said to assess executive constructs similar to those which the VLT was designed to measure.

Everyday measure of EF

The Dysexecutive Questionnaire (DEX; Wilson et al., 1996) is a 20-item checklist designed to measure the frequency of executive difficulties manifested in everyday life on a five-point scale ranging from "never" to "very often." Three executive factors are measured: cognitive, behavioral, and emotional. Higher scores on the DEX reflect poorer executive functioning. The independent-rater form was used as previous research has shown people with executive dysfunction may be poor informants of their own behavior due to reduced self-awareness (Bennett et al., 2005b; Burgess, Alderman, Evans, Emslie, & Wilson, 1998).

Experience with virtual reality technology

Participants self rated their level of experience in using virtual reality technology on a four point scale; 1 = never,

https://doi.org/10.1017/S1355617711001883 Published online by Cambridge University Press

Table 1. VLT: Subtasks and scoring criteria

VLT subtask	Definition	Subscore (range)	Functional task used to measure performance
Task Analysis	*The ability to consider the overall task requirements so that the most appropriate and successful behavioral approach to a task can be identified.	0–8	• Identifying the most logical and efficient order to perform the tasks on the "To Do List." • Identifying the most efficient order for the library books to be delivered to members' homes, while ensuring the task rules are adhered to.
Strategy Generation and Regulation	*The ability to generate and execute strategies to produce the optimal task response. This involves the ability to regulate performance and task demands and flexibly adapt or change strategies to ensure successful task completion.	0–8	• Generating an alternative solution to cooling the library when informed that the obvious solution (air conditioner) does not work. • Identifying how to photocopy a 3-page document when only 2 pieces of paper are available. • Generating an appropriate solution to ensuring 8 cups are placed on the table when only 7 cups are located on the cup tray. • Generating an appropriate solution to plugging in 3 appliances when only 2 power points are available.
Prospective Working Memory	*The ability to think ahead in the short term and hold possible responses online to enable selection of the most favorable behavioral response.	0–8	• Selecting the most appropriate catering menu given a set of required criteria. • Selecting the 5 most appropriate items for a library display given a set of required criteria.
Interference and Dual Task Management	*The ability to inhibit irrelevant or peripheral elements of an activity to ensure optimal task performance. Includes the ability to consider or perform more than one task simultaneously.	0–8	• Performing another task while waiting for the photocopier to warm up. • Ceasing a less important task when presented with an urgent task and returning to complete it later.
Response Inhibition	*The ability to stop or inhibit automatic responses to avoid inappropriate or unwanted responding.	0–8	• Adhering to task rule by inhibiting automatic response of answering the telephone. • Adhering to a task rule of not making personal phone calls when provided a phone messages to do so.
Time-based Prospective Memory	The ability to perform an action at a pre-determined specified time in the future.	0–8	• Turning the computer on at exactly 8:55 am. • Noting the non arrival of the food at 8:57 am.
Event-based Prospective Memory	The ability to perform an action in response to the occurrence of an external event.	0–8	• Crosses books off borrowing sheet when books are returned. • Places book on table when they are returned. • Documents the time a specified book is returned. • Checks the in-tray when walking past library desk.

2 = < weekly, 3 = < 1 hour per day, 4 = > 1 hour per day. Those who rated their level of experience as 1 or 2 were coded as having "low VR experience", whereas those who responded 3 or 4 were considered to have "high VR experience".

Procedures

Data collection occurred in the Library at Epworth Hospital over two 90-min sessions (including rest breaks) 1 week apart. Session one consisted of the VLT, Verbal Fluency, Zoo Map, and Brixton Spatial Anticipation Test. Session two consisted of the RLT, MSET, and Wisconsin Card Sorting Test. Administration order was counterbalanced to eliminate practice effects associated with performing the same task in the two different environments (virtual, real).

The VLT was run on a standard laptop and participants were trained in the navigation of the VE before its administration. Training time, ranging between three and fifteen minutes, was determined by the participant and was largely dependent on the participant's prior experience with VR software. Participants were provided a paper copy of the "Scenario Sheet" and "To Do List" and the researcher answered any questions before starting the task.

To obtain data regarding the intra- and inter-rater reliability of the VLT and RLT, the performance of 11 participants was videotaped and rated by two independent raters both at the time of task administration and 1 week later.

The DEX was completed and returned by mail by a "significant other" nominated by the participant, typically a family member or friend with whom they either lived or had at least weekly contact.

Data analyses were carried out using SPSS for Windows, version 18. Pearson's correlation coefficients were used to examine the intra- and inter-rater reliability of the VLT and RLT and to compare performance on the Library Task conducted in real and the VE. Independent samples *t* tests were

https://doi.org/10.1017/S1355617711001883 Published online by Cambridge University Press

used to compare the Control and TBI groups on the neuropsychological measures and VLT. Analysis of co-variance was conducted to examine the group difference in VLT scores after controlling for age, education, intelligence and verbal memory. Pearson's correlation coefficients were used to examine the relationship between (a) the VLT and executive measures, age, immediate attention, working memory and verbal memory, controlling for education and intelligence, and (b) the DEX and performances on the VLT and neuropsychological measures. One tailed tests and Fisher's transformations of r to Zr were used to test the significance of difference between two sets of correlations: (1) those between VLT and executive measures and (2) correlations between the VLT and a test of immediate attention. A standard multiple regression was performed to examine the relative contribution of executive measures, including the VLT to predict everyday EF as measured by the DEX. Alpha level was set at 0.05 throughout.

RESULTS

Intra- and Inter-rater Reliability of the VLT and RLT

Preliminary data investigating the reliability of the VLT and RLT, showed strong inter-rater reliability (rVLT = 1.0; $p < .001$; rRLT = 1.0; $p < .001$) and strong intra-rater reliability (rVLT = 1.0; $p < .001$; rRLT = 1.0; $p < .001$).

Correlation Between VLT and RLT Performance

The mean Total Scores on the VLT and RLT were 40.28 ($SD = 8.15$) and 41.35 ($SD = 8.48$), respectively. The examiner needed to assist a small number of TBI and Control participants who experienced some navigational difficulties; however the pattern of correlations of performances of the task in the two environments was similar for both groups. This suggests that the cognitive and physical impairments experienced by the TBI group did not impede their ability to use VR technology any more than Controls. As such, combined correlations for the two groups are presented. Performance in the real and VE was significantly correlated for the VLT and RLT Total Score ($r = .68$; $p < .01$), and for six of the seven subtests; Task Analysis ($r = .27$; $p = .04$), Strategy Generation and Regulation ($r = .77$; $p < .01$), Prospective Working Memory ($r = .53$; $p < .01$), Response Inhibition ($r = .54$; $p < .01$), Timed-based Prospective Memory ($r = .48$; $p < .01$), and Event-based Prospective Memory ($r = .73$; $p < .01$). Real and virtual performances were not significantly correlated for the Interference and Dual Task Management subtest ($r = -.10$; $p = .47$). Further investigation regarding this revealed that, whereas the groups did not significantly differ on the RLT Interference and Dual Task Management subtest, the Control group performed significantly better on this subtest than the TBI group on the VLT version (*M*TBI = 5.27; *SD*TBI = 1.86; *M*Control = 7.20; *SD*Control = 1.24; $t(58) = -4.74$; $p < .01$).

Comparison of TBI and Control Group Performance on Demographic Variables, Intelligence, Immediate Memory, Working Memory, and Verbal Memory Ability

The groups did not differ significantly in terms of age (TBI $M = 37.57$; $SD = 12.24$; Control $M = 32.10$; $SD = 12.34$; $t(58) = 1.72$; $p = .09$, $r = .22$), education (TBI $M = 13.33$; $SD = 2.58$; Control $M = 13.57$; $SD = 2{,}76$; $t(58) = -.34$; $p = .74$; $r = .04$), intelligence (TBI $M = 101.60$; $SD = 10.59$; Control $M = 104.73$; $SD = 7.68$; $t(52.94) = -1.31$; $p = .20$; $r = .13$), immediate memory (TBI $M = .01$; $SD = .90$; Control $M = .33$; $SD = 1.00$; $t(58) = -1.31$; $p = .19$; $r = .34$), working memory (TBI $M = .09$; $SD = 1.16$; Control $M = .53$; $SD = 1.00$; $t(58) = -1.57$; $p = .12$; $r = .41$), or verbal memory ability (TBI $M = 10.00$, $SD = 2.95$; Control $M = 10.67$; $SD = 2.32$; $t(58) = .97$; $p = .34$; $r = .13$).

Comparison of TBI and Control Group Performance on Neuropsychological Measures of EF

The performance of TBI and Control groups did not differ significantly on the EF tests, with the exception of the MSET, on which TBI participants performed more poorly than the control group.

Comparison of TBI and Control Group Performance on the Virtual Library Task of EF

The TBI group obtained significantly lower mean scores than the Control group on the VLT Total Score and on four of the VLT subtests, including VLT Prospective Working Memory, VLT Interference and Dual Task Management, VLT Time-based Prospective Memory and VLT Event-based Prospective Memory (Table 2). There were no significant group differences on the Task Analysis, Strategy Generation and Regulation, and Response Inhibition subtests of the VLT. Analysis of covariance (ANCOVA; Table 3) revealed that the TBI group obtained significantly lower scores on the VLT than the Control group after controlling for age, intelligence, working memory and verbal memory. Three of the four covariates made a significant contribution to the group differences in VLT scores including age, intelligence, and verbal memory performance.

Comparison of TBI and Control Group Performance on VR Experience

Chi square analyses revealed that the amount of participants' previous VR experience (low *vs*. high) did not differ between the groups, X2 (1, $N = 60$) = 1.67; $p = .20$). Independent samples t tests revealed participants with low VR experience did not perform significantly worse on the VLT than participants with high VR experience (*M*lowVR = 40.06; *SD*lowVR = 8.56; *M*highVR = 41.14; *SD*lowVR = 6.50; $t(58) = -.41$; $p = .69$).

https://doi.org/10.1017/S1355617711001883 Published online by Cambridge University Press

Table 2. Mean differences between groups on neuropsychological measures of EF and on the VLT ($N = 60$)

	TBI ($n = 30$)		Control ($n = 30$)					
Variable	*M*	(*SD*)	*M*	(*SD*)	*t*	*df*	*p*	*d*
Verbal Fluency	39.40	(12.68)	45.50	(10.97)	1.99	58	.05	.21
WCST-64	43.03	(8.51)	47.97	(10.83)	−1.95	57	.06	.25
Brixton Test	15.10	(6.98)	12.53	(5.54)	1.58	58	.12	.41
Zoo Map Test	3.60	(3.54)	4.00	(3.06)	−.51	58	.61	.07
Modified Six Elements	4.83	(1.34)	5.53	(0.86)	−2.41	49.41^	.02	.31
VLT Task Analysis	4.18	(1.88)	4.17	(1.97)	−1.07	58	.29	.00
VLT Strategy Gen. & Reg.	6.27	(1.51)	6.83	(1.32)	−1.55	58	.13	.20
VLT Prospective WM	5.60	(1.28)	6.37	(1.10)	−2.50	58	.02	.31
VLT IDTM	5.27	(1.86)	7.20	(1.24)	−4.74	50.66^	.00	.52
VLT Response Inhibition	6.27	(2.33)	6.27	(2.27)	.00	58	1.00	.00
VLT Time PM	4.93	(2.39)	6.33	(2.17)	−2.37	58	.02	.29
VLT Event PM	4.57	(2.39)	5.77	(1.92)	−2.14	58	.04	.27
VLT Total Score	37.08	(7.77)	43.48	(7.31)	−3.28	58	.00	.39

Notes. VLT Strategy Gen. & Reg. = VLT Strategy Generation & Regulation, VLT Prospective WM = VLT Prospective Working Memory, VLT IDTM = VLT Interference & Dual Task Management, VLT Time PM = VLT Time-based Prospective Memory, VLT Event PM = VLT Event-based Prospective Memory, ^This *df* value was adjusted to take account of a Levene's test showing violation of the homogeneity of variance assumption.

Convergent and Discriminant Validity

Convergent validity was examined by investigating the relationships between scores on VLT and scores on the neuropsychological measures. Support for the convergent validity of the VLT is provided by the moderate correlations between the VLT and three of the EF measures; Verbal Fluency, Zoo Map, and Modified Six Elements Test, controlling for education and IQ (Table 4).

Initial support for the discriminant validity of the VLT was provided by (1) the moderate correlations between the VLT and three of the EF tests, and (2) the non-significant correlation between the VLT and a test measuring immediate attention (Digits Forward). The correlation between the VLT and Digits Forward differed significantly from the correlation between the VLT and the Brixton Test ($Z = 2.17$; $p < .05$), and the correlation between the VLT and Digits Forward differed significantly from the correlation between the VLT and Zoo Map performance ($Z = 2.57$; $p < .01$). Moderate correlations were found between the VLT and age and measures of intelligence, working memory and verbal memory.

Table 3. Analysis of covariance for VLT Total Score

Source	*df*	MS	*F*	η^2	*p*
Age	1	171.27	7.62	.12	.01
Intelligence	1	100.86	4.49	.08	.03
Digits Backwards	1	14.98	.67	.01	.39
Logical Memory	1	128.75	5.73	.10	.03
Group	1	89.65	3.99	.07	.04
Error	54				

Note. $N = 60$.

Comparison of TBI and Control Group Everyday EF Performance as Measured by the Dysexecutive Questionnaire (DEX)

Independent sample *t* tests revealed that on average the TBI group experienced significantly more everyday executive dysfunction problems ($M = 30.00$; $SD = 14.15$) as measured by DEX independent rater responses, than the Control group ($M = 13.55$; $SD = 10.58$; $t(57) = 5.04$; $p < .01$; $r = .56$).

Relationships Between Scores on Neuropsychological Measures and the Virtual Library Task and Scores on the DEX

Pearson's correlation coefficients were calculated to examine the associations between scores on the EF measures and scores on the VLT. The pattern of correlations was similar for

Table 4. Correlations between virtual library task (VLT) and executive measures, age, immediate attention, working memory and verbal memory controlling for education and intelligence ($N = 60$)

	VLT	
	r	(*p*)
Verbal Fluency	.32	(.02)
WCST-64	.19	(.17)
Brixton Test	−.22	(.11)
Zoo Map	.29	(.03)
Modified Six Elements Test	.32	(.02)
Age	−.41	(.00)
Digits forward	.18	(.45)
Digits backwards	.11	(.45)
Logical Memory Test II	.33	(.01)

https://doi.org/10.1017/S1355617711001883 Published online by Cambridge University Press

Table 5. Pearson correlations between scores on neuropsychological measures of EF and the Virtual Library Task and scores on the DEX

	VF	WCS	Brix	Zoo Map	Mod SET	VLT Total	VLT TA	VLT SGR	VLT PWM	VLT IDTM	VLT Inhi	VLT TPM	VLT EPM
DEX	−.12	−.17	.23	.02	−.43**	−.38**	−.27*	−.19	−.07	−.45**	−.00	−.37*	−.24

Notes. VF = Verbal Fluency, WCS = Wisconsin Card Sorting Test computerized version, Brix = Brixton Spatial Anticipation Test, Mod SET = Modified Six Elements Test, VLT TA = VLT Task Analysis subtest, VLT SGR = VLT Strategy Generation and Regulation subtest, VLT PWM = VLT Prospective Working Memory subtest, VLT IDTM = VLT Interference and Dual Task Management subtest, VLT Inhi = VLT Inhibition subtest, VLT EPM = VLT Event-based Prospective Memory subtest, VLT TPM = VLT Time-based Prospective Memory subtest.
*$p < .05$, **$p < .01$.

each group; therefore, to maximize power combined correlations are presented in Table 5.

Significant correlations were found between DEX scores and only one of the neuropsychological measures, namely the MSET. Regarding the VLT, there were significant associations between the DEX score and the VLT Total Score and three subtests: Task Analysis, Interference and Dual Task Management, and Time-based Prospective Memory. The significant negative correlations indicate that poorer performance on the MSET and the VLT was indicative of a high number and/or severity of everyday executive problems being reported by the participants' significant other.

A standard multiple regression analysis was conducted to examine the relative contribution of all EF measures to scores on the DEX. According to Coakes and Steed (2003) at least five times more cases than independent variables are required to conduct multiple regression analysis. As such, only two predictor variables were included in the regression, on the basis that they were both designed with ecological validity in mind and that they correlated with the DEX: VLT Total Score and the MSET. Variables were entered simultaneously into the regression equation to ascertain the unique contribution of each to variance in DEX scores. A summary of the results of the regression analyses is presented in Table 6.

The total model was significant, $F(2,56) = 8.56$, $p < .01$, explaining 23.4% of the variance in DEX scores. At an individual level, both the MSET and the VLT were significant, with the Beta values indicating that the MSET was the stronger predictor of DEX scores.

DISCUSSION

The current study sought to examine whether performance on a VR task was similar to performance of the same task in the real world environment. As expected, scores on the VLT and the RLT were highly positively correlated. This finding is consistent with the results of previous research (Cushman & Duffy, 2008; Jansari et al., 2004; McGeorge et al., 2001; Rand et al., 2009; Zhang et al., 2003), suggesting that performance on the VLT is similar to performance on the RLT. This finding therefore supports the use of VE's for testing of patients with TBI. This is important because VR assessment has greater clinical utility than assessment in real world settings. The only exception to this was the lack of significant correlation between the real and virtual performances for the Interference and Dual Task Management subtest. The TBI group performed significantly more poorly than the Control group on the virtual version of this subtest but not on the real life version. The real life version may have involved more familiar navigation, whereas virtual navigation demands may have added to the complexity of this subtask, placing greater burden on the TBI participants. The verbal prompt provided on this version may have been more salient to TBI participants than the visual prompt provided in the VLT version. It is important to note that both versions of the task showed strong inter-rater reliability. In the present study VR experience did not impact on performance on the VLT.

Table 6. Summary of standard regression of Modified Six Elements test and Virtual Library Task (VLT) Total Score on Dysexecutive Questionnaire Score

Predictor variable	Raw correlation	B	Beta	*t*	*p*
Modified Six Elements test	−.43**	−4.20	−.34	−2.75	.01
VLT Total Score	−.38**	−.49	−.27	−2.19	.03

Note. $N = 60$, R Square = .25.

Support for the construct validity of the VLT as a measure of EF in participants with TBI was evidenced by the superior ability of the VLT to differentiate between patients with TBI and healthy Controls relative to the EF measures, even after controlling for age, education, intelligence, and verbal memory. Modest support for its convergent validity was also provided by the moderate correlations between the VLT and three of the five EF measures after controlling for education and IQ. These correlations, combined with the non significant correlation between the VLT and a measure of immediate attention and the significance of difference between pairs of correlations; namely the VLT and Digits Forward and (1) the VLT and the Brixton Test and (2) the VLT and Zoo Map performance, provide some support for the discriminant validity of the VLT. The moderate correlations between the VLT, and age, measures of intelligence, working memory and verbal memory supports previous studies which have reported that these constructs may influence performance on EF measures (Axelrod et al., 1996; Bennett et al., 2005a; Greve, Brooks, Crouch, Williams, & Rice, 1997). They underscore the importance of controlliong for these variables which have not been considered in many previous studies (Duncan, 2005).

As expected, and consistent with previously reported findings (Burgess et al., 1998; Chan & Manly, 2002;

https://doi.org/10.1017/S1355617711001883 Published online by Cambridge University Press

Dawson et al., 2009; Knight & Alderman, 2002; Wilson et al., 1996) the TBI group were reported by significant others to have significantly more executive difficulties in everyday life than the control group. Four of the neuropsychological measures did not identify this and failed to significantly differentiate between the two groups. This is consistent with several studies that have reported no significant differences between control and brain injured performances on Verbal Fluency (Alderman, et al., 2003; Jovanovski, 2004), Brixton Test of Spatial Anticipation (Draper & Ponsford, 2008), the WCST (Alderman et al., 2003; Dawson et al., 2009; Norris & Tate, 2000; Ord et al., 2009), and the Zoo Map Test (Evans, Chua, McKenna, & Wilson, 1997; Tyson, Laws, Flowers, Mortimer, & Schulz, 2008), although the Zoo Map Test has shown better discriminant validity in some previous research (Katz, Tadmor, Felzen, & Hartman-Maeir, 2007; Moriyama et al., 2002; Norris & Tate, 2000; Wilson, Evans, Emslie, Alderman, & Burgess, 1998). These findings add to a growing body of evidence regarding the limited ecological validity of traditional EF tests (Bennett et al., 2005b; Chan, 2001; Chaytor & Schmitter-Edgecombe, 2003; Chaytor et al., 2006; Manchester et al., 2004; Norris & Tate, 2000). The disparity between the demands of real-life and testing environments most likely accounts for this (Manchester et al., 2004; Morganti, 2004; Shallice & Burgess, 1991). That said, the multifactorial nature of the tasks, and the consequent possibility that the TBI patients did not have impairments on the components of EF measured by these tests cannot be ruled out.

The MSET did successfully differentiate the TBI and Control groups. Several studies have reported adequate construct validity of the MSET between healthy controls and clinical populations including alcoholics (Moriyama et al., 2002), TBI and multiple sclerosis (Norris & Tate, 2000), schizophrenia (Wilson et al., 1998), and acquired brain injury (Wilson et al., 1998). This task may be more sensitive to executive difficulties because it requires sustained cognitive effort over a 10-min period, incorporates subtasks simulating real world scenarios thereby providing good face validity, places significant demands on working memory, and requires an ability to multi-task and shift set.

The TBI group performed significantly worse on the VLT than the Control group overall, and on four out of the seven subtests; VLT Prospective Working Memory, VLT Interference and Dual Task Management, VLT Time-based Prospective Memory, and VLT Event-based Prospective Memory, suggesting that these tasks adequately taxed these executive sub-skills. The significant difference between the TBI and control groups in the Time and Event-based Prospective Memory subtests is consistent with the findings of previous studies using the Virtual Bungalow Task in patients with frontal lesions (Morris, Kotitsa, Bramham, Brooks, & Rose, 2002) and stroke (Brooks, Rose, Potter, Jayawardena, & Morling, 2004).

Both the MSET and the VLT were significantly and moderately correlated with the DEX, which supports the ecological validity of these tests. The tests together predicted 23.4% of the variance in DEX scores, providing support for a significant association between performance on these two tasks and everyday EF as measured on the DEX. Nevertheless there remains a substantial proportion of unshared variance in DEX scores, suggesting additional factors are contributing to the reported everyday executive difficulties. These may include: the level of environmental cognitive demand and compensatory strategy use (Chaytor et al., 2006), how often the person's executive function impairments negatively affect the significant other (Joyner, Silver, & Stavinoha, 2009), their degree of acceptance of cognitive and behavioral changes in the injured participant, their current stage of adjustment to the impact of the injury (Ponsford & Kinsella, 1991), emotional problems, pre-morbid functioning, health problems, and mobility status (Long & Collins, 1997). Future research examining ecological validity should attempt to control for the abovementioned sources of variation where possible.

Individually, the MSET made a significant and unique contribution to prediction of DEX scores. The failure of previous studies to find positive correlations between the MSET and overall DEX scores (Chan & Manly, 2002; Norris & Tate, 2000) may be due to the heterogeneity in their participants, who may not all have had executive dysfunction in the domains assessed by the tasks. However, the results of this study and that of Bennett et al. (2005b) who found moderately significant correlations ($r = -.36$) between TBI scores on the MSET and clinician-rated DEX suggests that this test is an ecologically valid measure in people with predominantly moderate to severe TBI.

The VLT also made a unique contribution to the prediction of DEX scores. The VLT appears to have greater face validity than the MSET. This is supported by the observations of Manly, Hawkins, Evans, Woldt, and Robertson (2002) that the MSET remains somewhat removed from everyday situations; "it is difficult to imagine many real situations where the parameters for switching from one task to another are so firmly set, and where the time allocated to each aspect is of such short duration" (2002, p. 279). Clinically the importance of face validity should not be underestimated; patients are more likely to accept feedback regarding cognitive impairments if the tools used to make decisions regarding their impairments are reflective of the everyday environments and scenarios. Another important advantage of the VLT over the MSET is that it attempts to provide operational definitions of the underlying components of EF which it purports to measure. Such information would be useful in assisting clinicians to tailor rehabilitation recommendations to a patient's specific EF impairments and strengths.

Several clinical implications arise from this study. We would caution against the use of Verbal Fluency, WCST, the Brixton Spatial Anticipation Test, and the Zoo Map test if a key function of the neuropsychological assessment is to extrapolate how cognitive behavioral deficits observed during assessment will impact on patients' functional executive abilities. We suggest that the VLT would also be an appropriate measure, potentially providing a more comprehensive assessment of EF than the MSET.

The present study is not without limitations. Given the moderate sample size it was only possible to examine the

https://doi.org/10.1017/S1355617711001883 Published online by Cambridge University Press

relative sensitivity of a limited number of executive tests. Thus one cannot rule out the possibility that other measures not included in this study may have shown greater sensitivity to everyday executive difficulties. However, a broad range of executive tasks was included. Development of an alternate version of the task for retest purposes will be a useful next step. The DEX was chosen as the ecological comparator for the neuropsychological measures and the VLT based on the recommendation of Chaytor and Schmitter-Edgecombe (2003) that neuropsychological measures of EF should be measured against a measure of everyday executive skill rather than many different global outcome measures. All but one participant lived in the community and, therefore, family- and friend-rated DEX scores were the most appropriate for this sample. However, the DEX is not without limitations and as previously discussed ratings made by relatives or friends may be influenced by a range of factors.

In addition to demonstrating that virtual reality assessments can be successfully administered to people with moderate to severe TBI, the current study provides evidence for the construct and ecological validity of the VLT; a newly developed assessment which aims to measure multiple components of EF in an integrated and lifelike manner. The inability of four of the five neuropsychological measures to (1) distinguish between the TBI and control groups and (2) correlate significantly with a measure of everyday EF, provides further evidence of the limited sensitivity and ecological validity of traditional pen and paper measures of EF. In contrast The MSET and the VLT appear to be sensitive and ecologically valid assessment tools. The VLT has the potential advantage over the MSET of providing objective measurement of various components of EF.

ACKNOWLEDGMENTS

We thank the participants in this study, who generously donated their time. We also thank Johnathan Wells for programming the virtual library environment. The development of the virtual library environment was supported by a grant from the Jack Brockhoff Foundation and completion of the study was supported by a grant from Monash University's Faculty of Medicine, Nursing and Health Sciences. There are no conflicts of interest relating to this manuscript.

REFERENCES

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

Axelrod, B.N., Goldman, R.S., Heaton, R.K., Curtiss, G., Thompson, L.L.., & Ghelune, G.J., et al. (1996). Discriminability of the Wisconsin Card Sorting Test using the standardization sample. *Journal of Clinical & Experimental Neuropsychology Society*, *18*, 338–342.

Bennett, P.C., Ong, B., & Ponsford, J. (2005a). Assessment of executive dysfunction following traumatic brain injury: Comparison of the BADS with other clinical neuropsychological measures. *Journal of the International Neuropsychological Society*, *11*(05), 606–613. doi:10.1017/S1355617705050721

Bennett, P.C., Ong, B., & Ponsford, J. (2005b). Measuring executive dysfunction in an acute rehabilitation setting: Using the dysexecutive questionnaire (DEX). *Journal of the International Neuropsychological Society*, *11*(4), 376–385.

Benton, A. (1968). Differential behavioural effects on frontal lobe disease. *Neuropsychologia*, *6*, 53–60.

Brooks, B.M., Rose, F.D., Potter, J., Jayawardena, S., & Morling, A. (2004). Assessing stroke patients' prospective memory using virtual reality. *Brain Injury*, *18*(4), 391–401.

Burgess, P.W., Alderman, N., Evans, J., Emslie, H., & Wilson, B.A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society*, *10*(4), 547–558.

Burgess, P.W., Alderman, N., Forbes, C., Costello, A., Coates, L.M., Dawson, D.R., … Channon, S. (2006). The case for the development and use of 'ecologically valid' measures of executive function in experimental and clincial neuropsychology. *Journal of the International Neuropsychological Society*, *12*(2), 194–209 .

Burgess, P.W., Alderman, N., Volle, E., Benoit, R.G., & Gilbert, S.J. (2009). Mesulam's frontal lobe mystery re-examined. *Resotrative Neurology and Neuroscience*, *27*(5), 493–506.

Burgess, P.W., & Shallice, T. (1996). *The Hayling and Brixton Tests: Test manual*. Edmunds, England: Themes Valley Test Company Limited.

Busch, R.M., McBride, A., Curtiss, G., & Vanderploeg, R.D. (2005). The components of executive functioning in traumatic brain injury. *Journal of Clinical & Experiemental Neuropsychology*, *27*, 1022–1032.

Chan, R.C. (2001). Dysexecutive symptoms among a non-clinical sample: A study with the use of the Dysexecutive Questionnaire. *British Journal of Psychology*, *92*(3), 551–565.

Chan, R., & Manly, T. (2002). The application of "dysexecutive syndrome" measures across cultures: Performance and checklist assessment in neurologically healthy and traumatically brain-injured Hong Kong Chinese volunteers. *Journal of the International Neuropsychological Society*, *8*(6), 771–780.

Chan, R., Shum, D., Toulopoulou, T., & Chen, E. (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*(4), 181–197.

Chaytor, N., Schmitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology*, *21*(3), 217–227.

Coakes, S.J., & Steed, L.G. (2003). *SPSS Analysis Without Anguish*. Queensland: John Wiley & Sons Australia.

Crawford, J.R., & Henry, J.D. (2005). *Assessment of executive dysfunction*. New York, NY: Oxford University Press.

Cushman, L.A., & Duffy, C.J. (2008). Detecting navigational deficits in cognitive aging and Alzheimer disease using virtual reality. *Neurology*, *71*, 888–895.

Damasio, A.R., & Anderson, S.W. (2003). The frontal lobes. In K.M. Heilman & E. Valenstein (Eds.), *Clinical Neuropsychology* (4th ed.). New York: Oxford University Press.

Dawson, D.R., Anderson, N.D., Burgess, P., Cooper, E., Krpan, K.M., & Stuss, D.T. (2009). Further development of the Multiple Errands Test: Standardized scoring, reliability, and ecological validity for the Baycrest version. *Archives of Physical Medicine and Rehabilitation*, *90*(11, Suppl. 1), S41–S51.

https://doi.org/10.1017/S1355617711001883 Published online by Cambridge University Press

Draper, K., & Ponsford, J. (2008). Cognitive functioning ten years following traumatic brain injury and rehabilitation. *Neuropsychology*, *22*(5), 618–625.

Duncan, J. (2005). Frontal lobe function and general intelligence: Why it matters. *Cortex*, *41*, 215–217.

Duncan, J., Johnson, R., Swales, M., & Freer, C. (1997). Frontal lobe deficits after head injury: Unity and diversity of function. *Cognitive Neuropsychology*, *14*(5), 713–741.

Elkind, J.S., Rubin, E., Rosenthal, S., Skoff, B., & Prather, P. (2001). A simulated reality scenario compared with the computerized Wisconsin Card Sorting Test: An analysis of Preliminary results. *CyberPsychology & Behavior*, *4*(4), 489–496.

Evans, J., Chua, S., McKenna, P., & Wilson, B. (1997). Assessment of the dysexecutive syndrome in schizophrenia. *Psychological Medicine*, *27*(3), 635–646.

Greve, K.W., Brooks, J., Crouch, J.A., Williams, M.C., & Rice, W.J. (1997). Factorial structure of the Wisconsin Card Sorting Test. *British Journal of Clinical Psychology*, *36*, 283–285.

Heaton, R.K. (2005). *Wisconsin Card Sorting Test-64: Computer version 2 – research edition (WCST-64CV2)*. Lutz, FL: Psychological Assessment Resources.

Jansari, A., Agnew, R., Akesson, K., & Murphy, L (2004). *Using virtual reality to create an ecologically valid measure of real world problems in patients with dysexecutive syndrome.* Paper presented at the Symposium on Neurological Rehabilitation: A Satellite Symposium to the Joint 27th Conference of the Australian Society for the Study of Brain Impairment and the 27th Mid-year Meeting of the International Neuropsychological Society Uluru, Australia.

Jovanovski, D. (2004). *Cognitive set shifting using functional magnetic reasonance imaging and virtual reality: A comparison between traditional and a novel ecologically valid exectuive function task*. Master of Arts, University of Toronto, Toronto.

Joyner, K.B., Silver, C.H., & Stavinoha, P.L. (2009). Relationship between parenting stress and ratings of executive functioning in children with ADHD. *Journal of Psychoeducational Assessment*, *27*(6), 452–464. doi:10.1177/0734282909333945

Katz, N., Tadmor, I., Felzen, B., & Hartman-Maeir, A. (2007). Validity of the Executive Function Performance Test in individuals with schizophrenia. *OTJR: Occupation, Participation and Health*, *27*(2), 44–51.

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

Knight, C., & Alderman, N. (2002). Development of a simplified version of the Multiple Errands Test for use in hospital settings. *Neuropsychological Rehabilitation*, *12*(3), 231–255.

Knight, R.G., & Titov, N. (2009). Use of virtual reality tasks to assess prospective memory: Applicability and evidence. *Brain Impairment*, *10*(1), 3–13.

Lamberts, K.F., Evans, J.J., & Spikman, J.M. (2010). A real-life, ecologically valid test of executive functioning: The executive secretarial task. *Journal of Clinical and Experimental Neuropsychology*, *32*(1), 56–65.

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

Long, C.J., & Collins, L.F. (1997). Ecological validity and forensic neuropsychological assessment. In R.J. McCaffrey, A.D. Williams, J.M. Fisher, & L.C. Laign (Eds.), *The practice of forensic neuropsychology: Meeting challenges in the courtroom*. New York: Plenum Press.

Manchester, D., Priestley, N., & Jackson, H. (2004). The assessment of executive functions: Coming out of the office. *Brain Injury*, *18*(11), 1067–1081.

Manly, T., Hawkins, K., Evans, J., Woldt, K., & Robertson, I.H. (2002). Rehabilitation of executive function: Facilitation of effective goal management on complex tasks using periodic auditory alerts. *Neuropsychologia*, *40*(3), 271–281.

Matheis, R.J. (2004). *Expanding the boundaries of neuropsychology: The application of VR for memory assessment*. Doctoral Thesis, Fairleigh Dickinson University, New Jersey.

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*, *10*(4), 375–383.

Morganti, F. (2004). Virtual interaction in cognitive neuropsychology. In G. Riva, C. Botella, P. Legeron, & G. Optale (Eds.), *Cybertherapy: Internet and virtual reality as assessment and rehabilitation tolls for clinical psychology and neuroscience*. Amsterdam: IOS Press.

Moriyama, Y., Mimura, M., Kato, M., Yoshino, A., Hara, T., Kashima, H., … Watanabe, A. (2002). Executive dysfunction and clinical outcome in chronic alcoholics. *Alcoholism: Clinical and Experimental Research*, *26*(8), 1239–1244 .

Morris, R.G., Kotitsa, M., Bramham, J., Brooks, B.M., & Rose, D. (2002). *Virtual reality investigation of strategy formation, rule breaking and prospective memory in patients with focal prefrontal neurosurgical lesions*. Paper presented at the International Conference of Disability, Virtual Reality & Associated Technology, Hungry.

Norris, G., & Tate, R.L. (2000). The Behavioural Assessment of the Dysexecutive Syndrome (BADS): Ecological, concurrent and construct validity. *Neuropsychological Rehabilitation*, *10*(1), 33–45.

Ord, J.S., Greve, K.W., Bianchini, K.J., & Aguerrevere, L.E. (2009). Executive dysfunction in traumatic brain injury: The effects of injury severity and effort on the Wisconsin Card Sorting Test. *Journal of Clinical and Experimental Neuropsychology*, *32*(2), 132–140.

Ponsford, J., & Kinsella, G. (1991). The use of a rating scale of attentional behaviour. *Neuropsychological Rehabilitation*, *1*(4), 241–257.

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.

Renison, B., Ponsford, J., Testa, R., Richardson, B., & Brownfield, K. (2008). *Virtual library task*. Melbourne: Monash University.

Shallice, T., & Burgess, P.W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, *114*(2), 727–741.

Shores, E.A., Marosszeky, J.E., Sandanam, J., & Batchelor, J. (1986). Preliminary validation of a clinical scale for measuring the duration of posttraumatic amnesia. *The Medical Journal of Australia*, *144*, 569–572.

Spooner, D.M., & Pachana, N.A. (2006). Ecological validity in neuropsychological assessment: A case for greater consideration in research with neurologically intact populations. *Archives of Clinical Neuropsychology*, *21*(4), 327–337.

Stuss, D.T. (2007). New approaches to prefrontal lobe testing. In B. L. Miller & J. Cummings (Eds.), *The human frontal lobes: Functions and disorders* (2nd ed., pp. 292–305). New York: Guilford Press.

https://doi.org/10.1017/S1355617711001883 Published online by Cambridge University Press

Testa, R., Bennett, P.C., & Ponsford, J. (2011). Factor Analysis of Nineteen Executive Function Tests in a Healthy Adult Population. *Manuscript accepted for publication to Archives of Clinical Neuropsychology. In Press*.

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.

Tyson, P.J., Laws, K.R., Flowers, K.A., Mortimer, A.M., & Schulz, J. (2008). Attention and executive function in people with schizophrenia: Relationship with social skills and quality of life. *International Journal of Psychiatry in Clinical Practice*, *12*(2), 112–119.

Wilson, B.A., Alderman, N., Burgess, P.W., Emslie, H., & Evans, J.J. (1996). *Behavioural assessment of the dysexecutive syndrome: Test manual*. England: Thames Valley Test Company.

Wilson, B.A., Evans, J.J., Emslie, H., Alderman, N., & Burgess, P.W. (1998). The development of an ecologically valid test for assessing patients with dysexecutive syndrome. *Neuropsychological Rehabilitation*, *8*(3), 213–228.

Zhang, L., Abreu, B.C., Seale, G.S., Masel, B., Christiansen, C.H., & Ottenbacher, K.J. (2003). A virtual reality environment for evaluation of a daily living skill in brain injury rehabilitation: Reliability and validity. *Archives of Physical Medicine and Rehabilitation*, *84*(8), 1118–1124.

https://doi.org/10.1017/S1355617711001883 Published online by Cambridge University Press