Simona Raspelli*
Federica Pallavicini

Applied Technology for
Neuropsychology Laboratory
Istituto Auxologico Italiano
Milan, Italy

Laura Carelli

Department of Human Sciences
University of Bergamo
Bergamo, Italy
and
Department of Neurology and
Laboratory of Neuroscience
"Dino Ferrari" Center
University of Milan
IRCCS Istituto Auxologico Italiano
Milan, Italy

Francesca Morganti

Department of Human Sciences
University of Bergamo
Bergamo, Italy

Elisa Pedroli
Pietro Cipresso

Applied Technology for
Neuropsychology Laboratory
Istituto Auxologico Italiano
Milan, Italy

Barbara Poletti
Barbara Corra
Davide Sangalli
Vincenzo Silani

Department of Neurology and
Laboratory of Neuroscience
"Dino Ferrari" Center
University of Milan
IRCCS Istituto Auxologico Italiano
Milan, Italy

Giuseppe Riva

Applied Technology for
Neuropsychology Laboratory
Istituto Auxologico Italiano
Milan, Italy
and
Department of Psychology
Catholic University of Milan
Milan, Italy

Validating the Neuro VR-Based Virtual Version of the Multiple Errands Test: Preliminary Results

Abstract

The purpose of this study was to establish ecological validity and initial construct validity of the virtual reality version of the Multiple Errands Test based on NeuroVR software as an assessment tool for executive functions. In particular, the Multiple Errands Test is an assessment of executive functions in daily life which consists of tasks that abide by certain rules and is performed in a shopping mall-like setting where there are items to be bought and information to be obtained. The study population included three groups: post-stroke participants ($n = 9$), healthy young participants ($n = 10$), and healthy older participants ($n = 10$). The general purpose of the study was investigated through the following specific objectives: (1) to examine the relationships between the performance of three groups of participants in the Virtual Multiple Errands Test (VMET) and in the traditional neuropsychological tests employed to assess executive functions; and (2) to compare the performance of post-stroke participants to those of healthy young and older controls in the Virtual Multiple Errands Test and in the traditional neuropsychological tests employed to assess executive functions. Correlations between Virtual Multiple Errands Test variables and some traditional executive functions measures provide preliminary support for the ecological and construct validity of the VMET; further performance obtained at the Virtual Multiple Errands Test provided a distinction between the clinical and healthy population, and between the two age control groups. These results suggest a possible future application of such an ecological approach for cognitive assessment and rehabilitation of stroke patients and elderly population with age-related cognitive decline.

1   Introduction

The term *executive functions* is an umbrella term comprising a wide range of cognitive processes and behavioral competencies which include verbal reasoning, problem solving, planning, sequencing, the ability to sustain attention, resistance to interference, utilization of feedback, multitasking, cognitive flexibility, and the ability to deal with novelty (Chan, Shum, Touloupoulou, & Chen, 2008).

Impairments of executive functions are common in neurological patients, in particular, those with frontal lobe damage consequent to traumatic brain injury and stroke (Baddeley & Wilson, 1988; Burgess, Veitch, Costello, & Shallice, 2000; Shallice & Burgess, 1991). Stroke is a common cause of death, and also

*Correspondence to s.raspelli@auxologico.it.

represents a very important cause of disability worldwide. It is well known that stroke affects both motor and cognitive aspects of human functioning, even if it is only in recent years that studies focused on the distribution of neuropsychological impairments in stroke patients, together with their role in predicting functional outcomes (Nys et al., 2007). Nys et al. highlighted a disorder in executive functioning, abstract reasoning, verbal memory, and/or language in 60–70% of the patients.

Individuals who suffer from executive function impairments present different problems, such as those of starting and stopping activities or the inability of mental and behavioral shifts; and demonstrate difficulties in activities of daily living (ADL) and instrumental activities of daily living (IADL; Shallice & Burgess, 1991; Chevignard et al., 2000; Fortin, Godbout, & Braun, 2003). In particular, McDowd, Filion, Pohl, Richards, and Stiers (2003) highlighted the role of attentional functioning, particularly divided attention and switching attention, in functional outcome and successful rehabilitation after stroke. Attentional abilities are not only a pure mental process, but play an important role in daily life functioning and in the rehabilitative process.

Studies performed in recent years revealed that deficits in executive function in the healthy elderly population appear to be associated with aging of the prefrontal cortex (Raz, 2000).

The assessment and rehabilitation of executive functions has been generally performed under typical clinical or laboratory settings, usually via pen and paper tasks rather than being presented in an actual or simulated manner. This aspect, together with the fact that opportunities for choices and decision-making are less available to patients within the clinical setting, makes these conditions unsatisfactory (Burgess et al., 2006; Lo Priore, Castelnuovo, & Liccione, 2003).

Indeed, the lack of ecological validity has been an important criticism for experimental tasks and traditional neuropsychological tests (Goldstein, 1996; Sbordone, 1996). Although many patients with frontal lobe lesions have been found to perform equally well, compared to controls, on traditional neuropsychological tests, they still experienced a lot of difficulty in everyday life activities (Shallice & Burgess, 1991). Conventional experimental tasks demand relatively simple responses to single events. On the contrary, more complex multi-step tasks in daily life may require a more complicated series of responses: goal setting and subgoals setting, prioritization of subgoals, triggering prospective memory to initiate subtasks when the conditions for them become ripe, and inhibition of irrelevant and inappropriate actions to different subtasks (Chan et al., 2008).

For these reasons, increasing the ecological validity of neuropsychological assessment is important, since this will increase the likelihood that a patient's cognitive and behavioral responses will replicate the response that would occur in real-life situations (Burgess et al., 2006). Regrettably, the assessment of executive function during typical daily life activities is still difficult (Rand, Rukan, Weiss, & Katz, 2009). Laboratory-based evaluations which simulate real-life tasks, such as the Behavioral Assessment of Dysexecutive Syndrome (BADS; Wilson, Alderman, Burgess, Emslie, & Evans, 1996) do exist. However, although the BADS has good validity (Wilson, Evans, Emslie, Alderman, & Burgess, 1998) and was recently found to predict function in a chronic schizophrenic sample (Katz, Tadmor, Felzen, & Hartman-Maeir, 2007), it still does not measure performance during real-life tasks, but rather during simulated life tasks. Functional instruments have been developed to assess executive function in real-life. One example is the Multiple Errands Test (MET; Alderman, Burgess, Knight, & Henman, 2003; Burgess et al.; Knight, Alderman, & Burgess, 2002; Shallice & Burgess, 1991), which is performed at a real shopping mall or in a hospital environment and involves the completion of various tasks, rules to adhere to, and a specified time frame. The assessment of executive functions in real-life settings has the advantage of giving a more accurate estimate of the patient's deficits than is possible within laboratory conditions (Burgess et al.). However, it is time-consuming and not always feasible in typical clinical settings (Chevignard et al., 2000; Fortin et al., 2003). Even the simplified versions of the MET, adapted especially to be performed in a hospital setting or in a nearby shopping mall, are time-consuming, primarily since patients must be taken to the setting where the assessment will be carried out and

should be able to walk independently in order to perform the assessment.

The application of virtual reality and the use of simulated environments, perceived by the user as comparable to real-world objects and situations, in the assessment and rehabilitation of executive functions may help to tackle the issues of ecological validity discussed earlier and overcome these limits (Chan et al., 2008; Rizzo & Kim, 2005).

Virtual reality's potential for rehabilitation assessment and intervention in general, and for cognitive rehabilitation specifically, is due to a number of unique attributes (Brooks & Rose, 2003; Riva et al., 2009; Rizzo & Kim, 2005): besides the opportunity for experiential, active learning which encourages and motivates the participant, the ability to objectively measure behavior in challenging but safe and ecologically-valid environments, while maintaining strict experimental control over stimulus delivery and measurement (Rizzo, Buckwalter, & Van der Zaag, 2002).

However, as underlined by Chan et al. (2008), this advanced technology should be carefully implemented in clinical setting, in particular with regard to those patients who are not familiar with computerized tests or who are anxious in undertaking a computerized test or in being tested in a semi-enclosed environment (Browndyke et al., 2002; Wiechmann & Ryan, 2003).

Rand, Rukan, Weiss, and Katz (2009) have developed a first version of the Virtual Multiple Errands Test (VMET) as an assessment tool for executive functions, within the virtual mall (Rand, Katz, Shahar, Kizony, & Weiss, 2005), a functional virtual environment currently consisting of a large supermarket which was programmed via GestureTek's IREX video capture virtual reality system. It was developed to provide post-stroke participants with the opportunity to engage in a complex, everyday task of shopping, in which their weak upper extremity and executive functions deficits can be trained. The VMall has been shown to provide an interesting and motivating task which can offer post-stroke participants an opportunity to practice different shopping tasks without leaving the treatment room (Rand, Katz, & Weiss, 2007); in this study, the VMall was found to be sensitive to differences in the shopping time and

number of mistakes between post-stroke participants, healthy young participants, and healthy older participants on a simple shopping task of four items (the Four-Item Test; Carelli et al., 2009; Raspelli et al., 2010).

The purpose of the current study was to establish ecological validity and initial construct validity of the VR version of the MET (Shallice & Burgess, 1991; Fortin et al., 2003), based on NeuroVR software as an assessment tool for executive functions. As opposed to GestureTek's IREX, which includes a camera which films the user and displays his or her image within the virtual environment and where the interaction is done using active movements, in NeuroVR, the participant is able to freely navigate in the various aisles with the aid of a joypad and to collect products (by pressing a button placed on the right side of the joypad), after having selected them with the viewfinder.

The study population included three groups: post-stroke participants, healthy young participants, and healthy older participants. The specific aims were to examine the relationships between the performance of three groups of participants in the VMET and at the traditional neuropsychological tests employed to assess executive functions and to compare the performance of post-stroke participants to those of healthy young and older controls in the VMET and in the traditional neuropsychological tests employed to assess executive functions.

2   Method

2.1 Participants

A total of 29 participants in three groups were included in the study, 9 post-stroke individuals and 20 healthy people in two age groups. The 9 stroke participants ranged in age from 50 to 70 years (mean age 62 years; *SD* 7.83). In addition, 20 healthy participants volunteered to participate in this study, including 10 young participants with an age range between 20 and 30 years (mean age 26 years; *SD* 1.95) and 10 older participants with an age range between 50 and 70 years (mean age 55 years; *SD* 6.03). All groups were fully independent in activities of daily living and instrumental activities of

Table 1. *Population Characteristics, Mean (SD)*

Group	Age (years)		Education (years)	
Patients ($n = 9$)	62	(7.83)	12.78	(3.56)
Adult control group ($n = 10$)	55	(6.03)	14	(2.1)
Young control group ($n = 10$)	26	(1.94)	17.1	(1.44)

daily living. Their demographic data are presented in Table 1.

Patients were excluded from the study who had a severe cognitive impairment (MMSE $\leq 18/30$) (M. F. Folstein, S. E. Folstein, & McHugh, 1975), a severe motor impairment which does not allow performance of the dual task procedure (Rand, Katz, Shahar, Kizony, & Weiss, 2005), auditory language comprehension difficulties (score at the ENB Token Test ≤ 26.5; Mondini, Mapelli, Vestri, & Bisiacchi, 2003), visual recognition impairments (score on Street's Completion Test $\leq 2.25/14$; Spinnler & Tognoni, 1987), and spatial hemi-inattention and neglect as assessed by the Star Cancellation Test within the Behavioural Inattention Test (BIT; Wilson, Cockburn, & Halligan, 1987). Patients also underwent an exhaustive traditional neuropsychological assessment.

Control subjects were excluded from the study who had a cognitive impairment (MMSE $\leq 24/30$) (Folstein et al., 1975), a motor impairment which does not allow performance of the dual task procedure (Rand, Katz et al., 2005), auditory language comprehension difficulties (score at the ENB Token Test ≤ 26.5; Mondini et al., 2003), visual recognition impairments (score on Street's Completion Test $\leq 2.25/14$; Spinnler & Tognoni, 1987), and spatial hemi-inattention and neglect as assessed by the Star Cancellation Test within the BIT (Wilson et al., 1987).

2.2 Instruments

2.2.1 The neuropsychological evaluation. A neuropsychological evaluation was conducted both for inclusion criteria and data collecting referred to patients' cognitive profile.

In particular, the following neuropsychological tests were employed: the Mini-Mental State Evaluation (M. F. Folstein et al., 1975), to assess the general cognitive level; the Star Cancellation Test within the BIT (Wilson et al., 1987), for visuo-spatial assessment; the Token Test within the Brief Neuropsychological Examination (ENB; Mondini et al., 2003), for auditory language comprehension difficulties; and Street's Completion Test, for object recognition and denomination (Spinnler & Tognoni, 1987).

As for the assessment of executive functions, the Test of Attentional Performance (TEA) (Zimmerman & Fimm, 1992) was employed with specific subtests to evaluate the ability to increase, waiting for a high priority stimulus, the level of attention and ability to keep it (a state of alertness); to pay attention to and to elaborate different information present at the same time (divided attention); to sustain selective attention (sustained attention); to control for input from different sensorial channel (intermodal comparison); to explore the visual field (visual exploration); to change the attentive focus (flexibility); to repress an inadequate reaction (go/no go); to focus attention, that is, the ability to reject irrelevant aspects of the stimulus (spatial incompatibility); to continuously control for the information flow through short term memory (working memory), and to shift the visual attentive focus without eye movements (attention shift).

In addition, for the assessment of executive functions, the Stroop Colour-Word Test (Stroop, 1935) was employed to evaluate frontal and inhibition abilities; the Iowa Gambling Task (Bechara et al., 1994), to evaluate the functional integrity of the orbito-frontal areas, through the simulation, in real time, of the personal ability in decision-making relative to the uncertainty of the premises and of their outcome, like a reward or a punishment; finally, the Dysexecutive Questionnaire (DEX; Wilson et al., 1998), to evaluate executive functions in everyday life.

The Activities of Daily Living (ADL) and Instrumental Activities of Daily Living (IADL) Tests (Katz, Ford, Moskowitz, Jackson, & Jaffe, 1963) were employed to assess activities of daily living, such as having a bath or getting dressed; and instrumental activities of daily living, such as using the telephone, doing some shopping,

Figure 1. *Two screen shots of the virtual supermarket.*

and keeping house, while the State and Trait Anxiety Index (STAI; Spielberger, Gorsuch, & Lushene, 1970) and the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) were used to evaluate the level of state and trait anxiety and of depression.

2.2.2 The Virtual Multiple Errands Test. The virtual environment employed in this study is a supermarket developed via NeuroVR[1] software and displayed on a desktop monitor. It consists of a Blender-based application[2] that enables active exploration of a virtual supermarket where users are requested to select and buy various products presented on shelves (see Figure 1). The user enters the supermarket and is presented with icons of the various items to be purchased.

With the aid of a joypad, the participant is able to freely navigate in the various aisles (using the up-down joypad arrows), and to collect products (by pressing a button placed on the right side of the joypad), after having selected them with the viewfinder. The virtual supermarket contains products grouped into the main grocery categories including beverages, fruits and vegetables, breakfast foods, hygiene products, frozen foods, garden products, and animal products. Signs at the top of each section indicate the product categories as an aid to navigation.

1. www.neurovr.org
2. www.blendernation.com

The original procedure of the Multiple Errands Test (MET; Shallice & Burgess, 1991) was modified to be adapted to the virtual scenario of the supermarket. It consists of some tasks (to buy some products from a shop and to obtain some information) that are performed in a mall-like setting or shopping center and abide by certain rules (e.g., to carry out all tasks but in any order; not to go into the same aisle more than once; not to buy more than two items per category of item).

In order to evaluate the VMET usability, a Likert scale ranging from 1 to 10 points was created to assess the dimensions of knowledge of technological means (computer; video games; joystick, and virtual reality); usability of the interface (difficulties during the experience; in using the joystick; in selecting products from aisle and in learning to move in the supermarket), and environmental content (difficulties in recognizing products on the aisle; evaluation of the products' organization in the supermarket; visibility of the signs at the top of each section).

2.3 Procedure

Participants were included in the study after a preliminary neuropsychological evaluation.

According to this preliminary evaluation and the exclusion criteria described above, participants who were deemed suitable for the study underwent a more exhaus-

tive neuropsychological assessment in order to obtain an accurate overview of their cognitive function. Moreover, participants were asked to complete the Virtual Multiple Errands Test (VMET) procedure after a training session. Indeed, a training period of about 15 min was first provided in a smaller version of the virtual supermarket environment in order to familiarize participants with both the navigation and shopping tasks.

Two sessions of about 60 min were scheduled for each patient; during the first session they underwent the complete neuropsychological assessment, while during the second session (held the following day) the VMET procedure within the virtual supermarket and TEA were administered. The order of administration of the two most time-consuming tests, VMET and TEA, was counterbalanced in the sample so that for half of the participants VMET was administered before TEA and for the other half TEA was administered before VMET.

2.3.1 Outcome Measures.
Neuropsychological tests scores were recorded and corrected for age, education level, and gender.

While completing the VMET procedure, the time of execution, total errors, partial task failures, inefficiencies, rule breaks, strategies, and interpretation failures were recorded. These are defined as follows.

- **Errors.** Errors are task failures that are omissions (defined as failing to attempt the task). For errors in executing the tasks, the scoring range was from 11 (the subject has correctly done the 11 tasks) to 33 (the subject has totally omitted the 11 tasks). In particular, the scoring scale for each task failure was from 1 to 3 (1 = the participant performed the task 100% correctly as indicated by the test; 2 = the participant performed aspects of the task, but the task was not completed 100% accurately; 3 = the participant totally omitted the task).
- **Inefficiencies.** Inefficiencies are defined as a failure to do more than one thing in one place when that is the only place to accomplish that task. Examples of the eight inefficiencies are not grouping tasks or not reading the instructions. For inefficiencies, the

scoring range was from 8 (great inefficiencies) to 32 (no inefficiencies). In particular, the scoring scale for each inefficiency was from 1 to 4 (1 = always; 2 = more than once; 3 = once; 4 = never).
- **Rule Breaks.** Rule breaks are defined as anything that violates the rules listed in the MET task list. Examples of the eight rule breaks are entering an area more than once or speaking to the examiner when not necessary. For rule breaks, the scoring range was from 8 (a large number of rule breaks) to 32 (no rule breaks). In particular, the scoring scale for each rule break was from 1 to 4 (1 = always; 2 = more than once; 3 = once; 4 = never).
- **Strategies.** Examples of the 13 strategies are planning before starting the tasks and marking off the tasks completed. The scoring range was from 13 (good strategies) to 52 (no strategies). In particular, the scoring scale for each strategy was from 1 to 4 (1 = always; 2 = more than once; 3 = once; 4 = never).
- **Interpretation Failures.** Interpretation failures offer insight into the type of errors and interpretation failures experienced by the subject in the testing situation. An example of the three interpretation failures is thinking that the tasks all had to be done in the order presented in the task list. The scoring range was from 3 (a large number of interpretation failures) to 6 (no interpretation failures). In particular, the scoring scale for each interpretation failure was from 1 to 2 (1 = yes; 2 = no).
- **Partial Task Failures.** For partial task failures, the scoring range was from 8 (no errors) to 16 (a large number of errors). As for partial task failures, the eight specific items, with a scoring range from 1 (yes) to 2 (no) for each one were: searched for item in the correct area; maintained task objective to completion; maintained sequence of the task; divided attention between components of task and components of other VMET tasks; organized materials appropriately throughout task; self-corrected upon errors made during the task; no evidence of perseveration; and sustained attention throughout the sequence of the task (not distracted by other stimuli).

3 Results

Data analysis was carried out using SPSS for Windows, version 17.0. Due to the small group sample size, nonparametric statistics were used. Specifically, Pearson's correlation coefficients were used to examine the relationships between the various scores of the neuropsychological tests employed to assess executive functions and the scores of the VMET for each group separately.

The comparison of the scores of the neuropsychological tests employed to assess executive functions and VMET between the post-stroke participants and both groups of healthy controls was performed using the Kruskal–Wallis procedure; in the case of significant differences, the Mann–Whitney procedure was used to determine the source of significance between each pair in the groups.

3.1 Descriptive Statistics

Table 2 shows the descriptive statistics with regard to the neuropsychological evaluation conducted for inclusion criteria for the three examined groups: patients and the two control groups (young and adult).

3.2 Correlations

With the aim of analyzing the relationships among the scores at VMET and at traditional tools for the measurement of executive functions within the patients group and the control groups, correlations analyses were carried out.

For patients, the following correlations emerged as significant between the VMET variables and the reaction times in some tests of the TEA.

- The time of execution of the VMET with the test for the state of alert with warning sign ($r = .762$, $p = .028$); the test of the intermodal comparison ($r = .81$, $p = .007$); and the test of audio divided attention ($r = .71$, $p = .047$).
- Total errors in VMET with the test of incompatibility (with time reaction as the measure; $r = -.75$, $p = .019$).

Table 2. *Descriptive Statistics for the Three Groups, Mean (SD)*

Group	MMSE	BIT	STREET	TOKEN	BDI	STAI TRAIT	STAI STATE	ADL	IADL
Patients ($n = 9$)	27.86 (1.62)	53.44 (1.13)	3.89 (2.54)	5 (.00)	10.71 (12.43)	47.5 (15.34)	39 (11.13)	6 (.00)	8 (.00)
Adult control group ($n = 10$)	28.92 (1.26)	54.00 (.00)	6.35 (2.5)	5 (.00)	6.2 (5.47)	40.3 (8.09)	36.4 (8.38)	6 (.00)	7.96 (.89)
Young control group ($n = 10$)	29.7 (.92)	54.00 (.00)	8.55 (2.95)	5 (.00)	6.5 (8.4)	42.9 (14.38)	37.4 (13.57)	6 (.00)	8 (.00)

- Inefficiencies in VMET with the test of attention shift with valid stimulus ($r = -.67$, $p = .045$) and without valid stimulus ($r = -.73$, $p = .026$).

For the control group made of adult participants, significant correlations emerged among some of the VMET variables; in particular, inefficiencies and rule breaks ($r = .76$, $p = .01$) and between the time of execution and total errors ($r = .678$, $p = .031$).

Finally, for the control group made of young participants, significant correlations emerged between interpretation failures and visual exploration with noncritical stimulus ($r = .63$, $p = .04$) and the time of execution ($r = -.71$, $p = .023$) and among some of the VMET variables and in particular between the time of execution and inefficiencies ($r = -.76$, $p = .013$), between inefficiencies and rule breaks ($r = .82$, $p = .004$), and interpretation errors ($r = .81$, $p = .005$).

3.3 Kruskal–Wallis and Mann–Whitney Tests

With the aim of analyzing the differences among the examined groups in VMET and in tests measuring executive functions, and in order to study the construct validity, the Kruskal–Wallis Test was first performed; finally, the Mann–Whitney Test for the direction of the results was performed.

Table 3 shows the emerged significant results.

4 Discussion

The construct validity of the VMET has been demonstrated by the significant correlations that emerged among the VMET and different tests employed for the measurement of executive functions within the groups of patients and healthy subjects. More specifically, significant correlations are those between the VMET and the scores of tests of TEA which measure executive functions within the different groups: intermodal comparison, spatial incompatibility, and divided attention. These tests are very sensitive to the correct ability to reject trivial aspects of stimuli (spatial incompatibility test) and to pay attention to and elaborate on different

information present at the same time (divided attention test); these are all basic components of executive functions.

A result opposite to expectations was the lack of significant correlations between variables of the VMET and the traditional tests measuring executive functions in the examined groups, the Stroop Test and the Iowa Gambling Task.

With regard to the lack of correlation with the Iowa Gambling Task, it is important to underline that different processes are involved into the two tests: while the VMET requires adherence to specific rules, the Iowa Gambling Task, which represents a test of affective decision-making, simulates in real time the ability to make decisions with uncertainty of premises and results, and lack of rules.

The lack of correlation between the VMET and the Stroop Test, but also between the VMET and the Test of Flexibility (TEA) in the different experimental groups could be explained considering that in the VMET procedure, there are nonconflicting instructions. In the future, the possibility to introduce supermarket announcements can be considered, in order to introduce information conflicting with the main task.

Finally, significant differences emerged among the three groups on some measures of the VMET and the other tests traditionally employed for the assessment of executive functions. As expected, patients made the greater number of errors, followed by adult and young control subjects (Raz, 2000). Moreover, patients employed the longest time for executing the VMET. This result could be interpreted considering the Likert scores for the assessment of the usability of the VMET. Indeed, contrary to patients and to adult control subjects, the 42% of the young control subjects had previously used virtual reality systems.

Moreover, 33% of patients had no previous knowledge of computers, while 28% of adult control subjects showed knowledge between 7 and 8 on the Likert score. Among those patients who had previously used a computer, the level of knowledge was low (less than 5).

Finally, patients showed the slowest reaction time in some tests of TEA battery, followed by the control groups, specifically in working memory, intermodal

Table 3. *Kruskal–Wallis and Mann–Whitney Tests**

	Patients	Adult control group	Young control group	Kruskal–Wallis H χ^2 (p)		Mann–Whitney
Time VMET	1102 (920)	645.00 (189)	433.7 (102.06)	13.6	.001	Patients < adults < young
Errors VMET	17.67 (4.09)	13.8 (1.81)	13.6 (1.26)	9.72	.008	Patients < adults < young
Mean intermodal comparison TEA	511.67 (75.72)	427.8 (93.53)	421.5 (76.42)	7.6	.022	Patients < adults < young
Mean flexibility TEA	1401.44 (574.87)	803.8 (150.25)	660.05 (143.25)	13.7	.001	Patients < adults < young
Mean spatial incompatibility TEA	556 (81.48)	533.4 (87.12)	437.1 (56.63)	10.3	.006	Patients < adults < young
Mean attention shift with nonvalid stimulus TEA	534.56 (225.22)	421.5 (85.29)	302.5 (37.34)	12.09	.002	Patients < adults < young
Mean attention shift with valid stimulus TEA	424.56 (110.53)	342 (49.22)	277.5 (34.47)	13.45	.001	Patients < adults < young
Mean working memory TEA	910.11 (338.93)	596.6 (143.66)	566.8 (109.01)	7.65	.022	Patients < adults < young

*For the patients, the adult control group, and the young control group, values given are mean (*SD*).

comparison, flexibility, spatial incompatibility, attention shift, and working memory tests (see Table 3).

5 Conclusions

Given the above mentioned results, it is possible to conclude that the VMET appeared sensible both to brain damage and aging, as already observed in previous studies with the original version of the MET.

The results of the present study may have important implications for rehabilitation of patients with cerebral lesions following stroke with regard to the possible ecological usefulness of the VR-based tools. Indeed, the literature shows many studies on damage in executive functions following stroke, employing functional magnetic resonance (fMRI), magnetic resonance (MRI), and clinical and functional measurements. However, little attention has been focused on ecological issues in the rehabilitation processes of stroke patients.

However, due to the small sample size, the obtained results should be considered preliminary: as an assessment tool, the VMET should be analyzed with regard to its temporal stability, namely test–retest reliability and criterion validity, before being ready for clinical application and for being applied to different clinical populations.

To conclude, this study provides preliminary data supporting the ecological and construct validity of the VMET as an assessment tool of executive functions and its role in differentiating between stroke patients and controls and between different age groups, with regard to the healthy population.

Acknowledgments

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*, 31–44.

Baddeley, A. D., & Wilson, B. (1988). Frontal amnesia and the dysexecutive syndrome. *Brain and Cognition, 7*, 212–230.

Bechara, A., Damásio, A. R., Damásio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50*(1–3), 7–15.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561–571.

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

Browndyke, J. N., Albert, A. L., Malone, W., Schartz, P., Paul, R. H., Cohen, R. A., et al. (2002). Computer-related anxiety: Examining the impact of technology-specific affect on the performance of a computerized neuropsychological assessment measure. *Applied Neuropsychology, 9*(4), 210–218.

Burgess, P. W., Alderman, N., Forbes, C., Costello, A., Coates, L. M. A., Dawson, D. R., Anderson, N. D., Gilbert, S. J., Dumontheil, I., & Channon, S. (2006). The case for the development and use of "ecologically valid" measures of executive function in experimental and clinical neuropsychology. *Journal of the International Neuropsychological Society, 12*, 194–209.

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

Carelli, L., Morganti, F., Poletti, B., Corrà, B., Weiss, P. L., Kizony, R., Silani, V., & Riva, G. (2009). A NeuroVR based tool for cognitive assessment and rehabilitation of post-stroke patients: Two case studies. *Studies in Health Technology Informatics, 144*, 243–247.

Chan, R. C. K., Shum, D., Toulopoulou, 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.

Chevignard, M., Pillon, B., Pradat-Diehl, P., Taillefer, C., Rousseau, S., Le Bras, C., & Dubois, B. (2000). An ecological approach to planning dysfunction: Script execution. *Cortex, 36*, 649–669.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the mental state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189–198.

Fortin, S., Godbout, L., & Braun, C. M. J. (2003). Cognitive structure of executive deficits in frontal lesioned head trauma patients performing activities of daily living. *Cortex, 39*(2), 273–291.

Goldstein, G. (1996). Functional considerations in neuropsychology. In R. J. Sbordone & C. J. Long (Eds.), *Ecological validity of neuropsychological testing* (pp. 75–89). Delray Beach, FL: GR Press/St. Lucie Press.

Katz, S., Ford, A. B., Moskowitz, R. W., Jackson, B. A., & Jaffe, M. W. (1963). Studies of illness in the aged. The index of ADL: A standardized measure of biological and psychosocial function. *JAMA, 185*, 914–919.

Katz, N., Tadmor, I., Felzen, B., & Hartman-Maeir, A. (2007). The Behavioral Assessment of the Dysexecutive Syndrome (BADS) in schizophrenia and its contribution to functional outcome. *Neuropsychological Rehabilitation, 17*, 192–205.

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*, 231–255.

Lo Priore, C., Casterlnuovo, G., Liccione, D., & Liccione, D. (2003). Experience with VSTORE: Considerations on presence in virtual environments for effective neuropsychological rehabilitation of executive functions. *CyberPsychology and Behavior, 6*(3), 281–287.

McDowd, J. M., Filion, D. L., Pohl, P. S., Richards, L. G., & Stiers, W. (2003). Attentional abilities and functional outcomes following stroke. *Journal of Gerontology, 58*, 45–53.

Mondini, S., Mapelli, D., Vestri, A., & Bisiacchi, P. (2003). *Esame Neuropsicologico Breve*. Milan, Italy: Raffaello Cortina Editore.

Nys, G. M. S., van Zandvoort, M. J. E., de Kort, P. L. M., Jansen, B. P. W., de Haan, E. H. F., & Kappelle, L. J. (2007). Cognitive disorders in acute stroke: Prevalence and clinical determinants. *Cerebrovascular Disease, 23*, 408–416.

Rand, D., Katz, N., Shahar, M., Kizony, R., & Weiss, P. L. (2005). The virtual mall: A functional virtual environment for stroke rehabilitation. *Annual Review of Cybertherapy and Telemedicine (ARCTT): A Decade of VR, 3*, 193–198.

Rand, D., Katz, N., & Weiss, P. L. (2007). Evaluation of virtual shopping in the VMall: Comparison of post-stroke participants to healthy control groups. *Disability and Rehabilitation, 13*, 1–10.

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

Raspelli, S., Carelli, L., Morganti, F., Poletti, B., Corra, B., Silani, V., & Riva, G. (2010). Implementation of the Multiple Errands Test in a NeuroVR-supermarket: A possible approach. *Annual Review of CyberTherapy and Telemedicine, 8*, 92–95.

Raz, N. (2000). Aging of the brain and its impact on cognitive performance: Integration of structural and functional findings. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed.) (pp. 1–90). Mahwah, NJ: Lawrence Erlbaum.

Riva, G., Carelli, L., Gaggioli, A., Gorini, A., Vigna, C., Algeri, D., Repetto, C., Raspelli, S., Corsi, R., Faletti, G., & Vezzadini, L. (2009). NeuroVR 1.5 in practice: Actual clinical applications of the open source VR system. *Studies in Health Technology and Informatics, 144*, 57–60.

Rizzo, A. A., Buckwalter, J. C., & Van der Zaag, C. (2002). Virtual environment applications in clinical neuropsychology. In K. Stanney (Ed.), *The handbook of virtual environments* (pp. 1027–1064). New York: Lawrence Erlbaum Associates, Inc.

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

Sbordone, R. J. (1996). Ecological validity: Some critical issues for the neuropsychologist. In R. J. Sbordone & C. J. Long (Eds.), *Ecological validity of neuropsychological testing* (pp. 15–41). Delray Beach, FL: GR Press/St. Lucie Press.

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

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. (1970). *The State Trait Anxiety Inventory (STAI). Test manual for form X* (R. E. Lazzari & P. Pancheri, Trans.). Palo Alto, CA: Consulting Psychologists Press.

Spinnler, H., & Tognoni, G. (1987). Standardizzazione e taratura italiana di test neuropsicologici. *Italian Journal of Neurological Sciences, 8*(6), 20–120.

Stroop, J. R. (1935). Studies of interference in serial verbal reaction. *Journal of Experimental Psychology, 18*, 643–662.

Wiechmann, D., & Ryan, A. M. (2003). Reactions to computerized testing in selection contexts. *International Journal of Selection and Assessment, 11*(2–3), 215–229.

Wilson, B. A., Alderman, N., Burgess, P. W., Emslie, H., & Evans, J. J. (1996). *Behavioural assessment of the dysexecutive syndrome.* London: Harcourt Assessment.

Wilson, B., Cockburn, J., & Halligan, P. W. (1987). *The Behavioural Inattention Test.* 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.

Zimmerman, P., & Fimm, B. (1992). *Test Batterie zur Aufmerksamkeitsprüfung (TAP).* Würselen: Psytest.