

# Psychometric Reliability of the NeuroVR-based virtual version of the Multiple Errands Test

Pietro Cipresso, Silvia Serino, Elisa Pedroli,  
Giuseppe Riva

Applied Technology for Neuro-Psychology Lab, IRCCS  
IRCCS Istituto Auxologico Italiano,  
Via Pellizza da Volpedo 41, 20149 Milano, Italy  
p.cipresso@auxologico.it

Giovanni Albani

IRCCS Istituto Auxologico Italiano,  
Department of Neurosciences, Istituto Auxologico  
Italiano,  
Piancavallo-Verbania, Italy

Giuseppe Riva

Psychology Department  
Catholic University of Milan  
Largo Gemelli, 1, 20123 Milan, Italy

**Abstract**—This study aims at testing the psychometric reliability of the NeuroVR-based virtual version of the Multiple Errands Test (VMET). In a first experiment two independent researchers scored 11 videos, each one of a subject running the VMET. In a second experiment 7 researchers scored two videos, each one of a subject running the VMET. Results showed the extent to which a disagreement becomes critical in VMET.

**Keywords**- *Psychometric Reliability; VMET; Virtual Reality; Executive Functions; NeuroVR; Multiple Errand Test.*

## I. INTRODUCTION

Recently, many researches have been conducted demonstrating the efficacy of using Virtual Reality (VR) in clinical setting, and in particular in psychotherapy, neuropsychology and neurorehabilitation. This approach is based on the use of a technology that allows to simulate daily life experiences through 3D interactive environments generated by the computer. Combining the simulation of 3D environments with audio tactile devices and immersive display like virtual head mounted display it's possible to induce in the user the feeling of interacting in a real environment and so to create situations and exercises which could help the therapeutic action, within the safe context of the therapist's laboratory. Moreover, the use of sensors of movements (such as the head-tracker) allows recording user's behaviors and to use these information to improve assessment and treatment decisions [1-6].

Thanks to these features, VR is becoming a tool more and more employed in the neuropsychological and neuro-rehabilitative fields. Within these contexts, a fundamental aim is to foster quantitative and qualitative improvements in daily-life activities (ADL) in order to promote an independent lifestyle [7-10].

A fundamental feature of VR to support this process is interaction: thanks to this feature, it is possible to lead patient to interact within the simulated environment and manipulate the objects inside through the use of haptic devices and input peripherals such as the data-glove and the head-tracker. The "controlled immersion" in the virtual environment allows the

patient to perform exercises specifically planned for this kind of approach (for example, to catch a virtual ball moving) and to learn again abilities compromised by neurological damages. Mobile Phone recent development further increased these possibilities [11-14]

Thanks to these features, VR makes possible to satisfy the main principles of an effective rehabilitative process, like the repeated practice, the feedback on the performance and the motivation to support patient's compliance with the prescription of the therapeutic protocol.

On the other side, VR leads the therapist to constantly monitor patient's performance and to quantify his/her improvements (or relapses) giving an objective evaluation of his/her state in time. Another important advantage regards the benefits of the experience of "presence", that is the perception of the simulation as it is a real experience [10]. Many studies suggest that virtual reality is, among the interactive media, the most able one to increase the sense of presence and so to transfer abilities and competencies got within the simulated environment in the real one. These features reflect in the results of the studies on the assessment of the efficacy of virtual reality as a neuro-rehabilitative technology. Specifically, VR is a tool potentially very useful to rehabilitate daily life activities which require specific executive and motor functions.

For example, it could be really useful for the assessment and rehabilitation of executive functions impairments [3]. The term "executive functions" refers to a set of behavioral competencies which include planning, sequencing, the ability to sustain attention, resistance to interference, utilization of feedback, the ability to co-ordinate simultaneous activity, cognitive flexibility and, more generally, the ability to deal with novelty. The neural substrates of these competencies are considered to lie in the prefrontal cortex. These abilities play a critical part in complex social behavior, help to suppress improper actions and to focus on purposeful information. The "dysexecutive syndrome" refers to a cluster of deficits in executive functions. Individuals who suffer from executive function impairments, including attention, planning, problem solving and behavioral control [15, 16], present problems of

## II. MATERIAL AND METHODS

starting and stopping activities, the difficulty in mental and behavioural shifts, increased distractibility and difficulties learning novel tasks despite the fact that their “basic” cognitive abilities are preserved [15]. Loss of executive functions is primarily a consequence of brain injury located in the prefrontal cortex area. The most common causes of executive functions impairments are neurological conditions including frontal lobe damage due to traumatic brain injury, pervasive CNS damage such as stroke [17] and those with specific pathologies such as Parkinson’s disease (PD).

On a methodological level, the assessment of executive functions under typical clinical or laboratory conditions are unsatisfactory for several reasons [18]. In such settings, planning, multi-tasking or problem solving are usually done via pen and paper tasks rather than being presented in an actual or simulated manner. There are also tools developed to measure executive deficits in situations similar to daily ones, such as such as the Behavioral Assessment of Dysexecutive Syndrome, the Dysexecutive Questionnaire (BADS & DEX) [19] the Multiple Errands Test (MET) [16]. More specifically, the MET is an assessment of executive functions in daily life originally developed by Shallice and Burgess [16] specifically for high functioning patients and adapted into the simple version [15] and the hospital version [20].

It consists of tasks that abide by certain rules and is performed in a mall-like setting or shopping center. After the tasks and the rules have been explained, patients are able to plan and choose the sequence of actions to complete the tasks. In this way the executive functions stimulated are numerous, from the ability to plan a sequence of actions, to problem solving and to cognitive and behavioral flexibility.

The tester follows the participant, recording mistakes of different kinds. Scoring of the MET relates to the numbers of mistakes of observed behavior in different categories including non efficiency, rule breaking and use of strategies mistakes in addition to the partial and complete mistakes of completing a task and the total number of mistakes.

To perform the Multiple Errands Test, the user has to go with the therapist to a real mall, with walking and mobility abilities required for both reaching the target place and moving into and around the mall. This could represent a problem if the patient has motor difficulties. In addition, this procedure could take a long time, for both patient and therapist, which could instead be used for more training in the rehabilitative setting. Finally, real shop features and structure are totally uncontrolled by the therapist, so that unpredictable and potentially disturbing changes may occur when the test takes place.

In this sense virtual reality has great potential for use in the rehabilitation of everyday life activities, involving cognitive and motor functions. The use of simulated environments, perceived by the user as comparable to real world objects and situations, can overcome the limits of the traditional tests employed to assess , by keeping intact its several advantages.

The studies support the hypothesis that virtual environments are very useful in assessment and rehabilitation since they increase the ecological validity and the efficacy of actual therapies.

### A. VMET Scoring procedure

For specification about the VMET procedure see the paper from Raspelli and colleagues [3]. Here we just gone deeply on the scoring procedure.

When the VMET test started, the examiner showed the shop, illustrating different shop sections. Then, the researcher gave different materials: a shopping list, a list of rules, a shop’s map, some information (opening and closing time, products in sale, etc.), papers, a pencil and a watch. Participants had always enough space on the table.

The researcher and the participant read together the following rules:

- you have to execute all the proposed tasks, but you can run them in any order;
- you cannot go in a place unless this is a part of a task;
- you cannot pass through the same passage more than once;
- you cannot buy more than two items per categories (look at the chart);
- take as few time as possible to complete this exercise however without hurry;
- do not talk to the researcher unless this is a part of the task;
- go to your "shopping cart" after 5 minutes from the beginning of the task and make a list of all the products that you bought.

The examiner started clocking without talking with subject not even to answer to the questions. The time stopped when the subject said "I finished". During the task the examiner took notes about subject’s behaviors and he paid attention to the different variables: Errors, Inefficiencies, Rule Breaks, Strategies and Interpretation Failures.

At the end, the researcher asked some information to the participant, who had to use all materials available and/or searching freely into the shop. These questions were:

- are there spoons on sale in the kitchen products department?
- which is the closing time of the supermarket?
- fruit is sold on how many shelves?
- how many departments are there in the supermarket?

The variables were coded as following.

**Errors:** Errors are task failures or partial omissions (defined as failing to attempt the task).

**Inefficiencies:** defined as a failure to do more than one thing in one place when is the only place to accomplish that task.

**Rule Breaks:** defined as anything that violates the rules listed in the MET task list.

**Strategies:** Examples of the 13 strategies are planning before starting the tasks and marking off the tasks completed.

**Interpretation Failures:** insight into the type of errors and interpretation failures experienced by the subject in the testing situation.

### B. Reliability Experiments

VMET is a test that might strongly depends on the scorer interpretation. To assess the variability among different scorers that came to use this test, we did two experiments:

**First experiment:** we asked to two independent scorer to score 11 videos, each one of a subject running the VMET.

**Second experiment:** we asked to 7 researcher to score two videos, each one of a subject running the VMET.

At the end of each scoring a researcher inserted data scored into a database for the statistical analysis.

## III. RESULTS

There are several methods and measures that can be used to calculate the test reliability [21]. For the First Experiment, with two scorer we used the following measures:

*Intra-class correlation coefficient (ICC)*, a measure of reliability, was used to assess the two observers agreement. ICC can be computed using a one-way or a two-way model.

*Lin's concordance correlation coefficient (CCC)* is the Pearson coefficient of correlation, and assesses the closeness of the data. [8-13] It is used to represent a Concordance Correlation Plot to measure the fit.

*information-based measure of disagreement (IBMD)* provides a useful tool to compare the degree of observer disagreement among different methods, populations or circumstances [18].

For the second experiment previous measures are critical, so we used the following: A corrected measure of disagreement denoted "corrected information-based measure of disagreement" (CIBMD) [18] based on the Shannon's notion of entropy [19], defined as the average amount of information contained in a variable.

Results of First Experiment are reported in Table I.

TABLE I. RELIABILITY STATISTICS FOR TWO SCORER DISAGREEMENT TESTS

	ICC (one-way)		ICC (two-way)		IBMD	
	ICC	95% CI	ICC	95% CI	IBMD	95% CI
<b>Errors</b>	0.88	(0.633, 0.966)	0.88	(0.616, 0.966)	0.06	(0.029, 0.109)
<b>Inefficiencies</b>	0.717	(0.264, 0.914)	0.732	(0.127, 0.927)	0.14	(0.074, 0.215)
<b>Rule Breaks</b>	0.696	(0.226, 0.907)	0.71	(0.178, 0.915)	0.1	(0.072, 0.164)
<b>Strategies</b>	0.266	(-0.343, 0.727)	0.243	(-0.445, 0.727)	0.15	(0.107, 0.226)
<b>Interpretation Failures</b>	0.286	(-0.324, 0.737)	0.306	(-0.258, 0.74)	0.1	(0.044, 0.189)

Results of Second Experiment are reported in Table II.

TABLE II. RELIABILITY STATISTICS FOR SCORERS DISAGREEMENT TESTS

	IBMD	
	IBMD	95% CI
<b>Errors</b>	0.1	(0.083, 0.126)
<b>Inefficiencies</b>	0.1	(0.087, 0.114)
<b>Rule Breaks</b>	0.06	(0.060, 0.063)
<b>Strategies</b>	0.09	(0.085, 0.100)
<b>Interpretation Failures</b>	0.09	(0.063, 0.135)

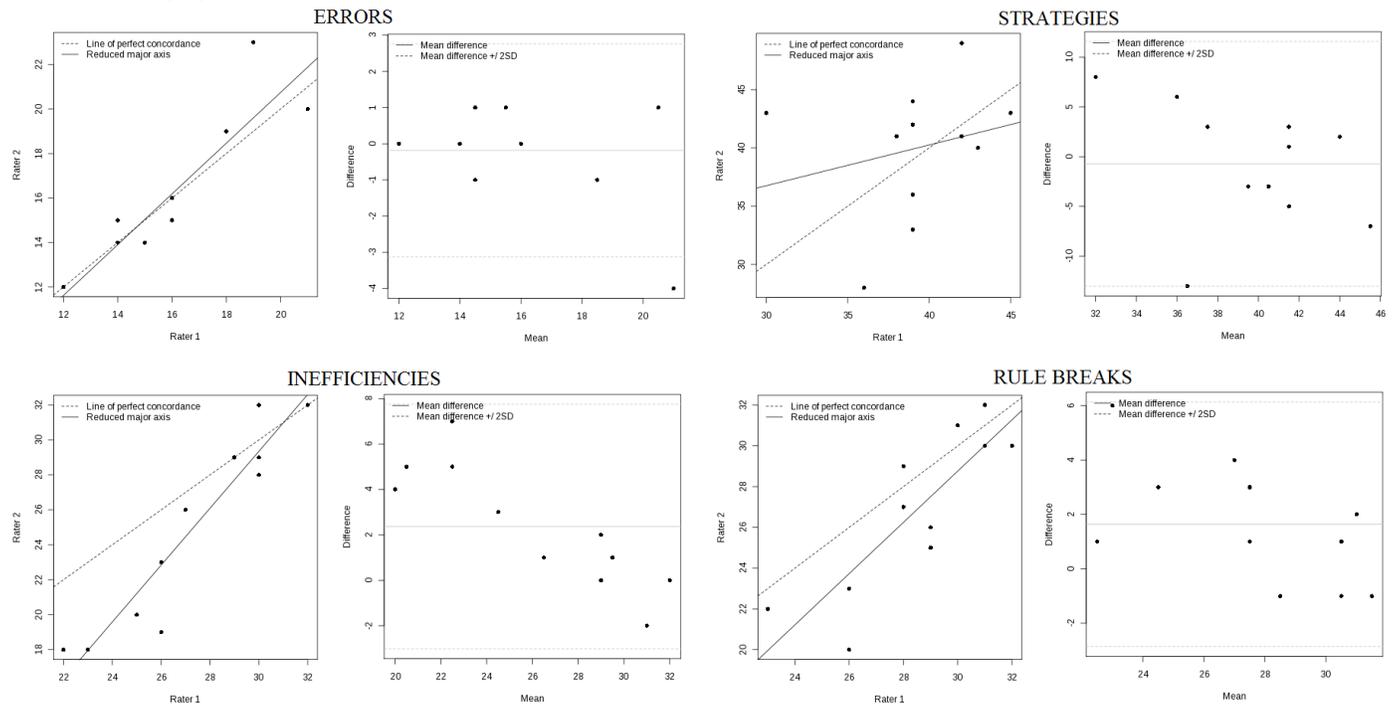


Figure 1. Concordance Correlation Plots (left side of each couple) and Bland And Altman Plot (right side of each couple).

#### IV. DISCUSSIONS

The use of virtual reality (VR) in clinical psychology has become more widespread [22]. The key characteristics of virtual environments for most clinical applications are the high level of control of the interaction with the tool, and the enriched experience provided to the patient [23]. Typically, in VR the patient learns to cope with problematic situations related to his/her problem. For this reason, the most common application of VR in this area is the treatment of anxiety disorders, i.e., fear of heights, fear of flying, and fear of public speaking [24, 25]. Since the clinical use of Virtual Reality has become more widespread, there is always more request of reliability and validity for the VR-based tests.

Our results showed a good reliability of the VMET Test. Even if a more extended data collection is needed, these results seems to be highly encouraging. Studies that make use of VMET scoring as here indicated can also consider to have just one scorer for their tests.

#### ACKNOWLEDGMENT (HEADING 5)

This work was supported by the Italian funded project "VRRehab. Virtual Reality in the Assessment and TeleRehabilitation of Parkinson's Disease and Post-Stroke Disabilities" - RF-2009-1472190.

#### REFERENCES

- [1] Riva G: Applications of virtual environments in medicine. *Methods Inf Med.* 2003;42(5):524-34.
- [2] Gaggioli A, Keshner EA, Weiss PL, Riva G (Eds.): *Advanced Technologies in Rehabilitation - Empowering Cognitive, Physical, Social and Communicative Skills through Virtual Reality, Robots, Wearable Systems and Brain-Computer Interfaces.* Amsterdam, IOS Press, 2009.
- [3] Raspelli S, Pallavicini F, Carelli L, Morganti F, Pedroli E, Cipresso P., Poletti B, Corra B, Sangalli D, Silani V, Riva G (2012). Validating the Neuro VR-based virtual version of the Multiple Errands Test: preliminary results. *Presence-Teleoperators And Virtual Environments*, vol. 21; p. 31-42, doi: 10.1162/PRES\_a\_00077
- [4] Cipresso P., Gaggioli A, Serino S, Pallavicini F, Raspelli S, Grassi A, Riva G (2012). EEG alpha asymmetry in virtual environments for the assessment of stress-related disorders. *Studies In Health Technology And Informatics*, vol. 173; p. 102-104, doi: 10.3233/978-1-61499-022-2-102
- [5] Villani D, Repetto C, Cipresso P., Riva G (in press). May I experience more presence in doing the same thing in virtual reality than in reality? An answer from a simulated job interview. *Interacting With Computers*, 24(4): 265-272; doi: 10.1016/j.intcom.2012.04.008
- [6] Pallavicini F, Cipresso P., Raspelli S, Grassi A, Serino S, Vigna C, Triberti S, Villamira M, Gaggioli A, Riva G (2013). Is Virtual Reality Always an Effective Stressors for Exposure Treatments? Some Insights from a Controlled Trial. *BMC Psychiatry*, vol. 13:52, doi: 10.1186/1471-244X-13-52
- [7] Cipresso P., La Paglia F., La Cascia C., Riva G., Albani G., La Barbera D. (in stampa). Break in volition: A virtual reality study in patients with

- obsessive-compulsive disorder. *Experimental Brain Research.* Doi: 10.1007/s00221-013-3471-y
- [8] Repetto C, Gaggioli A, Pallavicini F, Cipresso P., Raspelli S, Riva G (2013). Virtual reality and mobile phones in the treatment of generalized anxiety disorders: a phase-2 clinical trial. *Personal And Ubiquitous Computing*, 17: 2, pp. 253-260 February, doi: 10.1007/s00779-011-0467-0
- [9] Serino S, Cipresso P, Gaggioli A, Riva G (2013). The Potential of Pervasive Sensors and Computing for Positive Technology: The Interreality Paradigm. In: *Pervasive and Mobile Sensing and Computing for Healthcare, SSMI 2.* Edited by: Mukhopadhyay S C et al.. p. 207-232, Springer
- [10] Riva G, Gaggioli A, Grassi A, Raspelli S, Cipresso P, et al. (2011) NeuroVR 2--a free virtual reality platform for the assessment and treatment in behavioral health care. *Stud Health Technol Inform* 163: 493-495.
- [11] Villani D, Grassi A, Cagnetta C, Cipresso P., Toniolo D, Riva G (2012). The Effects of a Mobile Stress Management Protocol on Nurses Working with Cancer Patients: a Preliminary Controlled Study. *Studies In Health Technology And Informatics*, vol. 173; p. 524-528, doi: 10.3233/978-1-61499-022-2-524
- [12] Cipresso P., Serino S, Villani D, Repetto C, Sellitti L, Albani G, Mauro A, Gaggioli A, Riva G (2012). Is your phone so smart to affect your state? An exploratory study based on psychophysiological measures. *Neurocomputing*, vol. 84; p. 23-30, doi: 10.1016/j.neucom.2011.12.027
- [13] Villani D, Grassi A, Cagnetta C, Toniolo D, Cipresso P., Riva G (in press). Self-Help Stress Management Training Through Mobile Phones: An experience with oncology nurses. *Psychological Services*, ISSN: 1541-1559, doi: 10.1037/a0026459
- [14] A Gaggioli, G Pioggia, G Tartarisco, G Baldus, D Corda, P Cipresso, G Riva. A Mobile Data Collection Platform for Mental Health Research. *Personal and Ubiquitous Computing* 17: 2. 241-251 February 2013
- [15] 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.
- [16] Shallice, T., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, 114, 727-741.
- [17] Baddeley, Alan, and Barbara Wilson. "Frontal amnesia and the dysexecutive syndrome." *Brain and cognition* 7.2 (1988): 212-230.
- [18] Rand, Debbie, et al. "Validation of the Virtual MET as an assessment tool for executive functions." *Neuropsychological Rehabilitation* 19.4 (2009): 583-602.
- [19] Evans, J. J., et al. "Assessment of the Dysexecutive Syndrome." *Psychological Medicine* 27.3 (1997): 635-646.
- [20] Knight, Caroline, Nick Alderman, and Paul W. Burgess. "Development of a simplified version of the multiple errands test for use in hospital settings." *Neuropsychological Rehabilitation* 12.3 (2002): 231-255.
- [21] Henriques, Teresa, et al. "Information-based measure of disagreement for more than two observers: a useful tool to compare the degree of observer disagreement." *BMC medical research methodology* 13.1 (2013): 1-6.
- [22] Riva, G. (2005). Virtual reality in psychotherapy: review. *CyberPsychology & Behavior*, 8(3), 220-230; discussion 231-240.
- [23] Schultheis, M. T., & Rizzo, A. A. (2001). The Application of Virtual Reality Technology in Rehabilitation. *Rehabilitation Psychology*, 46(3), 296-311.
- [24] Emmelkamp, P. M. (2005). Technological innovations in clinical assessment and psychotherapy. *Psychotherapy & Psychosomatics*, 74(6), 336-343.
- [25] Wiederhold, B. K., & Rizzo, A. (2005). Virtual reality and applied psychophysiology. *Applied Psychophysiology and Biofeedback*, 30(3), 183-185.