A Virtual Reality Test for the Assessment of Cognitive Deficits: Usability and Perspectives

Elisa Pedroli, Pietro Cipresso, Silvia Serino,
Giuseppe Riva
Applied Technology for Neuro-Psychology Lab, IRCCS
IRCCS Istituto Auxologico Italiano,
Via Pellizza da Volpedo 41, 20149 Milano, Italy
e.pedroli@auxologico.it
p.cipresso@auxologico.it
s.serino@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— **The assessment of cognitive deficits, such as executive functions, can be critical in neuropsychology because is not easy to catch all the important cues of real-life behavior using the classic paper and pensil tests or even sophisticated laboratory tasks. On the other hand, using virtual simulations of prototypical daily situations, by the means of 3D virtual environments, it is possible to create situations which could help the therapeutic assessment, within the safe context of the therapist's laboratory. In this study, we described how we assessing executive functions using a Virtual Reality Test.**

Keywords: Usability, Psychometrics, Virtual Reality Test, VMET, Assessment, Cognitive Deficits

I. INTRODUCTION

In the assessment of cognitive deficits a key goal is to implement a task in daily activities to improve the quality of neuropsychological testing. A key feature of all VR applications that can support this process is interaction. Virtual environments (VE) are created to allow the user to interact with not only the VE but also with virtual objects within the environment. In some systems, the interaction may be achieved via a pointer operated by a mouse or joystick button. In other systems, a re-presentation of the user's hand (or other body part) may be generated within the environment where movement of the virtual hand is "slaved" to the user's hand allowing a more natural interaction with objects. Finally, while many applications of VR allow the user to control the viewpoint on the screen, third-person views or images of the users themselves that appear as players in the environment also provide the opportunity for interaction with the VE.

In recent years the field of VR has grown immensely. Practical applications for the use of this technology encompass many fields, from aviation training and military applications, to medicine, where surgeons can be trained in surgical techniques using VR systems. One of the newest fields to benefit from the advances in VR technology is that of assessment and rehabiltation [1-4].

Thank to these features, VR provides a unique medium suited to the achievement of several requirements for effective rehabilitation intervention: repetitive practice, feedback about performance, and motivation to endure practice [5-8]. Specifically, using VR it is also possible to drive and control exercises for patient rehabilitation within a functional, purposeful and motivating context [9-10]. More, several studies are exploring the capabilities of Mobile phones for the assessment and the inclusion of adapted VE [11-14].

An example of ecological task is the Multiple Errands Test (MET) [15, 16], which is performed at a real shopping mall or in a hospital environment and involves the completion of various tasks. 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 but could be a problem

Executive functions regards a wide range of cognitive processes like problem solving, planning, attention, sequencing, cognitive flexibility, multitasking and many others [17].

Recently, many studies demonstrated the efficacy of using Virtual Reality (VR) in clinical setting, and in particular for neuropsychological assessment. VR is based on the use of a technology that allow to simulate daily life experiences through 3D interactive environments generated by the computer. Enriching the simulation of 3D virtual environments with prototypical daily situations, it's possible to induce in the patient the feeling of interacting in a real environment and so to create situations which could help the therapeutic assessment, within the safe context of the therapist's laboratory. In particular, with this study, we aimed at assessing executive functions by using a virtual supermarket. The Virtual Multiple Errands Test was used as a validated protocol. Using virtual reality devices may be difficult for subject, for this reason we also decided to analyzed usability of this tool.

II. MATERIALS AND METHODS

A. The Virtual Multiple Errands Test

The virtual environment employed in this study is a supermarket developed via NeuroVR software [10] and displayed on a desktop monitor. It consists of a Blender-based

application that enables active exploration of a virtual supermarket where users are requested to select and buy various products presented on shelves. 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.

The original procedure of the Multiple Errands Test (MET) [16] 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).

B. Neuropsychological evaluation

A neuropsychological evaluation was conducted to detect the exclusion criteria and to have a cognitive profile of our participants.

We investigated different cognitive domains.

To assess the general cognitive level we employed the Mini-Mental Status Evaluations [18], for memory we used the Auditory-Verbal Learning Test and the Novelli's Battery composed Digit Span [19], Corsi's Span [19] and supra-span [20], Short Story [21].

For the assessment of executive function we used the Tower of London test and Verbal Fluency (semantic and phonemic) [21].

We used BIT and Benton's Judgment of Line Orientation to evaluated the visuospatial function and Similarities from WAIS-R and Laiacona's Naming Test to assessment language [22].

The Trial Making Test [23] and the Digit Span Backward from WAIS-R was used to evaluate attention and working memory.

To evaluate the level of state and trait anxiety and depression we used State and Trait Anxiety Index [24] and the Back Depression Inventory [25].

All the test score were recorded and correct for age, education level and gender.

C. Usability

The international standard, ISO 9241-11 provides guidance on usability and defines it as [26]:

"The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use".

ISO 9241-11 explains how to identify the information that it is necessary to take into account when specifying or evaluating usability in terms of measures of user performance and satisfaction [26]. Guidance is given on how to describe the context of use of the product and the measures of usability in an explicit way. It includes an explanation of how the usability of a product can be specified and evaluated as part of a quality system.

Usability can be identified through the following factors:

- *Effectiveness* - can users complete tasks, achieve goals with the product, i.e. do what they want to do?

- *Efficiency* - how much effort do users require to do this? (Often measured in time)

- *Satisfaction* – what do users think about the products ease of use?

These factors are affected by:

- *The users* - who is using the product? e.g. are they highly trained and experienced users, or novices?

- *Their goals* - what are the users trying to do with the product - does it support what they want to do with it?

- *The usage situation* (or 'context of use') - where and how is the product being used?

Usability should not be confused with "functionality," however, as this is purely concerned with the functions and features of the product and has no bearing on whether users are able to use them or not. Increased functionality does not mean improved usability.

There are a series of international standards for usability and user centered design.

To evaluate usability we used The System Usability Scale (SUS), a "quick and easy to use" measure developed by Brooke [27]. SUS is composed by ten item to evaluate usability defined as "the subjective perception of interaction with a system." The original SUS instrument [27], is composed of 10 statements that are scored on a 5-point scale of strength of agreement. Final scores for the SUS can range from 0 to 100, where higher scores indicate better usability. Because the statements alternate between the positive and negative, care must be taken when scoring the survey.

Products with scores less than 50 should be considered candidates for increased scrutiny and continued improvement and should be judged to be marginal at best, between 50 and 70 the product can be defined good and over 80 is excellent [28].

The advantages of this scale are that quick and easy to use by both study participants and administrators, that are the survey provides a single score on a scale that is easily understood by the wide range of people and that the survey is nonproprietary, making it a cost effective tool as well [28].

Figure 1. A screenshot of the Virtual Environment for the Multiple Errand Test.

III. PROCEDURE

Initially, the subjects underwent neuropsychological screening tests; if this preliminary evaluation didn't show any problem we continued with the administration.

After that, the subject started with experimental part. Before the test, the subject began a training phase in a different market to test the joypad use. In this phase the subject explored freely the environment for few minutes or as long as he learnt the use of joypad.

When the subject was ready, we started with the test in anther market. The examiner showed the shop, illustrating different sections. After that, the doctor gave different items: shopping list, a list of rules of conduct, a shop's map, some information (opening and closing time, products in sale, etc.), a white paper, a pencil and a watch. The subject had enough space on the table.

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 examiner asked some information to the subject, who had to use all materials available and/or searching freely into the shop.

All tests were corrected for age, gender and years of school.

IV. RESULTS

In Table I and Table II are indicated the main results arose from the usability tests for the 21 healthy subjects.

TABLE I. DESCRIPTIVE STATISTICS

Descriptive Statistics					
	N	Min	Max	Mean	Std. Dev.
Age	21	21	66	40.86	17.318
SUS Test	21	20.0	97.5	70.595	19.1034
Joypad	21	2	5	3.90	.944
Movement	21	1	5	4.00	1.304
Catch	21	2	5	4.19	1.030
Recognize	21	2	5	4.00	1.265
Organization	21	1	5	3.52	1.289
Charts	21	2	5	4.52	.981

TABLE II.   CORRELATIONS BETWEEN SUS TEST AND VIRTUAL REALITY SPECIFIC USABILITY FEATURES

		SUS Test	Joypad	Movement	Catch	Recognize	Organization	Charts
SUS Test	Pearson Correlation	1	.288	.657**	.464*	.222	.124	.516*
	Sig. (2-tailed)		.206	.001	.034	.332	.593	.017
	N	21	21	21	21	21	21	21
Joypad	Pearson Correlation	.288	1	.244	.122	.084	.084	.219
	Sig. (2-tailed)	.206		.287	.597	.718	.717	.341
	N	21	21	21	21	21	21	21
Movement	Pearson Correlation	.657**	.244	1	.112	-.061	-.030	.430
	Sig. (2-tailed)	.001	.287		.630	.794	.898	.052
	N	21	21	21	21	21	21	21
Catch	Pearson Correlation	.464*	.122	.112	1	.614**	.373	.391
	Sig. (2-tailed)	.034	.597	.630		.003	.096	.080
	N	21	21	21	21	21	21	21
Recognize	Pearson Correlation	.222	.084	-.061	.614**	1	.399	.202
	Sig. (2-tailed)	.332	.718	.794	.003		.073	.381
	N	21	21	21	21	21	21	21
Organization	Pearson Correlation	.124	.084	-.030	.373	.399	1	-.070
	Sig. (2-tailed)	.593	.717	.898	.096	.073		.764
	N	21	21	21	21	21	21	21
Charts	Pearson Correlation	.516*	.219	.430	.391	.202	-.070	1
	Sig. (2-tailed)	.017	.341	.052	.080	.381	.764	
	N	21	21	21	21	21	21	21

**. Correlation is significant at the 0.01 level (2-tailed).

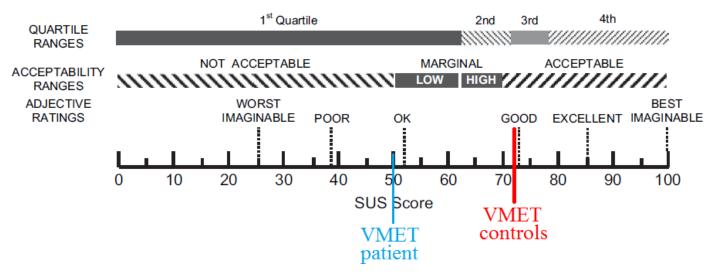*. Correlation is significant at the 0.05 level (2-tailed).

Figure 2.   The VMET Performances on the SUS Tests for the patient and the controls.

As represented in Figure 2, healthy participants gave a good usability for the VMET. From correlations table (See Table II) it is clear that Virtual Reality has itself some peculiarities that makes it unique and we strongly suggest for further works to evaluate out of SUS Test also these crucial aspects.

A. A case study: Usability in a Parkinson's Disease Patient

The clinical picture of Parkinson's Disease (PD) is characterized by a progressive deterioration of motor performances, with the occurrence of slowness and poverty of voluntary movements, expressionless face, "resting" tremor, stopped posture, festinating gait, axial instability, etc. Although the symptoms are improved by dopaminergic drugs such as L-dopa, over time the pharmacological therapy become less effective, and very disabling.

Specifically, this model of medical care contains important limitations: a) the follow up relies on a semi-quantitative evaluation (UPDRS, Unified Parkinson's disease Rating Scale [29], and Hoehn & Yahr staging of disease [30]) based on the experience of the neurologist; b) in many advanced PD cases, repeated daily assessments of motor symptoms would provide very useful information, unavailable using self-rating diaries; c) clinical picture assessed in an single outpatient check up in the medical office hardly represents the clinical status in fluctuating patients; d) the transport to the medical office of patients with severe disabilities may cause significant discomfort. On the other hand, VR methods can be proposed to improve some movement disorders of PD patients in a rehabilitative approach complementary to the pharmacological treatment. The following results (Table III), collected in a PD patient, showed that there needs more than an improvement to VMET, even if it can be consider a valid test when the patient is well trained to use the Virtual Reality settings.

TABLE III.   DESCRIPTIVE STATISTICS FOR THE PATIENT

Age	SUS Test	Joypad	Movement	Catch	Recognize	Organization	Charts
77	50	2	3	3	4	4	2

V.   DISCUSSIONS

User experience design challenges for devices and services concerning the older population are related to the ergonomics of the system, the restriction of the devices and technical infrastructures, and the demanding context of use.

Specifically thinking to VMET the Virtual Worlds used in the assessment need to have the following requirements that can be considered as needed features:

- **Ease of use**. The target users are clinicians and they do not need to be advanced users in virtual reality applications. The platform must allow them to easily tailor their scenarios to the specific purposes they have. And for the end user – the patients – interacting with the environments should also be as natural and intuitive as possible.

- **High quality visual content**. Even if they have different goals, these applications will be naturally compared to modern videogames. The quality of the

contents and the resulting visual experience must be high enough to result in a compelling and capturing interactive session.

- **Multiple visualizations of the user's body:** both through customizable 3D avatars and real time video capture of the body of the user that will be integrated directly in the virtual environment.

- **Strong sense of immersion**, supported by all possible techniques (ranging from motion tracking to stereoscopic vision).

- **Support for external devices**, including haptic ones.

- **Collection and integration of biosensor data.**

To further develop a usable system we will follow the five guidelines suggested by Dunlop and Webster [6]:

1) Designing for mobility: Users of mobile devices will need to work with small devices and are thus likely to have a far from ideal working environment. The environment will change drastically as the user moves.

2) Designing for a widespread population: Users will probably have no formal training in the technologies and consider them as devices to be used rather than computers.

3) Designing for limited input/output facilities: Small screen size, poor sound output quality, small keyboards in size and number of keys are often hard to use when the user is on the move.

4) Designing for (incomplete and varying) context information: Mobile devices can be made aware of their context (e.g. GPS). This gives new information to the systems but brings problems of implying task and user level activities from sensor information and unreliable or patchy sensor coverage.

5) Designing for user multitasking: Multitasking and support for task interruption is one of the keys to successful design with mobile devices. The opportunities for, and frequency of, interruption are likely to be much higher, given the context in which the devices will be used.

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, Cognetta 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, Cognetta 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] 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.

[18] 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.

[19] A. Orsini, D. Grossi, E. Capitani, M. Laiacona, C. Papagno, Dr. G. Vallar Verbal and spatial immediate memory span: Normative data from 1355 adults and 1112 children The Italian Journal of Neurological Sciences December 1987, Volume 8, Issue 6, pp 537-548

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

[21] Novelli, G.; Papagno, C.; Capitani, E.; Laiacona, M.; et al "Tre test clinici di memoria verbale a lungo termine: Taratura su soggetti normali. / Three clinical tests for the assessment of verbal long-term memory function: Norms from 320 normal subjects." Archivio di Psicologia, Neurologia e Psichiatria, Vol 47(2), Apr-Jun 1986, 278-296.

[22] A Orsini, C Laicardi "WAIS-R Contributo alla taratura italiana"-Firenze, Organizzazioni Speciali, 1997

[23] Reitan R. M. (1958) VALIDITY OF THE TRAIL MAKING TEST AS AN INDICATOR OF ORGANIC BRAIN DAMAGE. Perceptual and Motor Skills: Volume 8, Issue , pp. 271-276.

[24] 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.

[25] 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.

[26] ISO/IEC 9241 - Ergonomic requirements for office work with visual display terminals (VDTs)

[27] Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), Usability evaluation in industry (pp. 189–194). London: Taylor & Francis.

[28] Bangor, A. W. (2000). Display technology and ambient illumination influences on visual fatigue at VDT workstations. Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA.

[29] Martinez-Martin P, Gil-Nagel A, Gracia LM, Gomez JB, Martinez-Sarries J, Bermejo F. Unified Parkinson's Disease Rating Scale characteristics and structure. The Cooperative Multicentric Group. Movement disorders : official journal of the Movement Disorder Society. Jan 1994;9(1):76-83.

[30] Hoehn MM, Yahr MD. Parkinsonism: onset, progression, and mortality. 1967. Neurology. Nov 2001;57(10 Suppl 3):S11-26.

[31] Dunlop M, Brewster S, "The challenge of mobile devices for human-computer interaction" Personal and Ubiquitous Computing, 6 (2002), 235-236.