A Pilot Study on Interactive Multisensory Environments for Neuropsychological Assessment

Mathyas Giudici
mathyas.giudici@polimi.it
Politecnico di Milano
Milan, Italy

Eleonora Beccaluva
eleonora.beccaluva@polimi.it
Politecnico di Milano
Milan, Italy
Università degli Studi di Milano-Bicocca
Milan, Italy

Mattia Gianotti
mattia.gianotti@polimi.it
Politecnico di Milano
Milan, Italy

Jessica Barbieri
jessica.barbieri@polimi.it
Politecnico di Milano
Milan, Italy

Giacomo Caslini
giacomo.caslini@polimi.it
Politecnico di Milano
Milan, Italy

Franca Garzotto
franca.garzotto@polimi.it
Politecnico di Milano
Milan, Italy

ABSTRACT

Measuring cognitive functions is a complex and challenging process, and researchers usually conduct their study in experimental settings and with standardized tests. Such an approach is raising more and more criticism. Especially about the capability of psychology's laboratory experiments to provide generalizable and accurate indicators of impairments outside the laboratory boundaries in everyday life. This is acknowledged as the "real-world or lab" dilemma. Most tests traditionally performed using paper-and-pencil may fail to detect the individual's difficulties in the real world. The scientific community has launched a quest for new digitized, interactive, and more "ecological" versions. Our work investigates a novel approach to this topic that exploits interactive Multi-Sensory Environments (iMSE) and questions how iMSEs could contribute to the neuropsychology assessment field. The paper describes *NEP-Neuro-Psychological Suite*, a set of game-based activities in iMSE inspired by widely adopted neuropsychological tests. The suite provides a context in which existing or new neuropsychological tests can be experimented, extended, and modified with stimuli, contents, and tasks closer to real-world situations. We report an exploratory empirical study (N=22) on a well-known test to assess attention skills, the Stroop Test. The results, although very preliminary, provide insights into the effects of transposing classic tests into a novel form and the role that iMSEs can play in the debate about the "real-world or lab" dilemma.

AVI 2022, June 6–10, 2022, Frascati, Rome, Italy

ACM ISBN 978-1-4503-9719-3/22/06…$15.00
https://doi.org/10.1145/3531073.3531169

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; **Interaction paradigms**; *Accessibility technologies*; • **Applied computing → Interactive learning environments**.

KEYWORDS

Interactive Multisensory Environment, Neuropsychological Assessment, Neuropsychological Test, Stroop Test, Digitization of Neuropsychological Tests

ACM Reference Format:
Mathyas Giudici, Eleonora Beccaluva, Mattia Gianotti, Jessica Barbieri, Giacomo Caslini, and Franca Garzotto. 2022. A Pilot Study on Interactive Multisensory Environments for Neuropsychological Assessment. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces (AVI 2022), June 6–10, 2022, Frascati, Rome, Italy.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3531073.3531169

1 INTRODUCTION

Cognitive impairment (CI) is a term that encompasses the clinical state between normal cognition and dementia or cognitive disability [24]. When we use the term Cognitive Impairment, we often refer to the aging population. Nevertheless, this condition also affects young people with Neurodevelopmental Disorders (NDD) or subjects with brain injury and degenerative diseases. Frequently, subjects with CI present a significant memory deficit, which may be accompanied by involvement of other cognitive functions, such as language, visuospatial skills, executive functions, and reasoning skills [7]. The impairment level can range from Mild (subjects notice degradation or changes in cognitive functions but no restriction to their ability to perform everyday activities) to Severe (overall degradation of cognitive functions, understanding, speech, and the ability to talk or write). This latter stage results in the inability to live independently [25], with very high social costs: an increased toll for hospitals, state welfare, and 24/7 assistance [6]. Evidence from literature (e.g., [17]) indicates that early discovery of impairments and early treatment can provide longer and more autonomous life and can reduce the economic and social strain of health care. The Mild cognitive impairment stage is "ideal" for intervention with preventive therapies.

The neuropsychological assessment is the key element that allows detection and prompt treatment [2] and the neuropsychological tests are the tools used by experts to evaluate subjects' real capability [16]. However, new insights in research cast doubt on the actual ability of traditional tests to photograph the cognitive abilities of individuals outside the experimental settings [34]. Everyday life context is substantially different from the experimental one, and the border between an impaired and a non-impaired function is complex to detect, especially in Mild CI [1]. This is acknowledged as the "real-world or lab" dilemma, and the cardinal points of this dilemma [12] are mainly two. The first is a firm review of the very nature of the tests themselves. Traditional tests may be less effective in evaluating the decline of the subjects' abilities due to the attempt to ground the evaluation on their best and once-in-a-lifetime performance [12]. The second is skepticism about the assessments settings. During tests, patience is required to perform tasks sitting in front of a table, using paper and pencil materials, and performing specific cognitive tasks, one at a time and without any possible distraction [12]. This condition is impractical in daily-life scenarios. Cognitive abilities are usually stressed by multiple stimuli, happening simultaneously and very often involving movement in the space around the subject rather than sitting at a table in an aseptic and unfamiliar room [32]. In our opinion, the core question of the "real-world or lab" dilemma is: *"Do these tests miss something when it comes to detecting an individual's impairments outside the laboratory edges, in everyday life?"* [5]. We believe that most of the current tests in use may not be sensitive enough to accurately detect "real-world" impairments, especially at an early stage. We hypothesize that if a given test (with the same stimuli and task) is transposed from paper-and-pencil to full-body interaction, the subjects will be compelled to use more cognitive abilities simultaneously. This is more likely to mimic everyday life conditions and might act as a facilitator or an element of difficulty as it occurs to the subject in daily routine. To test our hypothesis, we created *NEP - NEuro-Psychological Test Suite* and in this paper, we will discuss the design, implementation, and results of one of the Suite activities: the Stroop Test [10], a neuropsychological test extensively used to assess selective and divided attention. We performed a pilot study with 22 healthy and neurotypical subjects comparing each subject's NEP-based neuropsychological assessment and corresponding traditional tests. The results of our study dis-confirm our initial hypothesis and reveal that, in this specific case, the classic paper-and-pencil version is comparable to the embodied version in the iMSE. This evidence contributes to the "real-world or lab" dilemma, suggesting that exploring new versions of the classical tests might not always be necessary.

1.1 State of the Art

The first attempts to exploit digital technology for neuropsychological assessment date back to the '80s [13], developing simple applications with the same contents and tasks of traditional paper-and-pencil tests. The aim was to facilitate the scoring activities reduce costs and time, and gather additional information automatically (e.g., reaction time) for data analysis [33]. The discussion on the ecological validity of neuropsychological tests has promoted a shift in the design paradigm of digitalized tests: from pure digital replicas to reinterpretations of the tests, in order to provide and experiment with a wider set of more controllable stimuli and tasks, and to create situations closer to the real world. Virtual Reality (VR) was the main solution used to implement this view [21]. Early VR-based tests were simulations of construct-driven tests extended with gamification elements. For example, Pugnetti et al. [26] report a series of experiments in Immersive VR using headset and controllers based on the Wisconsin Card Sorting Test in which the association between identical elements were performed using matching colors, shapes, or number of portholes on virtual doors, asking the user to navigate a virtual house. Several VR versions of tests introduce distractors in the form of controlled stimuli (e.g., phone ringing, a car driving by, popup images) to enhance realism [33]. For example, Henry et al. [11] describe a VR system, rendered through a VR headset, based on the Stroop Test [10] to measure selective attention, in which the subject is asked to concentrate on the information proposed in television while the phone is ringing. Another example is the Virtual Errands Test (VET) [19], which simulates work-related errands using virtual reality simulation in the browser from a PC and the Virtual MET test. The latter is based on the Multiple-Errand Test (MET) developed by Shallice and Burgess [31], which has shown good accuracy in detecting failures in executive function and the capability of predicting difficulties in everyday life contexts. In MET, the tasks assigned to the user take place in a real shopping context. Its VR version simulates the same physical space and tasks and avoids the obvious practical limitation of the traditional test. The Virtual Errands Test (VET) [19] simulated work-related errands and was proved to lead to measures comparable to the traditional non-digital version. Serino et al. [30] investigate the use of 360° video in an immersive VR rendered through a headset for function-let assessment of episodic memory and pinpoint the potential of these media (immersive, realism, egocentrism) to enhance the ecologic validity of the test. A crucial example is Perrochon et al. [23], who intended to explore a full-body interaction approach to the Stroop Test using an electronic walkway. Instead of pointing to the word, as in the paper-and-pencil test, subjects complete the task by stepping on the correct words over the walkway. The electronic walkway was also used to digitalize the Trail Making Test [22], switching the paradigm of linking points using a paper-and-pencil tool demanding the subject to walk over them.

2 THE NEURO-PSYCHOLOGICAL TEST SUITE

2.1 Neuro-Psychological Test Suite in a nutshell

NEP (Neuro-Psychological Test Suite) is a set of activities for iMSEs inspired by *traditional* standardized neuropsychological tests acknowledged to measure memory and attention and designed in collaboration with domain experts. The Suite activities take place in the iMSE depicted: the Magic Room [9]. The logic of the tasks is the same as some widely used traditional tests (Digit Span Test [20], Trail Making Test [27], Stroop Test [10], and Corsi Test [14]). Still, NEP tests take the form of immersive game experiences involving wall or floor projections, smart objects, full-body interaction, and stimulation on multiple sensory channels.

GREEN	BLUE	BLUE	BLUE
RED	RED	RED	GREEN
BLUE	RED	BLUE	RED
RED	GREEN	GREEN	BLUE
GREEN	BLUE	RED	RED

Figure 1: Example of traditional paper-and-pencil Stroop Test

All NEP activities can be configured using a tablet. Particularly, for each game, it is possible to select several configurable parameters such as time of exposure to the stimulus, type of stimuli, number of stimuli, and level of difficulty to address all subjects' profiles and abilities. The tablet interface also enables the caregiver to control the flow (select, start, repeat, pause, stop, quit). Figure 2 shows the activity outline, features, and graphic content of the test presented in this paper.

2.2 The Interactive Multi-Sensory Environment

A Multisensory Environment (MSE) is a dedicated physical space equipped with physical materials and devices that provide stimulation for multiple senses. The most widely known MSE is Snoezelen which is used, among others, for the elderly with dementia [15] in many therapeutic and educational contexts. More recently, MSEs have evolved to *interactive* Multisensory Environments (iMSEs), where digital devices are integrated into physical materials, and the whole space sense user's actions and controls multisensory stimuli generation. An example is the *Magic Room* [8, 9], which is an iMSE that involves floor and wall projections, motion sensors, and a variety of smart appliances (lights and aroma emitters) and smart objects (e.g., digitally enhanced toys). The Magic Room provides a set of immersive interactive multisensory experiences for both neurotypical children and children with NDD. The activities are designed to promote socialization and inclusion, relaxation, or specific cognitive skills. Activities can be pre-customized to address the needs of each specific user or user group, and the caregiver can manually control their stimuli and interaction flow at run-time using a tablet-based application. An increasing number of studies [8, 15] have empirically investigated iMSEs as means to promote engagement, curiosity, learning, or socialization among both typical and atypical persons, benefits that also have a theoretical foundation (particularly, in the embodied cognition theory [35] and the sensory integration theory [18]). iMSEs have been widely used also in other contexts with fragile populations, like children with NeuroDevelopmental Disorders (e.g.,[9]) for learning, rehabilitation, and well-being purposes. In addition, iMSEs for the elderly have been investigated in several settings with encouraging results [3, 4]. However, the role of full-body interaction and immersive environments as assessment tools is scarcely investigated.

2.3 The Stroop Test Activity

In this paper, we present the iMSE version of the Stroop Test [10] that is widely used to assess *selective and divided attention*. In the *classical version* the test is structured by three sub-trials where patients are required to read, as quickly as possible, three charts. In the first two, subjects must read colors names printed in black ink and name different color patches. In the third, called *Interference* (Figure 1), the names of the colors are printed in an inconsistent color ink (for instance, the word *green* is printed in blue ink). Patients have to name the color of the ink instead of reading the word. In a nutshell, subjects are required to perform a less automated task (i.e., naming ink color) while inhibiting the interference arising from a more automated task (i.e., reading the word). This difficulty in inhibiting the more automated process is called the Stroop Effect [29]. We developed an iMSE version of the Stroop Test, called *Basic Interactive Stroop Test*, inspired by Perrochon et al. [23]. In the *Basic Interactive Stroop Test* (Figure 2), we present the same set of schema reported in the traditional test (Figure 1). The interaction, instead, is entirely modified: words are displayed on the floor projection, and the subjects must move into the room to catch them. Color names and colored patches are organized in rows, and they slide from the top to the bottom of the projection area under the participant's feet. The user must walk onto the correct sliding item before exiting the projected field. Instruction is provided by recorded audio. The interaction of the original test, based on vocal responses, is substituted by a motor interaction.

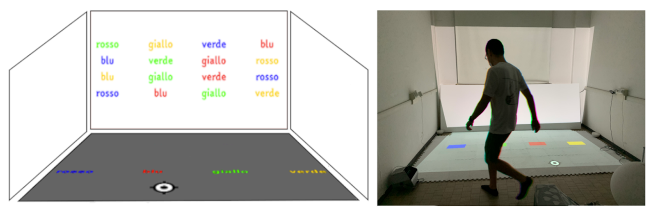

Figure 2: Basic Interactive Stroop Test

3 THE PILOT STUDY

We conducted a preliminary pilot study on healthy volunteers to establish a baseline of neurotypical performance. This is common practice as it provides normative data for the atypical population.

3.0.1 Goal and experimental design. Our goal was to answer the question: *Do iMSE tests and traditional tests deliver the same performances scores and, allegedly, measure the same cognitive abilities?* We were expecting different *performances scores* between the traditional test and its iMSE version. This pilot study features the administration of neuropsychological tests described above in two experimental conditions: *Condition A* - traditional paper-and-pencil test; *Condition B*- multi-sensory Smart iMSE test (*Basic Interactive Stroop Test*). The experimental design is a within-subject with one fixed factor: condition (traditional method vs. multisensory). Therefore, analyses involved a within design, whose dependent variable will be participants' performance (errors). For each participant, we gathered 6 scores.

3.1 Materials and Methods

3.1.1 Participants. The study involved 22 healthy Italian native speaker subjects (11 females and 11 males), mean age was 27 y.o. (range 24-37, SD 2.8). In addition, experimental conditions were evenly balanced.

3.1.2 Procedure. Participants underwent one session in which they were exposed to both experimental conditions according to their assigned order. The two experimental conditions were carried out within our lab. In the setting of *Condition A*, subjects were seated in a room with a table and a chair. The examiner, an experienced psychologist, stood at one side of the table, facing the subject. General biographical information was first collected for each participant, and then they were administered the traditional test. All materials were standardized, printed, and stored in separate packages. The duration of the session was approximately 5 minutes. In the *Condition B* setting, subjects were let into the iMSE. One of the researchers and the experienced psychologist were present within the room at all times. Once the room session was initiated, the system provided the instructions, and, if needed, further information was given by the psychologist. In addition, all of the subjects' responses were annotated in an organized note sheet. The duration of the session was approximately 10 minutes.

3.2 Data analysis

We computed different indexes for the classical test and the iMSE activity. We performed a first descriptive statistical analysis and quartile distributions to verify whether a subject's score ranked the same or a different one. We postulated that if subjects have a score corresponding to a quartile in the classical test, they will also have it in the iMSE activity if such activity measures the same ability. We calculated the percentages of subjects that had obtained the same quartile positioning. Then, we performed a second statistical analysis to explore significant correlations between subjects' scores in the two conditions. We performed Spearman's rank correlation and paired-sample t-Test on tests scores with a significant statistical correlation. All the analyses were performed using JAMOVI (Version 1.2.27) [28].

3.2.1 Tasks performances. Of all the 3 sub-trials comparison of all test, 2 have percentages of matching quartile above 77%, indicating that subjects performances in the iMSE activities are similar to the ones in classical tests. *Basic Interactive Stroop Test* that measured errors sub-trial *color name* 100% and *interference* 95,45%. The remaining sub-trial reports percentage of matching quartile positions below 77%, suggesting that a significant number of subjects' performance scores in this sub-trial were different in the two conditions.

3.2.2 Correlation and Statistical analysis. We performed multiple Spearman's rank correlations between scores obtained by the participants in both iMSE Activities and the classical test. The results show a statistically significant correlation in *Basic Interactive Stroop Test* (*interference*) and its traditional version ($r = 0.52$, $p < 0.014$). Finally, we performed a t-test to investigate the significance of subjects' mean performances in the two conditions for the test mentioned above: $t(21)=1{,}88$, $p=.001$.

3.3 Discussions

Our study presents several limitations; one above all is in the sample size. The selection and enrollment of the participants were strongly limited by the COVID-19 pandemic, the Omicron variant, the vaccination campaign, and social restriction and social distancing rules. Moreover, the activities proposed by our battery are not very, if at all, compatible with advanced CIs (e.g., Severe Impairment) or with significant physical (e.g., paralysis) or sensory impairments (e.g., blindness or deafness). They require subjects to be able to move in space independently, coordinate and execute movements with their upper limbs, and have unaltered sensory competencies (vision and hearing). Evidence collected so far, although very preliminary, provides an empirical contribution to the ongoing critical discussion about traditional neuropsychological tests and their ecological validity ("real-world or lab" dilemma) [5]. *Basic Interactive Stroop Test* (measuring errors in the color name and interference sub-trials) appears comparable in the two versions. Subjects' performances are ranked in the same quartile in both traditional settings and iMSE, meaning subjects made the same (or very close) number of errors regardless of interaction modality or environment. Thus, this result points to the assumption that both embodied and the classical version tap on the same cognitive functions. Moreover, the role of external stimuli, such as the presence of multi-channel stimulation (e.g., lights and sounds), did not act as a facilitator or obstacle for subjects in performing the tasks. These data confirm that the classic pencil-and-paper Stroop Test is predictive of subjects' performances to assess selective and divided attention.

4 CONCLUSIONS AND FUTURE WORK

We have presented an iMSEs activity of *NEP (Neuro-Psychological Test Suite)*, inspired by a traditional standardized neuropsychological test, the Stroop Test, acknowledged to be an effective tool to measure attention. With our work, we would like to offer two main contributions. The first is technological as, to the best of our knowledge, this is the first attempt to create a suite of neuropsychological tests for iMSEs. The second is to bring interesting insight into the ongoing debate over the "real-world or lab" dilemma. The findings of our study dis-confirm our hypothesis and provide evidence of similarity in scores between the traditional test and the NEP-based version. The pilot study has highlighted that modifying the interaction mechanism in the Stroop Test does not affect the test scores in healthy adults. iMSEs are invaluable tools for creating an environment to study and manipulate the effect of more "ecological" stimuli that resemble those present in everyday life. However, as far as the Stroop Test is concerned, the role of the "ecological" transposition, the presence of daily and familiar stimuli, and the immersive environment did not have any measurable effect on the subjects' performance. Still, the NEP battery needs more in-depth investigation. Further studies are required to confirm and corroborate our results and investigate how elderly with dementia or Alzheimer or subjects with NDD perform in the NEP activities.

We aim to continue the study and collect more data for factor analysis on all variables involved, and we are planning to investigate the relationship between user experience and test performance.

REFERENCES

[1] Christopher M Aanstoos. 1991. Experimental psychology and the challenge of real life. (1991).

[2] Erin D Bigler. 2012. Symptom validity testing, effort, and neuropsychological assessment. *Journal of the International Neuropsychological Society* 18, 4 (2012).

[3] Ian Burns, Helen Cox, and Helen Plant. 2000. Leisure or therapeutics? Snoezelen and the care of older persons with dementia. *International Journal of Nursing Practice* 6, 3 (2000), 118–126.

[4] Sally Wai-chi Chan, David R Thompson, Janita PC Chau, Wilson WS Tam, Ivy WS Chiu, and Susanne HS Lo. 2010. The effects of multisensory therapy on behaviour of adult clients with developmental disabilities—A systematic review. *International Journal of Nursing Studies* 47, 1 (2010), 108–122.

[5] Naomi Chaytor and Maureen Schmitter-Edgecombe. 2003. The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology review* 13, 4 (2003), 181–197.

[6] Antonio Di Carlo, Marzia Baldereschi, Luigi Amaducci, Stefania Maggi, Francesco Grigoletto, Guglielmo Scarlato, Domenico Inzitari, and Italian Longitudinal Study on Aging Working Group. 2000. Cognitive impairment without dementia in older people: prevalence, vascular risk factors, impact on disability. The Italian Longitudinal Study on Aging. *Journal of the American Geriatrics Society* 48, 7 (2000), 775–782.

[7] Erika M Ebly, David B Hogan, and Irma M Parhad. 1995. Cognitive impairment in the nondemented elderly: results from the Canadian Study of Health and Aging. *Archives of neurology* 52, 6 (1995), 612–619.

[8] Franca Garzotto, Eleonora Beccaluva, Mattia Gianotti, and Fabiano Riccardi. 2020. *Interactive Multisensory Environments for Primary School Children.* Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376343

[9] Franca Garzotto and Mirko Gelsomini. 2018. Magic Room: A Smart Space for Children with Neurodevelopmental Disorder. *IEEE Pervasive Computing* 17, 1 (Jan. 2018), 38–48. https://doi.org/10.1109/MPRV.2018.011591060

[10] Charles J Golden and Shawna M Freshwater. 1978. Stroop color and word test. (1978).

[11] Mylène Henry, Christian C Joyal, and Pierre Nolin. 2012. Development and initial assessment of a new paradigm for assessing cognitive and motor inhibition: the bimodal virtual-reality Stroop. *Journal of neuroscience methods* 210, 2 (2012).

[12] Gijs A Holleman, Ignace TC Hooge, Chantal Kemner, and Roy S Hessels. 2020. The 'real-world approach'and its problems: A critique of the term ecological validity. *Frontiers in Psychology* 11 (2020), 721.

[13] Robert L. Kane and Gary G. Kay. 1992. Computerized assessment in neuropsychology: A review of tests and test batteries. *Neuropsychology Review* 3, 1 (March 1992), 1–117. https://doi.org/10.1007/BF01108787

[14] Roy PC Kessels, Martine JE Van Zandvoort, Albert Postma, L Jaap Kappelle, and Edward HF De Haan. 2000. The Corsi block-tapping task: standardization and normative data. *Applied neuropsychology* 7, 4 (2000), 252–258.

[15] G. E. Lancioni, A. J. Cuvo, and M. F. O'Reilly. 2002. Snoezelen: an overview of research with people with developmental disabilities and dementia. *Disability and Rehabilitation* 24, 4 (Jan. 2002), 175–184. https://doi.org/10.1080/09638280110074911

[16] Muriel Deutsch Lezak, Diane B Howieson, David W Loring, Jill S Fischer, et al. 2004. *Neuropsychological assessment.* Oxford University Press, USA.

[17] Rebecca G Logsdon, Laura E Gibbons, Susan M McCurry, and Linda Teri. 2002. Assessing quality of life in older adults with cognitive impairment. *Psychosomatic medicine* 64, 3 (2002), 510–519.

[18] Donlin M Long. 2003. Perceptual Learning.

[19] Peter McGeorge, Louise H Phillips, John R Crawford, Sharin E Garden, Sergio Della Sala, Alan B Milne, Steven Hamilton, and John S Callender. 2001. Using virtual environments in the assessment of executive dysfunction. *Presence* 10, 4 (2001), 375–383.

[20] Marco Monaco, Alberto Costa, Carlo Caltagirone, and Giovanni Augusto Carlesimo. 2013. Forward and backward span for verbal and visuo-spatial data: standardization and normative data from an Italian adult population. *Neurological Sciences* 34, 5 (May 2013), 749–754. https://doi.org/10.1007/s10072-012-1130-x

[21] Thomas D Parsons. 2015. Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in human neuroscience* 9 (2015), 660.

[22] Anaick Perrochon and Gilles Kemoun. 2014. The Walking Trail-Making Test is an early detection tool for mild cognitive impairment. *Clinical Interventions in Aging* (Jan. 2014), 111. https://doi.org/10.2147/CIA.S53645

[23] Anaick Perrochon, Gilles Kemoun, Watelain, and Berthoz. 2013. Walking Stroop carpet: an innovative dual-task concept for detecting cognitive impairment. *Clinical Interventions in Aging* (March 2013), 317. https://doi.org/10.2147/CIA.S38667

[24] Ronald C Petersen, Rachelle Doody, Alexander Kurz, Richard C Mohs, John C Morris, Peter V Rabins, Karen Ritchie, Martin Rossor, Leon Thal, and Bengt Winblad. 2001. Current concepts in mild cognitive impairment. *Archives of neurology* 58, 12 (2001), 1985–1992.

[25] Ronald C Petersen, Glenn E Smith, Stephen C Waring, Robert J Ivnik, Eric G Tangalos, and Emre Kokmen. 1999. Mild cognitive impairment: clinical characterization and outcome. *Archives of neurology* 56, 3 (1999), 303–308.

[26] Luigi Pugnetti, Laura Mendozzi, Achille Motta, Annamaria Cattaneo, Elena Barbieri, and Aaron Brancotti. 1995. Evaluation and retraining of adults' cognitive impairments: Which role for virtual reality technology? *Computers in Biology and Medicine* 25, 2 (1995), 213–227.

[27] Ralph M Reitan. 1958. Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and motor skills* 8, 3 (1958), 271–276.

[28] Murat Doğan Şahin and Eren Can Aybek. 2019. Jamovi: an easy to use statistical software for the social scientists. *International Journal of Assessment Tools in Education* 6, 4 (2019), 670–692.

[29] Federica Scarpina and Sofia Tagini. 2017. The stroop color and word test. *Frontiers in psychology* 8 (2017), 557.

[30] Silvia Serino and Claudia Repetto. 2018. New Trends in Episodic Memory Assessment: Immersive 360° Ecological Videos. *Frontiers in Psychology* 9 (2018), 1878. https://doi.org/10.3389/fpsyg.2018.01878

[31] TIM Shallice and Paul W Burgess. 1991. Deficits in strategy application following frontal lobe damage in man. *Brain* 114, 2 (1991), 727–741.

[32] Donna M Spooner and Nancy A Pachana. 2006. Ecological validity in neuropsychological assessment: A case for greater consideration in research with neurologically intact populations. *Archives of clinical neuropsychology* 21, 4 (2006).

[33] Sonia Valladares-Rodríguez, Roberto Pérez-Rodríguez, Luis Anido-Rifón, and Manuel Fernández-Iglesias. 2016. Trends on the application of serious games to neuropsychological evaluation: A scoping review. *Journal of Biomedical Informatics* 64 (Dec. 2016), 296–319. https://doi.org/10.1016/j.jbi.2016.10.019

[34] Barbara A Wilson. 1993. Ecological validity of neuropsychological assessment: Do neuropsychological indexes predict performance in everyday activities? *Applied and Preventive Psychology* 2, 4 (1993), 209–215.

[35] Margaret Wilson. 2002. Six views of embodied cognition. *Psychonomic bulletin & review* 9, 4 (2002), 625–636.