

# Comparing adaptive cognitive training in virtual reality and paper-and-pencil in a sample of stroke patients

Ana Lúcia Faria †  
Madeira Interactive Technologies  
Institute da Universidade da Madeira  
Faculdade de Psicologia e de Ciências  
da Educação da Universidade de  
Coimbra  
Funchal, Portugal  
ana.faria@m-iti.org

Teresa Paulino †  
Madeira Interactive Technologies  
Institute  
Faculdade de Ciências Exactas e da  
Engenharia  
Universidade da Madeira  
Funchal, Portugal  
teresa.paulino@m-iti.org

Sergi Bermúdez i Badia  
Madeira Interactive Technologies  
Institute  
Faculdade de Ciências Exactas e da  
Engenharia  
Universidade da Madeira  
Funchal, Portugal  
sergi.bermudez@m-iti.org

**Abstract**— The growing number of people with cognitive deficits creates an urgent need for new cognitive training solutions. Paper-and-pencil tasks are still widely used for cognitive rehabilitation despite the proliferation of new computer-based methods, like VR-based simulations of ADL's. The health professionals' resistance in adopting new tools might be explained by the small number of validation trials. Studies have established construct validity of VR assessment tools with their paper-and-pencil version by demonstrating significant associations with their traditional construct-driven measures. However, adaptive rehabilitation tools for intervention are mostly not equivalent to their counterpart paper-and-pencil versions, which makes it difficult to carry out comparative studies. Here we present a 12-session intervention study with 31 stroke survivors who underwent different rehabilitation protocols based on the same content and difficulty adaptation progression framework: 17 performed paper-and-pencil training with the Task Generator and 14 performed VR-based training with the Reh@City. Results have shown that both groups performed at the same level and there was not an effect of the training methodology in overall performance. However, the Reh@City enabled more intensive training, which may translate in more cognitive improvements.

**Keywords**—cognitive rehabilitation, paper-and-pencil, virtual reality, personalization.

## I. INTRODUCTION

Cognitive deficits affect a person's capability to live independently and are present in 3-19% of people older than 65 years [1]. Between 2015 and 2050, the proportion of the world's older adults is estimated to almost double from about 12% to 22% [2]. This will raise the numbers of age-associated diseases, like stroke and dementia, which already have 15 and 50 million new cases every year, respectively [3], [4]. These facts created an urgent need for intensive and personalized cognitive training solutions to maximize neural plasticity and, consequently, improve functional independence [5].

Cognitive exercises, including computer-based programs, have been used to improve specific neuropsychological processes, predominantly attention, memory, and executive functions [6], [7]. Despite many descriptions of particular programs and interventions, limited data on the effectiveness of cognitive rehabilitation is available because of the heterogeneity of participants, interventions, and outcomes [8]. In what concerns interventions, still today, paper-and-pencil tasks are the most commonly used methods to train cognitive functions in clinical settings [9]. Although with

established clinical validity and reduced cost [10], paper-and-pencil methods are mostly planned and delivered based on the clinician experience and lack a solid theoretical framework for intervention personalization [11]. Additionally, rehabilitation with these tasks has shown to have a limited transfer to performance in activities of daily living (ADL) [9].

Over the last years, rehabilitation tools based on virtual reality (VR) have been developed and validated as promising solutions to improve cognitive functions [12], [13]. VR-based methods have shown potential to be ideal environments to incorporate cognitive tasks within the simulation of ADL's [14], [15], [16] offering immersive and ecologically valid experiences capable of promoting enjoyment and adherence [20]. However, there is still an insufficient number of trials to clinically validate these methods [19], which together with the difficulties in adopting new technologies [20], limits the acceptance of these methods by health professionals who choose to continue performing paper-and-pencil interventions.

Regardless of the purported advantages of virtual environments, there are several critical areas that require further development. One area of note is the need to bridge widely accepted paper-and-pencil methodologies with VR-based ADL's simulations. In the field of cognitive assessment, a considerable number of studies have compared VR neuropsychological assessment tools with their paper-and-pencil original versions [25]. Raspelli and colleagues (2012) evaluated a virtual version of the Multiple Errands Test (MET), the Virtual Multiple Errands Test (VMET), with the purpose of establishing ecological and construct validity as an assessment tool for executive functions. The MET 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 post-stroke participants and healthy adults. Correlations between VMET variables and some traditional executive functions paper-and-pencil measures provided preliminary support for its ecological and construct validity [24]. Nir-Hadad and colleagues (2015) examined the discriminant, construct-convergent and ecological validity of the Adapted Four-Item Shopping Task, an assessment of the shopping Instrumental Activity of Daily Living (IADL). Stroke and healthy participants performed the shopping task in both the Virtual Interactive Shopping environment and a real shopping environment. The shopping task outcomes were compared to paper-and-pencil measures of executive functions. The findings provided initial support for the

† Contributed equally.

validity of the Adapted Four-Item Shopping Task as an IADL assessment that requires the use of executive functions for people with stroke [25]. Vourvopoulos and colleagues (2014) developed a VR-based simulation of activities of daily living (ADLs) within a city where stroke participants had to accomplish several goals; and found a strong correlation between the VR score and the Mini-Mental State Examination cognitive screening test for clinical assessment of cognitive function in several domains [26]. Parsons and colleagues (2018) compared the performance of healthy participants on a virtual apartment-based Stroop with traditional (multi-item) and computerized (single item) modalities. Results suggested the potential of the Virtual Apartment Stroop task to distinguish between prepotent response inhibition and resistance to distractor inhibition in young adults [23]. Costa and colleagues (2018) compared the performance obtained on the assessment of perception of spatial abilities in an immersive VR spatial task and its correspondent paper-and-pencil version and found that the VR task is ecologically more valid since it is closer to real life [27].

The previously referred studies lead us to conclude that researchers of VR-based cognitive assessments have sought to establish construct validity by demonstrating significant associations between construct-driven virtual environments with other traditional construct-driven measures [28]. So, how can we validate VR-based cognitive rehabilitation? There is a rising number of VR-based rehabilitation tools, most of them incorporating personalization and adaptation, but no study has explored associations between adaptation in traditional paper-and-pencil and VR training. This comparison is challenging since interventions are planned and delivered according to health professionals' clinical experience, which involves a large variety of paper-and-pencil tasks with different difficulty levels and customizable or adaptive VR systems. One solution would be to have an objective difficulty adaptation framework to be applied in a set of paper-and-pencil training tasks and then compare it with content equivalent VR-based tasks using the same difficulty adaptation framework. This comparison would allow identifying the specific contributions of VR over clinically accepted paper-and-pencil, which could promote the adoption of VR technologies by health professionals.

Based on the NeuroRehabLab's cognitive rehabilitation adaptation framework, which established quantitative and task-specific guidelines to personalize training difficulty to the patient profile [29], we developed two content equivalent tools and clinically validated them with stroke patients: a web-based paper-and-pencil Task Generator [30] and a VR-based simulation of different ADL's, the Reh@City [26], [16]. To compare these two rehabilitation methods, we performed an intervention study with stroke survivors in order to explore three main questions:

- a) Are paper-and-pencil and VR training performances equivalent?
- b) Is performance modulated by the difficulty adaptation or the training methodology?
- c) Which training method is more intensive?

## II. METHODS

### A. Participants

Participants were selected in the Physical Medicine and Rehabilitation department from the Madeira Health Service (Portugal). In total, we have selected 35 outpatients based on the following inclusion criteria: no more than 75 years old; first ischemic stroke episode and at least at 6 months post-stroke (chronic phase); self-reported cognitive complaints; no hemi-spatial neglect as assessed by the clinicians with the Line Bisection test [31]; capacity to be seated and ability to read and write. The study was approved by the Madeira Health Service Ethical Committee (reference number: 13/2016), and all the participants gave informed consent before participation.

Table I presents the mean values (standard deviations) of the demographic characteristics (age, gender, education) and clinical information (stroke type and location and time post-stroke) for the two groups. No differences between groups were found with the Mann-Whitney test.

TABLE I. DEMOGRAPHICS CHARACTERISTICS OF THE PARTICIPANTS.

	Reh@City (N=14)	Task Generator (N=17)	MW	p.
Age (years)	59.14 ± 11.81	65.00 ± 6.20	83.500	.107
Gender (M/F)	5/9	11/7	94.000	.235
Education (years)	8.00 ± 5.32	5.50 ± 3.15	100.500	.338
Stroke type (I/H/NS)	12/2/0	14/3/1	115.000	.694
Stroke localization (R/L/NS)	11/3/0	9/6/3	85.500	.125
Months post- stroke	45.93 ± 43.56	21.33 ± 12.88	89.500	.168

### B. Protocol

An intervention study was performed between June 2016 and January 2019. A total of 35 stroke survivors met the eligibility criteria and had shown motivation to participate. Intervention allocation was made through randomization of the Madeira island counties: participants from Porto Moniz, Calheta, Ribeira Brava, Santana, Câmara de Lobos, and west Funchal would perform the paper-and-pencil intervention and; participants from São Vicente, Ponta do Sol, Santa Cruz, Machico and east Funchal would perform the VR intervention. When recruitment stopped, there were 18 participants in the paper-and-pencil group, and 17 in the VR one (Fig. 1).

All participants were assessed through the Montreal Cognitive Assessment (MoCA) [32] before and after the intervention. Each participant went through a set of 12 30-minute sessions with a frequency of 3 per week. On each session, the participant was assigned a set of cognitive tasks individually personalized according to the participant cognitive levels previously assessed through the MoCA. There was no predefined number of tasks that the participant had to complete; tasks were performed on each session at the participant's own pace. On both groups, the intervention consisted of fulfilling tasks, and at the end of each set, the difficulty level for the following set of tasks was calculated based on the participant's performance. If the user obtained an average performance lower than 50%, the difficulty was

reduced in 0.5 points (out of 10), if higher than 70%, the difficulty was increased in the same amount; otherwise, the difficulty value remained the same. This difficulty parameter is further explained in the next subsection.

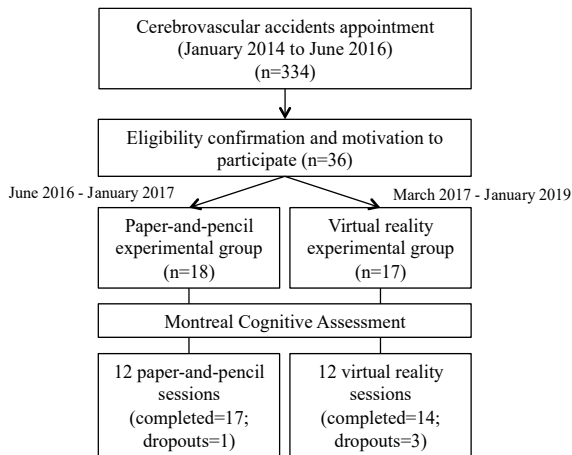


Fig. 1. Diagram showing an overview of the intervention study.

### C. Tools

Two tools were developed to create personalized cognitive rehabilitation: a web tool that generates paper-and-pencil tasks named the Task Generator (TG) [30] and the Reh@City (RC) [16], a virtual reality system that integrates the simulation of tasks based on ADL's on a virtual environment. The two tools are described as follows:

**Task Generator:** The TG is a web-based tool that allows the automatic generation of paper-and-pencil cognitive tasks tailored for each user profile. The TG consists of 11 cognitive tasks: cancellation (example on Fig. 2), numeric sequences, problem resolution, association, comprehension of contexts, image pairs, word search, mazes, categorization, action sequencing, and memory of stories. A brief description of each task can be found in Table II(A). The personalization of tasks depends on the user levels of the following cognitive domains: attention, memory, executive function, and language. These levels are found through the MoCA with values varying between 1 and 10 with 0.5 intervals, where 10 represents the highest value that is possible to score. For instance, the maximum value that is possible to achieve on the attention domain of the MoCA is 6; this result is then normalized to the TG scale, corresponding to the maximum value of 10.

The process is similar for the remaining domains: memory, executive function, and language, which can hold the maximum values of 11, 7, and 6, respectively. The user profile levels are manually set using sliders for each domain. One additional parameter, the difficulty, is used to adjust the cognitive tasks based on user performance. The initial value of the difficulty is found by normalizing the MoCA's total score to the 10-point TG scale. This value varies over sessions based on the average scores obtained on each set of tasks, which needs to be manually calculated.

**Reh@City:** The RC system is a virtual environment that consists of a city with different locations where the user can perform cognitive tasks equivalent to the ones generated by the TG [33]. Like in the TG, tasks are personalized for each user profile based on the MoCA's results of the same

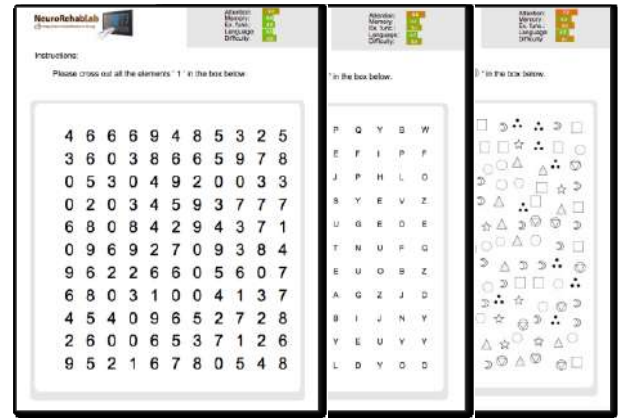


Fig. 2. TG Cancellation tasks with different difficulty adaptations.

cognitive domains: attention, memory, executive function, and language. MoCA's results are also normalized to the same scale used in the TG (from 1 to 10, with 0.5 intervals). The RC performs the normalization process automatically without the need for manual calculation. The RC system is also able to save the user levels avoiding the need for configuration on each intervention session. The cognitive tasks are spread over eight locations in the city: bank, clothes shop, home, kiosk, park, supermarket, post office, and pharmacy. The RC tasks were ideated to be as equivalent as possible to the ones found in the TG. Eight tasks have been implemented, and a brief description of each one can be found in Table II(B). Tasks are presented as requests that the participant needs to fulfill. Each task starts by navigating to a specific location and then executing the task on that location. The process is repeated until all expected locations to complete tasks have been visited. RC automatically adjusts each set of tasks based on user performance using the principles mentioned in the protocol subsection of this paper. Each set consists of 7 tasks plus the navigation in the city, which is equivalent to the mazes task in the TG.

Contrary to the TG, the RC allows evolving to difficulty levels above the maximum value of the scale, which is 10. In these circumstances, removing "helpers" that assist the participant when performing the tasks increases the difficulty. For instance, when reaching level 10, the task request is only visible for a few seconds, the participant needs to memorize what task needs to fulfill. At level 10.5 a mini-map that enables to have a broader overview of the path is no longer visible, and at level 11, city signs that indicate directions to the locations are removed. Fig. 3 indicates the helpers that have been mentioned.

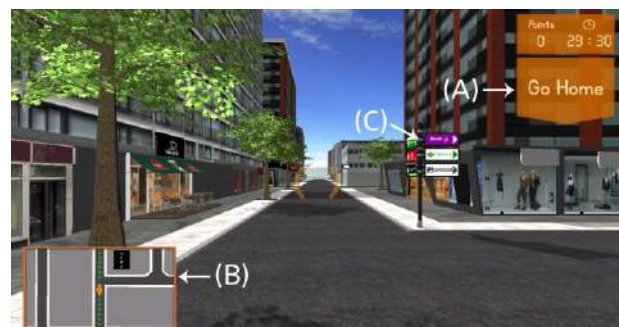


Fig. 3. RC street showing helpers that are removed after difficulty level 10: (A) Task request text, (B) mini-map, and (C) city signs.

For instance, the cancellation task, in the TG consists of a set of numbers, letters or symbols where the participant is required to circle or cross a specific given item. The number of items and if they are organized or not are set by the difficulty level. Fig. 2 shows some examples of paper-and-pencil tasks generated by the TG tool. In RC, the cancellation task can be found in two different locations: the pharmacy and the post office. A set of shelves with products are presented, and similarly, the participant is required to find one or more items. Items are randomly displayed on the shelves; the number of items is also set by the difficulty level (Fig. 4).



Fig. 4. Cancellation task at the pharmacy location in RC.

TABLE II. COGNITIVE TASKS DESCRIPTION OF EACH TOOL: (A) TASK GENERATOR AND (B) REH@CITY

Task	(A) Task Generator	(B) Reh@City
<b>Cancellation</b>	Find a specific letter, number, or symbol on an assorted sheet	Find specific items at the pharmacy or post office
<b>Numeric Sequences</b>	Fill in the missing numbers on numeric sequences	Fill in the missing numbers on a numeric sequence at the ATM
<b>Problem resolution</b>	Solve mathematical calculations	Choose the correct invoice after shopping at the supermarket
<b>Association</b>	Match related image pairs	Not applicable
<b>Comprehension of contexts</b>	Mark true or false on affirmations concerning a given contextual image	Not applicable
<b>Image pairs</b>	Memorize a set of image pairs and recall each pair when not visible	Find matching cards in a memory game at the park
<b>Word search</b>	Find words on a sheet of assorted letters	Not applicable
<b>Mazes</b>	Find the correct path from the entry to the exit	Navigate in the city through the shortest path until finding a given location
<b>Categorization</b>	Name the category of each image on a set of images	Select items of a given category at the clothes shop
<b>Action sequencing</b>	Order a set of actions in a manner that makes sense	Select the steps in the right order to accomplish a given task at home
<b>Memory of stories</b>	Read a story and then afterward answer a set of questions concerning the story	Read the text of a newspaper at the kiosk and then answer a set of questions when reaching the next location

#### D. Data Analysis

Data from the TG was manually inserted in table sheets with the information of each session per participant including the difficulty level, and the percentage of performance obtained on each task. The RC automatically generated log files in CSV and XML formats, which enabled easy importing into Excel table sheets. RC creates two types of files, one ready for analysis with data summaries, and a highly detailed type of log files with data saved at the software frame rate (mostly 30FPS). These data had to go through a manual verification and rectification process because, during the intervention, the software crashed a few times. After this initial process, the means per participant considering the performance obtained on each task, on each set of tasks, in all intervention tasks (overall performance), the number of tasks performed, and the difficulty level evolution over sessions was computed for both groups. To compare the intensity of training in both methodologies, we used the number of tasks sets, the highest difficulty level achieved, and the total number of tasks performed by each participant. All statistical analyses were performed using SPSS software (version 20, SPSS Inc., Chicago IL, USA). As a criterion for significance, we used an  $\alpha$  of 0.05. Normality of data was assessed with the Kolmogorov-Smirnov (KS) test. As some data were not normally distributed, the Mann-Whitney (MW) test was used to compare the between-group differences from baseline to the end of the study. To analyze the effect of the difficulty adaptation and the training methodology in performance we did a general linear model univariate analysis. Since we only had homogeneity of variances, as assessed by the Levene's test, for the overall performance and not for all training tasks, we did not perform a general linear model multivariate analysis to analyze each task performance separately. Instead, we have compared them through a non-parametric Mann-Whitney analysis.

### III. RESULTS

#### a) Are paper-and-pencil and VR training performances equivalent?

Regarding the twelve training sessions overall performance, no significant statistical differences were found ( $U=91.000$ ,  $Z=-1.111$ ,  $p=.266$ ) between the two groups (TG:  $Mdn=79.91$ ,  $IQR=72.11-86.12$ ; RC:  $Mdn=77.63$ ,  $IQR=71.09-81.47$ ) (Fig. 5).

#### b) Is performance modulated by the difficulty adaptation or the training methodology?

A Levene's test showed that the variances for performance ( $F(1,204)=.838$ ,  $p=.361$ ) and difficulty adaptation ( $F(1,204)=.113$ ,  $p=.737$ ) were equal between groups. After complying with the homogeneity assumption, we performed a general linear model univariate analysis. According to the obtained results the main effect of the adapted difficulty was significant ( $F(10)=1.992$ ,  $p=.036$ ) but not the training methodology (TG versus RC) ( $F(1)=.079$ ,  $p=.779$ ). The interaction of these two factors was also not significant,  $F(10,1)=.621$ ,  $MS=109.710$ ,  $p=.795$ .

Concerning performance by each task separately, no significant differences were found for the Numeric Sequences, Mazes, Categorization, and Memory of Stories.

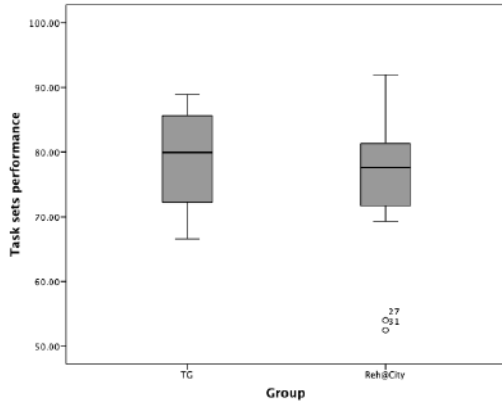


Fig. 5. Comparison of the overall task performance per training modality.

However, the Mann-Whitney test indicated that the participants who went through the TG intervention achieved a significantly higher performance in both Cancellation ( $U=66.000$ ,  $Z=-2.167$ ,  $p=.030$ ) and Action Sequencing ( $U=55.000$ ,  $Z=-2.555$ ,  $p=.011$ ) tasks. On the other hand, the participants who performed the RC training performed significantly better in the Problem Resolution ( $U=24.000$ ,  $Z=-3.773$ ,  $p<.001$ ) and Image Pairs tasks ( $U=69.000$ ,  $Z=-1.985$ ,  $p=.047$ ). In these four tasks where there were statistical differences (Table III), the TG group had higher performances for the Cancellation and Action Sequencing tasks while the RC group had higher performances for the Problem Resolution and Image Pairs task. So, there is no clear specific preference for any approach, which is consistent with the lack of significant differences between overall performances in both training methodologies.

TABLE III. RESULTS OF THE MANN-WHITNEY TEST ON MEDIAN SCORES OF PERFORMANCE OBTAINED ON EACH TASK (IQR BETWEEN BRACKETS).

	Task Generator	Reh@City	MW
Cancellation	99.75 (96.42-100)	89.44 (68.17-100)	.030*
Numeric Sequences	89.00 (77.89-93.38)	83.95 (61.50-90.48)	.204
Problem Resolution	44.45 (27.21-59.45)	81.53 (71.43-94.64)	<.000*
Association	100 (92.82-100)	Not applicable	Not applicable
Comprehension of Contexts	88.09 (82.43-93.40)	Not applicable	Not applicable
Image Pairs	54.67 (29.80-64.59)	66.34 (60.24-73.97)	.047*
Word Search	93.33 (88.25-99.00)	Not applicable	Not applicable
Mazes	83.33 (69.50-100)	88.59 (78.90-96.40)	.719
Categorization	92.85 (79.14-95.66)	86.02 (76.68-91.97)	.321
Action Sequencing	83.34 (59.92-100)	67.86 (35.42-73.56)	.011*
Memory of Stories	72.50 (59.49-83.96)	73.96 (61.67-85.73)	.648

\* Statistically significant difference.

Note: Association, Comprehension of Contexts and Word Search tasks are only available in the TG tool.

### c) Which training modality is more intensive?

Both groups evolved in difficulty level in similar ways. However, the RC group evolved slightly faster and attained higher levels (Fig. 6).

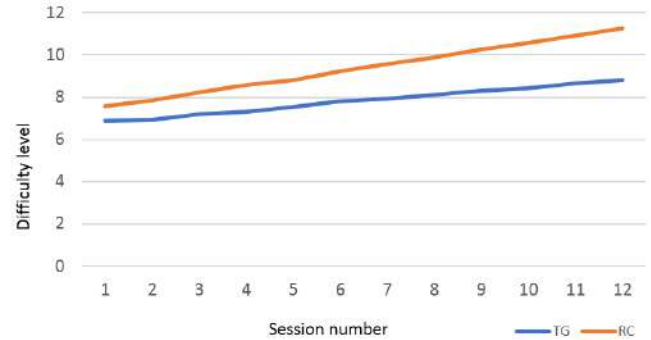


Fig. 6. Comparison of the means of difficulty evolution over the 12 sessions of the TG and the RC.

As described before, RC allows progression to higher levels than 10 while in the TG, even if the participant could further progress, there is a ceiling effect and training can only be generated until the maximum level of 10. However, only two out of seventeen participants of the TG group managed to reach the maximum difficulty and were limited in progressing in the last training sessions. While in the RC group, eleven out of fourteen managed to surpass the difficulty level of 10.

According to Table IV, the three task performance parameters (number of sets, last set difficulty, the total number of tasks) are higher in the VR training modality and significantly different from the paper-and-pencil one. The RC allows solving more tasks and subsequently completing more task sets, which results in a more progressive difficulty adaptation for the same amount of time.

TABLE IV. RESULTS OF MANN-WHITNEY TEST ON MEDIAN SCORES OF TASK PERFORMANCE PARAMETERS (IQR BETWEEN BRACKETS)

	Task Generator	Reh@City	MW
Number of tasks' sets	5 (5-6)	10 (9-13.50)	<.001*
Last set difficulty	9 (7.75-9.75)	11.50 (10.25-12.13)	.001*
Total number of tasks	54 (49.5-64.5)	82.5 (68.75)	.001*

\* Statistically significant difference.

## IV. DISCUSSION

Paper-and-pencil tasks are still widely used in cognitive rehabilitation despite the proliferation of computer-assisted [7], [6] and other multimedia methods, like VR-based simulations of ADL's [14], [15], [16]. This might be due to a reduced number of trials validate these methods [19] and by the difficulties in new technologies adoption by health professionals [20]. A significant number of studies has established construct validity of VR neuropsychological assessment tools with their paper-and-pencil version [25] by demonstrating significant associations with their traditional construct-driven measures [28]. Unfortunately, the clinical validation of a VR-based rehabilitation system is limited if only assessed by baseline and post-intervention neuropsychological outcome measures, which are paper-and-

pencil based and lack ecological validity [28]. Hence, the specific role of adaptation in VR and paper-and-pencil cognitive rehabilitation remains unexplored. To our knowledge, no other study exists comparing VR and with widely accepted paper-and-pencil in an adaptive rehabilitation protocol.

Here we presented a comparison of two content equivalent rehabilitation methodologies based on the same difficulty adaptation framework [29]: a web-based paper-and-pencil Task Generator and a VR-based simulation of different ADL's, the RC. The original paper-and-pencil tasks inspired tasks in RC. Sometimes, enhancing the ecological validity of some tasks in VR required adjusting some elements according to what was possible. For instance, the number of targets and distractors in a cancellation task. However, the training adaptation was implemented in the very same way and using the same computing models and difficulty progression rules. A twelve-session intervention study with stroke patients has led us to three main conclusions concerning the content equivalence of both training modalities.

First, according to the overall training performance comparison and despite the differences in training, the personalization and adaptation framework used led to similar cognitive training performances. Hence, there were no differences between groups, meaning that both tools delivered adaptive content of equivalent difficulty.

Second, we wanted to understand if performance was modulated by the implemented adaptation of task difficulty or by the training technology used, being it paper-and-pencil or VR. Our results show a significant effect on the performance of the difficulty adaptation but not of the training methodology, which further strengthens the equivalence of both training methodologies. By specifically comparing both groups' individual tasks performance, we have found significant differences that were consistent with the implementation adjustments we had to perform in VR. For instance, the performance obtained in the Cancellation task was significantly higher in the TG group, this may be due to the number of elements and targets which was much reduced in the RC. To illustrate this discrepancy, the same task with the same level of difficulty, in the RC could have only one target among 20 distractors. While in the paper-and-pencil task would have 15 targets among 120 distractors, by failing to find the correct target in the RC would lead to 0% performance, while by failing one in the TG task would not translate in the same percentage in performance. In the Problem Resolution task, the RC group obtained significantly higher performance. This is due to the number of calculations to solve equivalent tasks, which is also reduced in the RC. The Image Pairs task performance was also significantly higher in the RC group, this may be due to the task itself, which is slightly different in each method: in the TG, the participant was required to remember the pairs of unrelated images, and on the RC, is required to find identic pairs of images in a game. Finally, in what concerns the Action Sequencing task, the TG group performed at a higher level because there was no strict rule in how to order the actions, as long as the ordering was logic, as opposed to RC, where tasks were required to be selected in a specific programmed order. If one step failed, it counted immediately to the overall task performance. However, despite these

implementation differences, they did not have a statistical impact on overall performance.

Third and last, our findings, concerning the training intensity of each methodology, show that the VR-based group performed a larger number of tasks and therefore also more task sets and finished at higher difficulty levels. This could lead us to conclude that VR allows a more effective training by enabling more repetitions in the same amount of time, turning the training more intensive. However, these results can also be interpreted by the implementation' discrepancies. Not all the TG tasks have been implemented in the RC, the result in terms of the number of sets was expected since the participant had three tasks less to accomplish to complete each set. Regarding the higher level of difficulty attained by the RC group, this can be influenced by multiple factors such as an easier interaction in VR compared to paper-and-pencil, motivational factors of gaming in VR, computer automation of task delivery and also by embedding tasks in ecologically meaningful contexts.

## V. CONCLUSION

The presented study, besides its limitations, is the first to compare adaptive paper-and-pencil training with content equivalent VR-based ADL's simulation, by using the same personalization and difficulty adaptation framework within a longitudinal clinical intervention.

Findings of this study support that despite the necessary differences in task implementations, both groups performed at the same level and there was not an effect of the training methodology in overall performance. Moreover, our results contribute with new evidence and provide a further understanding of the impact of using adaptation in VR simulations of ADL's in the rehabilitation of cognitive deficits, instead of paper-and-pencil. Although there are not established clinically important differences for cognitive assessment and rehabilitation outcome measures, we can conclude that the RC offered a more intensive training leading to more task repetitions and higher difficulty adaptation progression, which we believe can be translated in more cognitive improvements.

Nevertheless, there is still a need for further research considering larger samples and more comparative studies with other cognitive rehabilitation tasks.

## ACKNOWLEDGMENTS

This study was supported by the European Commission through the RehabNet project (Neuroscience-Based Interactive Systems for Motor Rehabilitation) under grant 303891 RehabNet FP7-PEOPLE-2011-CIG, by LARSyS UID/EEA/50009/2019, MACBIOIDI (INTERREG program MAC/1.1.b/098), by FCT through the BRaNT project (PTDC/CCI-COM/31046/2017), and by the Agência Regional para o Desenvolvimento da Investigação, Tecnologia e Inovação.

The authors would like to thank Joana Alegria and Carolina Jorge for the collaboration in data collection.

## REFERENCES

- [1] W. H. Organization and others, "The World Health Report 2008: Primary health care (now more than ever)," 2014.
- [2] "Mental health of older adults." [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/mental-health-of-older-adults>. [Accessed: 30-Jan-2019].
- [3] J. MacKay and G. A. Mensah, *The Atlas of Heart Disease and Stroke*. World Health Organization, 2004.
- [4] "Dementia." [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/dementia>. [Accessed: 30-Jan-2019].
- [5] B. A. Wilson, "Compensating for Cognitive Deficits Following Brain Injury," *Neuropsychol. Rev.*, vol. 10, no. 4, pp. 233–243, Dec. 2000.
- [6] J. Solana *et al.*, "Improving Brain Injury Cognitive Rehabilitation by Personalized Telerehabilitation Services: Guttman Neuropersonal Trainer," *Biomed. Health Inform. IEEE J. Of*, vol. 19, no. 1, pp. 124–131, 2015.
- [7] V. Tedim Cruz *et al.*, "Implementation and Outcomes of a Collaborative Multi-Center Network Aimed at Web-Based Cognitive Training – COGWEB Network," *JMIR Ment. Health*, vol. 1, no. 1, Nov. 2014.
- [8] M. M. Sohlberg and C. A. Mateer, *Cognitive Rehabilitation: An Integrative Neuropsychological Approach*. Guilford Publications, 2017.
- [9] T. D. Parsons, "Neuropsychological Rehabilitation 3.0: State of the Science," in *Clinical Neuropsychology and Technology*, Springer International Publishing, 2016, pp. 113–132.
- [10] T. D. Parsons, "Ecological Validity in Virtual Reality-Based Neuropsychological Assessment," in *Encyclopedia of Information Science and Technology*, Hershey, PA: Information Science Reference., 2015, pp. 1006–1015.
- [11] F. Gracey and B. A. Wilson, "Theoretical approaches to cognitive rehabilitation," in *Clinical Neuropsychology: A Practical Guide to Assessment and Management for Clinicians*, L. H. Goldstein and J. E. McNeil, Eds. John Wiley & Sons, Ltd, 2013, pp. 463–466.
- [12] R. D. Luca *et al.*, "Effects of virtual reality-based training with BTs-Nirvana on functional recovery in stroke patients: preliminary considerations," *Int. J. Neurosci.*, vol. 128, no. 9, pp. 791–796, Sep. 2018.
- [13] M. G. Maggio *et al.*, "Virtual Reality and Cognitive Rehabilitation in People With Stroke: An Overview," *J. Neurosci. Nurs. J. Am. Assoc. Neurosci. Nurses*, Jan. 2019.
- [14] M. D. Navarro, R. Lloréns, E. Noé, J. Ferri, and M. Alcañiz, "Validation of a low-cost virtual reality system for training street-crossing. A comparative study in healthy, neglected and non-neglected stroke individuals," *Neuropsychol. Rehabil.*, vol. 23, no. 4, pp. 597–618, 2013.
- [15] N. Josman, R. Kizony, E. Hof, K. Goldenberg, P. L. Weiss, and E. Klinger, "Using the Virtual Action Planning-Supermarket for Evaluating Executive Functions in People with Stroke," *J. Stroke Cerebrovasc. Dis.*, vol. 23, no. 5, pp. 879–887, May 2014.
- [16] A. L. Faria, A. Andrade, L. Soares, and S. B. i Badia, "Benefits of virtual reality based cognitive rehabilitation through simulated activities of daily living: a randomized controlled trial with stroke patients," *J. NeuroEngineering Rehabil.*, vol. 13, p. 96, 2016.
- [17] "Cognitive assessment and rehabilitation in virtual reality: theoretical review and practical implications | Directory of Open Access Journals." [Online]. Available: <https://doaj.org/article/465e35cf44af41ea93345ea61a331961>. [Accessed: 23-Jan-2019].
- [18] "Virtual Reality as a Vehicle to Empower Motor-Cognitive Neurorehabilitation." [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6224455/>. [Accessed: 03-Dec-2018].
- [19] K. E. Laver, B. Lange, S. George, J. E. Deutsch, G. Saposnik, and M. Crotty, "Virtual reality for stroke rehabilitation," *Cochrane Database Syst. Rev.*, vol. 11, p. CD008349, Nov. 2017.
- [20] WHO, "WHO | Assistive devices/technologies: what WHO is doing," *WHO*, 2011. [Online]. Available: <http://www.who.int/disabilities/technology/activities/en/>. [Accessed: 09-May-2016].
- [21] T. Ogourtsova, P. Archambault, S. Sangani, and A. Lamontagne, "Ecological Virtual Reality Evaluation of Neglect Symptoms (EVENS): Effects of Virtual Scene Complexity in the Assessment of Poststroke Unilateral Spatial Neglect," *Neurorehabil. Neural Repair*, vol. 32, no. 1, pp. 46–61, Jan. 2018.
- [22] A. Neguț, S.-A. Matu, F. A. Sava, and D. David, "Task difficulty of virtual reality-based assessment tools compared to classical paper-and-pencil or computerized measures: A meta-analytic approach," *Comput. Hum. Behav.*, vol. 54, pp. 414–424, Jan. 2016.
- [23] T. D. Parsons and M. D. Barnett, "Virtual apartment stroop task: Comparison with computerized and traditional stroop tasks," *J. Neurosci. Methods*, vol. 309, pp. 35–40, Nov. 2018.
- [24] S. Raspelli *et al.*, "Validating the Neuro VR-based virtual version of the Multiple Errands Test: preliminary results," *Presence Teleoperators Virtual Environ.*, vol. 21, no. 1, pp. 31–42, 2012.
- [25] S. Y. Nir-Hadad, P. L. Weiss, A. Waizman, N. Schwartz, and R. Kizony, "A virtual shopping task for the assessment of executive functions: Validity for people with stroke," *Neuropsychol. Rehabil.*, vol. 0, no. 0, pp. 1–26, Nov. 2015.
- [26] A. Vourvopoulos, A. L. Faria, K. Ponnam, and S. Bermúdez i Badia, "RehabCity: Design and Validation of a Cognitive Assessment and Rehabilitation Tool through Gamified Simulations of Activities of Daily Living," in *11th International Conference on Advances in Computer Entertainment Technology*, Funchal, Portugal, 2014.
- [27] R. Q. M. da Costa *et al.*, "A Comparison Between an Immersive Virtual Reality Spatial Task and its Correspondent Paper-and-Pencil Version with one's Perception of Spatial Abilities," *Alzheimers Dement. J. Alzheimers Assoc.*, vol. 14, no. 7, p. P427, Jul. 2018.
- [28] T. D. Parsons, "Neuropsychological Assessment 3.0," in *Clinical Neuropsychology and Technology*, Springer, 2016, pp. 65–96.
- [29] A. L. Faria, M. S. Pinho, and S. B. i Badia, "Capturing Expert Knowledge for the Personalization of Cognitive Rehabilitation: Study Combining Computational Modeling and a Participatory Design Strategy," *JMIR Rehabil. Assist. Technol.*, vol. 5, no. 2, p. e10714, 2018.
- [30] A. L. Faria and S. Bermúdez i Badia, "Personalizing paper-and-pencil training for cognitive rehabilitation: a feasibility study with a web-based Task Generator," presented at the ICAP – International Conference on Applied Psychology and Human Behavior, Funchal, Portugal, 2018.
- [31] T. Schenkenberg, D. C. Bradford, and E. T. Ajax, "Line bisection and unilateral visual neglect in patients with neurologic impairment," *Neurology*, vol. 30, no. 5, pp. 509–509, May 1980.
- [32] Z. S. Nasreddine, I. Collin, H. Chertkow, N. Phillips, H. Bergman, and V. Whitehead, "Sensitivity and specificity of the Montreal Cognitive Assessment (MoCA) for detection of mild cognitive deficits," *Can J Neurol Sci*, vol. 30, no. 2, p. 30, 2003.
- [33] T. Paulino, A. Faria, and S. Bermudez i Badia, "Reh@City v2.0: a comprehensive virtual reality cognitive training system based on personalized and adaptive simulations of activities of daily living," presented at the Experiment@ International Conference 2019 (exp.at'19), Madeira Island, Portugal, 2019.