Are 3D virtual environments better than 2D interfaces in serious games performance? An explorative study for the assessment of executive functions.

Irene Alice Chicchi Giglioli[1*], Carla de Juan Ripoll[1], Elena Parra[1], and Mariano Alcañiz Raya[1]

[1] Instituto de Investigación e Innovación en Bioingeniería (I3B), Universitat Politècnica de València, València, 46002, Spain.

*Corresponding author

1

Abstract

Executive functions refer to higher-order cognitive processes that supervise and guide goal-directed and adaptive behaviors in response to everyday situations. The traditional measures used to assess executive functions include paper-and pencil tests and/or computerized tests that have been found to have a moderate level of ecological validity in predicting real-world performance. Serious games (SG) represent a novel methodological approach, allowing investigating subjects' performance in real-simulated situations. Serious games are computer games which primary purposes include investigating human behaviors and changes. Furthermore, SG can also vary according to the technology used and the interaction. Indeed, a SG can be rendered via a non-immersive screen-based (2D) or via an immersive virtual reality game (3D).

Starting from these premises, we compared a narrative-contextualized SG in 2D and 3D, correlating them with traditional tests. Findings showed different condition correlations with the traditional tasks and the comparison between the two systems have revealed that 3D is able to generate lower reaction times, higher correct answers, and lower perseverative responses in attentional abilities, inhibition control, and cognitive shifting than 2D condition. The present study yielded evidence on the use of more ecological tools to identify the functional cognitive status in real-simulated contexts along with traditional evaluation.

Keywords

1. Introduction

Executive functions (EFs) refer to higher-order cognitive processes that involve symbolic operations, such as attention, memory, processing speed, inhibition control, planning, cognitive flexibility, and concrete operations, which guide goal-directed and adaptive behaviors in response to specific situations (Baddeley, 1981; Hughes, 2013). The current approach to assess EFs refers to use standardized measurement tools, as paper-and pencil and computerized tests, consisting of a set of predefined stimuli delivered in a controlled laboratory environment via paper-and-pencil and/or via computer systems. For example, the Trail Making Test (TMT) is a reliable and valid measurement consisting of two parts: in the part A participants have to sequentially connect numbers and in the part B they have to alternatively connect numbers and letters. The aim of the test is to assess the cognitive ability to perform sequencing and visual search, as well as cognitive flexibility and attentional abilities (Reitan, 1958). Other measurements include the Winsconsin Card Sort Test (WCST ; Heaton et al., 1993) to assess cognitive flexibility and attentional processes, the Stroop Test (Stroop, 1935) for attentional abilities and inhibition control, and the Tower of London (Burgess et al., 2006) for planning abilities. Although these tools have proved reliability and validity, various behavioral studies found to have a moderate level of ecological validity in predicting real-life performance (Chaytor & Schmitter-Edgecombe, 2003; Elkind et al., 2001), showing that low scores on standardized measures do not inevitably entail poor executive functioning in real life and high scores on standardized tests do not reflect good executive performance in real-life activities (Barker et al., 2004; Chevignard et al., 2000; Manchester et al., 2004; Renison et al., 2012)

In order to fill this gap, in the last decades, Serious Games (SG) are representing a novel methodological approach that has attracted the attention in neuropsychological research (Fleming et al., 2014, 2017). Serious games are computer games that are not just designed for fun purposes and the primary purpose is training, education, intervention, promoting behavioral

changes and investigating human behavior (Connolly et al., 2012; Fleming et al., 2014, 2017).

Furthermore, SG cannot only vary according to the purpose but can also vary according to the technology used and the interaction. Indeed, a SG can be rendered via a non-immersive screen-based (2D) in which the interaction is due to a keyboard or a mouse. The technological advances allow also creating Immersive Virtual Reality Games (3D-IVRG) in which participants are fully immersive in a simulated synthetic environment provided by Head-Mounted Displays (HMD), in which audio, olfactory, and vibrotactile stimuli can be added and the interaction can be rendered with both hands using specific controllers. The technological immersion allows to participants to feel themselves inside the simulated environment, perceiving it as it were real, as well as the interaction devices and stimuli allow to participants to act and interact with objects and situations as they were into the real ones (Slater, 2009). Despite the immersion advantage, 3D-IVRG present some disadvantages in term of technology costs. Indeed, a 3D-IVRG system requires as well as a personal computer, as a 2D system, a powerful graphic card that could support a Head Mounted Display (HMD) for the visualization and interaction with the environment.

As mentioned above, 2D and 3D-SG are promising tools for neuropsychological assessment of EFs and various environments have been mainly developed and tested for 3D systems, such as virtual beach (Elkind et al., 2001), classroom (Climent & Banterla, 2012; Diaz-Orueta et al., 2014; 2019; Henry et al., 2012; Iriarte et al., 2012; Parsons et al., 2007; ;), mall (Rand et al., 2007; 2009), and office or building (Matheis et al., 2007; Pugnetti et al., 1995; 1998).

More in detail, Pugnetti et al. (1995; 1998) developed one of the first virtual building, WCST-based, in which participants should achieve the exit door matching stimuli in the environment, such as shape, colour, and numbers. Results showed that navigational factor in the virtual environment confused the results on cognitive flexibility. Another study on cognitive flexibility developed a virtual beach scenario, in which participants had to deliver sodas, frisbees,

popsicles, and beach balls to umbrellas (Elkind et al., 2001). Results showed that the system was not able to discriminate between patients with impairments and healthy subjects. More recently, other virtual environments based on distractor stimuli, have been developed and tested to improve neuropsychological assessment (Climent & Banterla, 2012; Diaz-Orueta et al., 2014; 2019 Henry et al., 2012; Iriarte et al., 2012; Parsons et al., 2007;). Various virtual classroom environments, including simulated-real elements, such as desks, children, teacher, and a whiteboard where the tasks were administered have been tested on cognitive constructs. For example, results on children with attention deficit and hyperactivity disorder vs. children with a typical development showed significant differences in errors and omissions and body movement (Parsons et al., 2007).

Furthermore, to our knowledge, a few studies have compared 2D and 3D systems for the assessment of EF. For example, Lo Priore et al. (2003) compared skin conductance response (SCR) between a 3D and a 2D store, in which patients and healthy subjects had to explore the environments and solve six task sequences, ordered by complexity and created to stimulate executive functions, programming, categorical abstraction, short-term memory and attention. The results showed a significantly higher SCR during the 3D-store condition than 2D-store, suggesting a higher individual engagement and activation.

Starting from these premises and according to the 3D disadvantages we have created a narrative-contextualized SG in 2D and 3D, EF-based, comparing the two conditions performance (2D vs. 3D) and correlating them with standardized tasks performance.

2. Materials and Methods

2.1. Subjects

Ninety-four healthy subjects participated to the data analysis. 47 subjects participated to the 3D condition study. For 2D condition group, we performed a random-stratified sampling over

the database of our previous research (Chicchi Giglioli et al., 2018) to obtain a sample (n=47) similar in sociodemographic characteristics (age, gender, education, and level of use of technologies) to the 3D condition group. 3D condition's participants were recruited through local advertisement among college students and administration and workers' staff of the Polytechnic University of Valencia. To be included in the study analysis, participants were required to have a score higher than 24 in the "Mini-Mental State Examination" (MMSE) (Flostein et al., 1975). Before participating to the study, each subject received written information about the research and was required to give written consent for inclusion in the experiment. The study obtained the ethical approval by the Ethical Committee of the Polytechnic University of Valencia. Table 1 shows descriptive data of participants according the two conditions.

Table 1
Demographic data (Mean, Standard Deviation (SD), *n*, and %] values) of participants (n=47 for condition).

2.2.   Questionnaires

Participants completed four questionnaires. First, they responded to a sociodemographic questionnaire about age, gender, education and level of use of technologies. Second, subjects completed the following self-report instruments: Cognitive Flexibility Scale (CFS) (Martin & Rubin, 1995), which includes 12 items that participants evaluated using a six-point rating scale ranging from 1 ("Strongly disagree") to 6 ("Strongly agree") and a score of 60 or more indicates that the individual has a high cognitive flexibility; Attentional Control Scale (ACS) (Derryberry & Reed, 2002), which comprises 20 items scored from 1 ("Almost never") to 4 ("Always") and higher scores show a great ability to maintain voluntarily attention in a task, while low values are related to greater attention stiffness; and Barratt Impulsiveness Scale (BIS) (Barratt, 1959) which includes 30 items that participants assessed using a four-point rating scale (0 - "Rarely or never", 1 – "Occasionally", 3 – "Often" and 4 – "Always or almost always"). A score of 72 or more means that individual is highly impulsive. Between 52 and 71

is considered within the normal limits of impulsivity. A score below 52 represents a subject with a high control of impulsivity.

2.3. Standard tasks

Participants completed a total of 6 standardized tasks (ST). Computerized versions of ST have been administered to participants: Dot Probe Task version published by Miller and Fillmore (2010), the neutral pictures (20 in total) were selected from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008); Go/NoGo Task (Fillmore et al, 2006); Stroop Test, (Stroop 1935); Trail Making Task, paper-and-pencil-based version published by Reitan (1958); Wisconsin Card Sorting Test (Grant & Berg, 1948); and Tower of London - Drexler (TOLDX, Culbertson & Zilmer, 1999). The standard tasks were randomly presented and performed on a personal computer.

2.4. Serious game scenario

SG consisted of a simulation of narrative-contextualized situations, settled in a spaceship, which aim was to discover a new land for living, composed by eight missions (tasks) (Fig. 1). These tasks were used to assess attention, inhibition control, impulsivity, planning, and cognitive flexibility. Table 2 provides a description of the SG tasks, the ST, the outcome measures and the EFs tested. Participant was the protagonist of the simulation and the narration drive him/her in the different situations and activities. While participants performed the activities, the systems recorded the following parameters for each task: execution times, latency times, and the correct answers. Furthermore, in accordance with the specificity of each SG-EFs component, specific parameters have been collected (e.g. perseverative responses in CF2 and CF3).

The SG was played by participants using two different interfaces, one based on a personal computer and a keyboard (2D condition) and another one where the participants wore an HMD

device (HTC VIVE, https://www.vive.com/eu/product/), performing the tasks in a three-dimensional virtual environment (3D condition). The virtual screening system was developed using Unity 5.5.1f1 software, applying c# programming language using the Visual Studio tool.

Figure 1: Screen shot of the SG tasks.

Table 2
Descriptions of tasks, the standardized tasks (ST), outcome measures, and the EFs assessed.

2.5. Experimental procedure

The study consisted of two parts: part A and part B. In the part A, participants completed the four questionnaires (demographic, CFS, ACS and BIS) and performed the standardized tasks on a personal computer. The standardized tasks were randomized for each participant. In the part B, subjects performed the 2D serious games using a personal computer (2D condition) or the 3D serious games using a Head Mounted Display HTC VIVE (3D Condition).

2.6. Data analysis

The analyses were performed using SPSS version 22.0 (Statistical Package for the Social Sciences for Windows, Chicago, IL) for PC. First, we verified the assumptions of normality applying Kolmogorov Smirnov ($p > .05$) and the internal consistency of the scales was assed via Cronbach's alpha. Second, it has been verified the normal cognitive functioning of the subjects (MMSE>24; CFS, ACS and BIS). Third, Pearson correlations were computed between psychological questionnaires, ST and the SG performance in both conditions. Finally, paired t-test was conducted to compare differences between conditions. The level of significance was set at $\alpha = 0.05$.

3. Results and Discussion

3.1 Results

After confirming the assumption of normality (Kolmogorov Smirnov $p > .05$) on the distribution of the questionnaires and ST scores, we assessed the internal consistency of the self-report scales (Cronbach´s alpha αattention = .826, αcognitive flexibility = .701, αimpulsiveness = .746; bootstrap 95%).

Regarding the cognitive functioning of the two groups, mean total scores on CFS showed that subjects had a normal cognitive flexibility: 2D-CFS (M=53.78, SD= 6.57), 3D-CFS (M=56.38, SD=5.83); the mean total values on attentional control showed a normal functioning: 2D-ACS (M=56.59, SD=8.14), 3D-ACS (M=55.79, SD=8.46); and for impulsivity a low mean scores have been obtained: 2D-BIS (M=42.82, SD=11.87), 3D-BIS (M= 45.71, SD=12.91) indicating that subjects had a high control of impulsivity. Table 3 and 4 show descriptive data for the ST and 2D/3D-Serious Games.

Table 3
Mean, Standard Deviation (SD) and [Range] values for Standard Tasks and Serious Games variables (2D condition)

Table 4
Mean, Standard Deviation(SD) and [Range] values for Standard Tasks and 3D-Serious Games variables (3D condition)

3.1.1. Questionnaires – Serious Game correlations

Table 5 reports Pearson correlations calculated for each questionnaire and the serious games performance in the two conditions.

Table 5
Pearson correlations between Questionnaires and Serious Games Variables (2D and 3D condition).

3.1.2. Standard Tasks – Serious Game correlations

Pearson correlations calculated for each standard task and serious games in the two conditions have been reported in Table 6.

Table 6
Pearson correlations between Standard Tasks and Serious Games Variables.

3.1.3. Condition comparison: 2D versus 3D

A paired t-test has been conducted to compare behavioral responses in 2D and 3D conditions. There were significant differences in the scores, especially in attentional Serious Games (see Table 7).

Table 7
Significant differences between 2D and 3D performance

3.2. Discussion and conclusions

The main aim of this study was to investigate the plausibility and feasibility of a serious game comparing 2D versus a 3D system, to integrate the traditional evaluation of EFs with a more ecologically valid assessment.

Our results yielded four main findings. First, 2D condition showed lower latency time and higher correct answer than traditional tasks and with respect to attentional abilities, which decreased in 3D condition together with higher scores than 2D. Regarding inhibition control, 2D and 3D condition showed higher latency times and lower correct answers than traditional

tasks but in 3D condition latency time's decrease and correct answers increase than 2D condition. As for cognitive shifting, 3D condition showed higher latency times and correct answers than 2D condition and both conditions showed higher latency times and correct answers than traditional tasks. Planning abilities in 3D condition showed lower scores and times than 2D condition.

Second, we found a few correlations between questionnaires and serious game using both systems. However, the 3D condition provided more correlations and moderate than 2D condition. These results could depend on that executive functions are traditionally measured and assessed using implicit tasks and less with questionnaires. Third, the correlation results between 2D and 3D serious game and standardized tasks showed that 3D condition is more able to detect higher correlations than 2D condition.. Fourth, the comparison between the two systems have revealed that 3D is able to generate lower reaction times, higher correct answers, and lower perseverative responses in attentional abilities, inhibition control, and cognitive shifting than 2D condition. According to the results,   3D condition seems to allow at participants acting and interacting with objects and situations in a more naturalistic way (Slater, 2009).

Even though the present findings are relevant, they present some limitations. First, healthy subjects that composed the sample limited the sensitivity of the results. Second, considering the use of virtual reality, it would be important to also assess the individual's perception of usability and presence. Further studies are required to examine plausibility, feasibility of the two systems in accordance with EFs, mainly regarding its sensitivity, including clinical populations, as well as its reliability and validity according to the different criterions and the distinctive components of cognitive functioning. Nevertheless, the present study suggested that individuals provided a better performance using a 3D system than a 2D system. .

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Funding Statement

References

Baddeley, A. (1981). The cognitive psychology of everyday life. *Br. J. Psychol*., *72*, 257–269. doi: 10.1111/j.2044-8295.1981.tb02184.x

Barker, L. A., Andrade, J., & Romanowski, C. A. J. (2004). Impaired implicit cognition with intact executive function after extensive bilateral prefrontal pathology: A case study. *Neurocase, 10*(3), 233-248.

Barratt, E. S. (1959). Anxiety and impulsiveness related to psychomotor efficiency. *Perceptual and motor skills*, *9*(3), 191-198.

Burgess, P. W., Alderman, N., Forbes, C., Costello, A., LAURE, M. C., Dawson, D. R., ... & Channon, S. (2006). The case for the development and use of "ecologically valid" measures of executive function in experimental and clinical neuropsychology. *Journal of the international neuropsychological society, 12*(2), 194-209.

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology review*, *13*(4), 181-197.

Chaytor, N., Schmitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of clinical neuropsychology*, *21*(3), 217-227.

Chelune, G. J., & Moehle, K. A. (1986). Neuropsychological assessment and everyday functioning.

Chevignard, M., Pillon, B., Pradat-Diehl, P., Taillefer, C., Rousseau, S., Le Bras, C., & Dubois, B. (2000). An ecological approach to planning dysfunction: script execution. *Cortex*, *36*(5), 649-669.

Giglioli, I. A. C., de Juan Ripoll, C., Parra, E., & Raya, M. A. (2018). EXPANSE: A novel narrative serious game for the behavioral assessment of cognitive abilities. *PloS one*, *13*(11), e0206925.

Climent, G., & Bánterla, F. (2011). Nesplora classroom, ecological assessment of attentional processes. *Theoretical Manual. Donostia, Spain: Nesplora*.

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & education*, *59*(2), 661-686.

Culberston, W. C., & Zillmer, E. A. (1999). Tower of London-Drexel (TOLDX). *Examiners's manual. Research version. Canadá: Multi-Health Systems Inc*.

Derryberry, D., & Reed, M. A. (2002). Anxiety-related attentional biases and their regulation by attentional control. *Journal of abnormal psychology*, *111*(2), 225.

Díaz-Orueta, U., Garcia-López, C., Crespo-Eguílaz, N., Sánchez-Carpintero, R., Climent, G., & Narbona, J. (2014). AULA virtual reality test as an attention measure: Convergent validity with Conners' Continuous Performance Test. *Child Neuropsychology*, *20*(3), 328-342.

Díaz-Orueta, U. (2017). Advances in Neuropsychological Assessment of Attention. *The Role of Technology in Clinical Neuropsychology*, 103.

Elkind, J. S., Rubin, E., Rosenthal, S., Skoff, B., & Prather, P. (2001). A simulated reality scenario compared with the computerized Wisconsin Card Sorting Test: An analysis of preliminary results. *CyberPsychology & Behavior*, *4*(4), 489-496.

Fillmore, M. T., Rush, C. R., & Hays, L. (2006). Acute effects of cocaine in two models of inhibitory control: implications of non-linear dose effects. *Addiction*, *101*(9), 1323-1332.

Fleming, T. M., Cheek, C., Merry, S. N., Thabrew, H., Bridgman, H., Stasiak, K., ... & Hetrick, S. (2014). Serious games for the treatment or prevention of depression: a systematic review.

Fleming, T. M., Bavin, L., Stasiak, K., Hermansson-Webb, E., Merry, S. N., Cheek, C., ... & Hetrick, S. (2017). Serious games and gamification for mental health: current status and promising directions. *Frontiers in psychiatry*, *7*, 215.

Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of experimental psychology*, *38*(4), 404.

Heaton, R. K., & Staff, P. A. R. (1993). Wisconsin card sorting test: computer version 2. *Odessa: Psychological Assessment Resources*, *4*, 1-4.

Henry, M., Joyal, C. C., & Nolin, P. (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), 125-131.

Hughes, C. (2013). Executive function: Development, individual differences, and clinical insights. In *Neural Circuit Development and Function in the Brain*, 429-445.

Iriarte, Y., Diaz-Orueta, U., Cueto, E., Irazustabarrena, P., Banterla, F., & Climent, G. (2016). AULA—advanced virtual reality tool for the assessment of attention: normative study in Spain. *Journal of Attention Disorders*, *20*(6), 542-568.

Lang, P. J. (2005). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. *Technical report*.

Li, J., Theng, Y. L., & Foo, S. (2016). Effect of exergames on depression: a systematic review and meta-analysis. *Cyberpsychology, Behavior, and Social Networking*, *19*(1), 34-42.

Lo Priore, C., Castelnuovo, G., Liccione, D., & Liccione, D. (2003). Experience with V-STORE: considerations on presence in virtual environments for effective neuropsychological rehabilitation of executive functions. *Cyberpsychology & behavior*, *6*(3), 281-287.

Manchester, D., Priestley, N., & Jackson, H. (2004). The assessment of executive functions: Coming out of the office. *Brain injury*, *18*(11), 1067-1081.

Martin, M. M., & Rubin, R. B. (1995). A new measure of cognitive flexibility. *Psychological reports*, *76*(2), 623-626.

Matheis, R. J., Schultheis, M. T., Tiersky, L. A., DeLuca, J., Millis, S. R., & Rizzo, A. (2007). Is learning and memory different in a virtual environment?. *The Clinical Neuropsychologist*, *21*(1), 146-161.

Merry, S. N., Stasiak, K., Shepherd, M., Frampton, C., Fleming, T., & Lucassen, M. F. (2012). The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. *Bmj*, *344*, e2598.

Miller, M. A., & Fillmore, M. T. (2010). The effect of image complexity on attentional bias towards alcohol-related images in adult drinkers. *Addiction*, *105*(5), 883-890.

Nyhus, E., & Barceló, F. (2009). The Wisconsin Card Sorting Test and the cognitive assessment of prefrontal executive functions: a critical update. *Brain and cognition*, *71*(3), 437-451.

Parsey, C. M., & Schmitter-Edgecombe, M. (2013). Applications of technology in neuropsychological assessment. *The Clinical Neuropsychologist*, *27*(8), 1328-1361.

Parsons, T. D., Bowerly, T., Buckwalter, J. G., & Rizzo, A. A. (2007). A controlled clinical comparison of attention performance in children with ADHD in a virtual reality classroom compared to standard neuropsychological methods. *Child Neuropsychology*, *13*(4), 363-381.

Pugnetti, L., Mendozzi, L., Motta, A., Cattaneo, A., Barbieri, E., & Brancotti, A. (1995). Evaluation and retraining of adults' cognitive impairments: Which role for virtual reality technology?. *Computers in Biology and Medicine*, *25*(2), 213-227.

Pugnetti, L., Mendozzi, L., Attree, E. A., Barbieri, E., Brooks, B. M., Cazzullo, C. L., ... & Rose, F. D. (1998). Probing memory and executive functions with virtual reality: Past and present studies. *CyberPsychology & Behavior*, *1*(2), 151-161.

Rand, D., Katz, N., & Weiss, P. L. (2007). Evaluation of virtual shopping in the VMall: Comparison of post-stroke participants to healthy control groups. *Disability and rehabilitation*, *29*(22), 1710-1719.

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

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

Renison, B., Ponsford, J., Testa, R., Richardson, B., & Brownfield, K. (2012). The ecological and construct validity of a newly developed measure of executive function: the virtual library task. *Journal of the International Neuropsychological Society*, *18*(3), 440-450.

Rizzo, A., John, B., Newman, B., Williams, J., Hartholt, A., Lethin, C., & Buckwalter, J. G. (2013). Virtual reality as a tool for delivering PTSD exposure therapy and stress resilience training. *Military Behavioral Health*, *1*(1), 52-58.

Roepke, A. M., Jaffee, S. R., Riffle, O. M., McGonigal, J., Broome, R., & Maxwell, B. (2015). Randomized controlled trial of SuperBetter, a smartphone-based/internet-based self-help tool to reduce depressive symptoms. *Games for health journal*, *4*(3), 235-246.

Slater, M., Usoh, M., & Steed, A. (1994). Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, *3*(2), 130-144.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3549-3557.

Stuss, D. T., Benson, D. F., Weir, W. S., Naeser, M. A., Lieberman, I., & Ferrill, D. (1983). The involvement of orbitofrontal cerebrum in cognitive tasks. *Neuropsychologia*, *21*(3), 235-248.

Stuss, D. T., & Alexander, M. P. (2000). Executive functions and the frontal lobes: a conceptual view. *Psychological research*, *63*(3-4), 289-298.

Table 1

Demographic data (Mean, Standard Deviation (SD), n, and % values) of participants (n=47 for condition).

Condition	2D			3D		
Demographic	Mean (*SD*)	*n*	%	Mean (*SD*)	*n*	%
Age	31.60(*8.76*)			28.68(*11.18*)		
Gender (M/F)						
Male		17	36%		20	43%
Female		30	64%		27	57%
Level of use of technologies (H/L)						
High		26	55%		24	51%
Low		21	45%		23	49%
Education						
High School Degree		16	34%		12	25%
Bachelor's degree		21	45%		23	50%
Master's degree		10	21%		12	25%

Table 2

Descriptions of the SG tasks, the standardized tasks (ST), outcome measures, and the EFs assessed.

SG tasks	Standardized Task (ST)	Outcome Measures	EFs assessed
Task 1 (AT1): "The takeoff: you are the pilot and you have to take off the spaceship. To take off you should follow the earth planet images that appear in front of you"	Dot Probe Task	Latency times Correct Answers	Attention Inhibition control
Task 2 (AT3): "Aliens attack: in the space that are a lot of elements and in this moment your spaceship is attacked by aliens and you have to avoid and kill the aliens"	Go/NoGo Task	Latency times Correct Answers	Attention Inhibition control
Task 3 (AT4): "The oxygen valve has broken! You have to repair it but the valve is closed in a strongbox. The strongbox has a code that you have to unlock"	Stroop Test	Latency times Correct Answers	Attention Inhibition control
Task 4 (CF1): "Water and Food: the water and food supply is almost all gone. To obtain water you have to pump up the level and for food you have to cultivate"	TMT A-B	Total Time A Total Time B	Attention Cognitive shifting
Task 5 (CF2): "The orchard is empty" You should grow up new plants. You have 4 kinds of plants and you have to decide in which group of plants you joint the new plant"	WSCT	Latency time Correct answers Perseverative responses	Cognitive shifting
Task 6 (CF3): "Without fuel: your fuel supply is finished. To obtain fuel you should activate the turbine. For activating you have to combine different elements two by two"	WSCT	Latency time Correct answers Perseverative responses	Cognitive shifting
Task 7 (PL1): "Lock up: you are lock up in a room and you have to use and combine different objects that you find in the room to open the door"	Tower of London	Total score Initial time Execution time Total time	Planning
Task 8 (AT2): "Resources: you have achieved the new planet and you have to manage the resources. To manage the resources, you should select the correct elements that you need to live"	Dot Probe Task	Latency times Correct Answers	Attention

Table **3**

Mean (Standard Deviation: SD) and [Range] values for Standard Tasks and Serious Games variables (2D condition)

Standard Tasks	Mean	(SD)	[Range]
Dot Probe Task			
Correct answers (%)	.988	(.016)	[.925-1]
Latency time (ms)	.451	(.045)	[.371-.544]
Go/Nogo Task			
Correct answers (%)	.986	(.018)	[.912-1]
Latency time correct answers - go (ms)	.407	(.040)	[.328-.562]
Stroop Test			
Correct answers (%)	.991	(.015)	[.935-1]
Latency time (ms)	1.321	(.261)	[.899-2.004]
Trail Making Task			
Total time (A)	35.153	(7.791)	[21.800-58.655]
Total time (B)	49.922	(11.044)	[32.447-82.249]
Wisconsin Card Sorting Test			
Correct answers (%)	.647	(.228)	[.167-.900]
Latency time (ms)	.995	(.374)	[.193-1.778]
Perseverative responses (count)	19.511	(19.3413)	[0-77]
Tower of London			
Total score	26.362	(2.453)	[19-30]
Initial time (ms)	10.057	(5.902)	[3.718-31.275]
Execution time (ms)	19.927	(5.345)	[8.998-33.702]
Total time (ms)	298.726	(87.240)	[175.545-638.906]
SG (2D condition)	Mean	(SD)	[Range]
AT1			
Correct answers (%)	.986	(.018)	[.900-1]
Latency time (ms)	.445	(.044)	[.352-.527]
AT2			
Correct answers (%)	.984	(.023)	[.864-1]
Latency time (ms)	.443	(.057)	[.363-.676]
AT3			
Correct answers (%)	.953	(.100)	[.468-1]
Latency time correct answers - go (ms)	.642	(.175)	[.210-1.067]
AT4			
Correct answers (%)	.959	(.086)	[.597-1]
Latency time (ms)	2.281	(.848)	[1.175-3.922]
CF1			
Total time (A)	50.627	(13.965)	[24.183-93.955]
Total time (B)	50.751	(14.683)	[22.984-89.992]
CF2			
Correct answers (%)	.567	(.258)	[.094-.884]
Latency time (ms)	.945	(.416)	[.140-1.808]
Perseverative responses (count)	27.617	(24.334)	[5-112]
CF3			

Correct answers (%)	.628	(.119)	[.258-.833]
Latency time (ms)	.751	(.348)	[.186-1.601]
Perseverative responses (count)	4.872	(.448)	[3-5]
PL1			
Total score	11.723	(1.192)	[9-15]
Initial time (ms)	14.258	(16.852)	[1.798-74.930]
Execution time (ms)	300.238	(115.874)	[50.853-535.502]
Total time (ms)	314.496	(117.012)	[53.152-538.713]
Correct answers (%)	.628	(.119)	[.258-.833]
Latency time (ms)	.751	(.348)	[.186-1.601]
Perseverative responses (count)	4.872	(.448)	[3-5]

Table **4**

Mean (Standard Deviation: SD) and [Range] values for Standardized Tasks and 3D-Serious Games variables (3D condition)

Standardized Tasks	Mean	(SD)	[Range]
Dot Probe Task			
Correct answers (%)	.964	(.144)	[.013-1]
Latency time (ms)	.464	(.064)	[.352-.668]
Go/Nogo Task			
Correct answers (%)	.989	(.014)	[.936-1]
Latency time correct answers - go (ms)	.395	(.036)	[.331-.517]
Stroop Test			
Correct answers (%)	.989	(.020)	[.887-1]
Latency time (ms)	1.253	(.261)	[.919-2.072]
Trail Making Task			
Total time (A)	37.384	(10.757)	[25.273-78.179]
Total time (B)	51.011	(11.520)	[32.193-85.786]
Wisconsin Card Sorting Test			
Correct answers (%)	.708	(.177)	[.134-.896]
Latency time (ms)	1.137	(.273)	[.296-1.599]
Perseverative responses (count)	16.787	(13.763)	[5-77]
Tower of London			
Total score	26.000	(3.967)	[11-30]
Initial time (ms)	12.183	(7.354)	[2.301-38.432]
Execution time (ms)	20.816	(7.396)	[7.080-37.435]
Total time (ms)	350.199	(123.597)	[125.710-677.913]
SG (3D condition)	Mean	(SD)	[Range]
AT1			
Correct answers (%)	.985	(.039)	[.750-1]
Latency time (ms)	.367	(.048)	[.199-.498]
AT2			
Correct answers (%)	.973	(.031)	[.875-1]
Latency time (ms)	.397	(.036)	[.336-.483]
AT3			
Correct answers (%)	.963	(.031)	[.872-1]
Latency time correct answers - go (ms)	.427	(.035)	[.358-.513]
AT4			
Correct answers (%)	.990	(.018)	[.887-1]
Latency time (ms)	1.747	(.330)	[1.175-2.693]
CF1			
Total time (A)	53.744	(24.337)	[22.764-138.092]
Total time (B)	51.847	(13.873)	[33.588-97.681]
CF2			
Correct answers (%)	.702	(.213)	[.078-.909]
Latency time (ms)	1.086	(.289)	[.110-1.536]
Perseverative responses (count)	15	(19.116)	[0-118]
CF3			

	Correct answers (%)	.652	(.111)	[.391-.845]
	Latency time (ms)	.797	(.541)	[.248-3.678]
	Perseverative responses (count)	4.979	(.254)	[4-6]
PL1				
	Total score	11.745	(1.343)	[8-15]
	Initial time (ms)	13.474	(15.974)	[1.106-69.543]
	Execution time (ms)	292.263	(106.577)	[110.771-499.882]
	Total time (ms)	305.736	(109.974)	[116.277-518.200]

	Correct answers (%)	.652	(.111)	[.391-.845]
	Latency time (ms)	.797	(.541)	[.248-3.678]
	Perseverative responses (count)	4.979	(.254)	[4-6]

Table **5**

Pearson correlations between questionnaires and serious games variables for condition (2D and 3D condition).

	ATT	CF	IM	Imo	Ico	Inp
2D CONDITION						
2D-AT4_LT						-.404**
2D-CF2_PR						.289**
3D CONDITION						
3D-AT1_CA					-.310*	
3D-AT3_CA					-.300*	
3D-CF2_CA		.335*				
3D-CF2_LT						-.311*
3D-CF2_PR			.304*			.361*
3D-CF3_LT		-.310*				

Note. *p<.05 **p<.01*. ATT = Attentional Control Scale, CF = Cognitive Flexibility Scale, IM = Barratt Impulsiveness Scale, Imo = motor impulsivity, Ico = cognitive impulsivity, Inp = Non-planning impulsivity. CA = Correct answers, LT = Latency time, PR = Perseverative responses, ST = Switch time, NST = Non-switch time.

Table **6**

Pearson correlations between Standard Tasks and Serious Games Variables.

	1		2		3		4		5			6		
	CA	LT	CA	LT	CA	LT	TT-A	TT-B	CA	LT	PR	IT	ET	TT
2D														
AT1														
CA		.363*												
LT		.761**		.482**			.303**		-.369*					
AT2														
CA	.542**		.294*			-.335*								
LT	-.323*	.581**		.392**		.377**								
AT3														
LT				.305*										
AT4														
CA							.320*	.298*						
LT	-.375**												.422**	
CF1														
TT(A)														.341*
TT(B)			.297*						.299*	.357*				.315*
CF2														
CA		-.463**							.410**	.315*	.335*			
LT		-.320*							.422**	.477**	.348*			
PR		.345*												
CF3														
CA									.489**	.467**				
LT							.318*		.339*	.442**				
PL1														
TS					.403**									-.291*
IT														.384**
3D														
AT1														
LT		.542**		.451**	.293*				.337*					
AT2														
CA		.329*												
LT		.657**		.532**					.336*					
AT3														
CA			.483**							.308*	-.323*	.301*		
LT		.361*												
AT4														
CA			.353*		.719**	.314*								

LT	.353*			.539**	.436**	.439**	-.343*				.456**
CF1											
TT(A)						.396**	-.362*		.389**		.372*
TT(B)			.303*	.302*							
CF2											
CA		-.318*				-.533**	.513**		-.406**		
LT								.294*			
PR					.288*	.309*	-.548**		.361*		
CF3											
CA	-.298*										
PR					.519**					.341*	
PL1											
TS						-.401**					
IT											.372*
ET						.526**			.290*		.463**

Note. *p<.05 **p<.01. 1 = Dot Probe Task, 2 = Go/Nogo Task, 3 = Stroop Test, 4 = Trail Making Test, 5 = Wisconsin Card Sorting Task, 6 = Tower of London. CA = Correct answers, LT = Latency time, PR = Perseverative responses, ST = Switch time, NST = Non-switch time, TT = Total time, ET = Execution time, IT = initial time.

Table **7**

Significant differences between
2D and 3D performance.

Variable	Mean	p
AT1_LT		
2D	.445	.000
3D	.367	
AT2_CA		
2D	.984	.038
3D	.973	
AT2_LT		
2D	.443	.000
3D	.397	
AT3_LT		
2D	.642	.000
3D	.427	
AT4_CA		
2D	.959	.021
3D	.990	
AT4_LT		
2D	.2.281	.000
3D	1.730	
CF2_CA		
2D	.567	.007
3D	.702	
CF2_PR		
2D	27.62	.006
3D	15.00	

Note. CA = Correct answers, LT = Latency time, PR = Perseverative responses.