Taking Neuropsychological Test to the Next Level: Commercial Virtual Reality Video Games for the Assessment of Executive Functions

Federica Pallavicini(✉), Alessandro Pepe, and Maria Eleonora Minissi

Department of Human Sciences for Education "Riccardo Massa", University of Milan-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy
federica.pallavicini@gmail.com

Abstract. Virtual reality and video games are increasingly considered as potentially effective tools for the assessment of several cognitive abilities, including executive functions. However, thus far, only non-commercial contents have been tested and virtual reality contents and video games have been investigated separately. Within this context, this study aimed to explore the effectiveness in the assessment of executive functions using a new type of interactive content - commercial virtual reality games - which combines the advantages of virtual reality with that of commercial video games. Thirty-eight participants completed the Trial Making Test as traditional commonly used assessments of executive functions and then played the virtual reality game Audioshield using an HTC Vive systems. Scores on the Trial Making Test (i.e., time to complete part A and B) were compared to scores obtained on Audioshield (i.e., number of orbs hit by the players and technical score). The results showed that: (a) performance on the Trial Making Test correlated significantly with performance on the virtual reality video game; (b) scores on Audioshield can be used as a reliable estimator of the results of Trial Making Test.

Keywords: Virtual reality · Virtual reality video games · Video games · Cognitive assessment · Executive functions

1 Introduction

1.1 Current Issues in the Assessment of Executive Functions

The term executive functions refers to a wide range of high-level cognitive processes in the prefrontal lobe [1] associated with reactive inhibition and the regulation of goal achievement behavior [2]. Inhibitory control (i.e., resisting one's initial impulse or a strong pull to do one thing and instead act more wisely) [3], working memory (i.e., holding information in mind while performing one or more mental operations) [4], and cognitive flexibility (i.e., the ability to flexibly adjust to changing demands or priorities, to look at the same thing in different ways or from different perspectives, as required for set shifting or task switching) [5] have been identified as the three main components of executive functions [6]. Executive functions are essential for many everyday life

M. Antona and C. Stephanidis (Eds.): HCII 2019, LNCS 11573, pp. 133–149, 2019.
https://doi.org/10.1007/978-3-030-23563-5_12

activities across a wide range of contexts, such as the ability to work, function independently at home, or maintain social relationships [3, 7]. Furthermore, these cognitive skills have been found to be predictive of health, wealth, academic success, and quality of life throughout life [8].

Given the relevance of executive functions in daily life and the great complexity of cognitive impairments associated with them, an accurate and efficient assessment is of particular relevance [9]. The most widely used tools for the executive functions assessment are paper-and-pencil, and computer-based tests such as the Trial Making Test (TMT) [10], the Stroop test [11], and the Wisconsin Card Sorting Test (WCST) [12, 13]; however, these tests have several important limitations, including:

- Low accessibility: Traditional neuropsychological instruments have been developed to be used in clinical settings; hence, they are costly and require the presence of a health professional (i.e., neurologist or neuropsychologist). Accordingly, these measures are difficult to access and are often applied only retrospectively, that is, after the cognitive impairment has been detected. This delay has a serious negative influence on the effectiveness of the intervention and the treatment options [14]. In addition, most of these tests have been designed and tested on a population with medium educational levels; therefore, they are not suitable for individuals with less education [15];
- Low ecological validity: Most neuropsychological tests are unable to measure complex real-life activities [16] and fail to predict the way in which individuals manage these activities in their daily living [17]. For example, performance on traditional tests of executive functions, such as the WCST and the Stroop test, does not reflect activities of daily living and do not predict functional performance across a range of real-world situations [18]. In addition, although commonly used for clinical and research purposes, traditional neuropsychological tasks are thought to lack the stresses and distractions of real-life settings [19]
- Low quality of assessment data: Neuropsychological tests are often viewed as repetitive, frustrating, and boring, leading to individuals' disengagement, which may negatively affect the quality of assessment data [20].

In response to the limits of conventional assessments of executive functions, neuropsychologists are increasingly emphasizing the need for a new generation of "function-led" tests that would be able to asses real-world functioning and tap into a number of executive domains [21, 22]. A number of ecological tasks have already been developed for the evaluation of executive functions, such as the Multiple Errand Test (MET) [23], or the Behavioral Assessment of the Dysexecutive Syndrome (BADS) [24]. Even if this new type of tasks provided more accurate estimates of the individual functional performance in activities of daily living compared to traditional neuropsychological tests [18] they would have several limitations that restrict their adoption. Among the main one, ecological tasks are time-consuming and require the presence of a specialist [25]. In addition, data collected in the real-world environment are frequently confounded by numerous factors, which cannot be captured by these measures [26].

1.2 Information and Communication Technologies for the Neuropsychological Evaluation of Executive Skills

With the aim of overcoming the existing problem in neuropsychological assessment of executive functions, Information and Communication Technologies (ICT) are increasingly employed, particularly virtual reality and video games, which are described in detail in the following paragraphs.

Virtual Reality. By definition, virtual reality allows users to navigate and interact with a three-dimensional computer-generated environment in real-time [27]. This technology has been successfully applied in the assessment of cognitive skills, including executive functions, for decades [28, 29], when it was still much less advanced from the point of view of both hardware and software. Several studies have reported that virtual reality was effective in offering a more realistic, ecologically valid alternative to traditional neuropsychological tasks assessing executive functions [22, 30], providing a balance between naturalistic observations and the need to control the key variables [31]. Virtual environments, in fact, thanks to the possibility to provide various complex conditions [32] and to record and analyze individual performance [33], combine the rigor and the control of the laboratory measures with simulations that reflect real-life situations [34].

Several ad hoc virtual contents have been developed for the assessment of executive functions. In particular, different virtual shopping environments have been tested over the last decade to evaluate executive functions deficits in people with frontal head injuries or stroke, such as the Virtual Multiple Errand Test (VMET) [35, 36], the Adapted Four-Item Shopping Task [37], and the Virtual Supermarket (VSM) [38]. Another virtual reality tool that has been developed to test EF is the Jansari Assessment of Executive Functions (JEF) [39], which requires individuals to complete different tasks in an office, the Virtual Library Task (VLT) [40], that required individuals to complete tasks as if they were running a library, and RehabCity [41], which assessed executive functions by asking the users to accomplish goals that require problem-solving. Furthermore, a recent study reported the efficacy of three immersive virtual reality tasks (i.e., a seating arrangement task, an item location task, and a virtual parking simulator) in assessing executive functions in a healthy population of younger and older adults [42].

Video Games. Every day, about 2.6 billion people worldwide, 61% male and 39% female, with a mean age of 34 years old, play video games on their personal computers, mobile phones, or consoles [43]. One of the main reason for their success and diffusion is related to the fact that computer games are designed to be motivating and challenging [44], providing easy access to "fun" and to a sense of engagement and self-efficacy [45]. This ability to motivate the user [46], as well as to elicit positive emotions [47] and a short-term increase in subjective well-being [48], has in the last years begun to be exploited for purposes beyond entertainment, including the mental health panorama [49, 50]. In the neuropsychological field, in particular, video games appear relevant also because of their nature of requiring complex cognitive skills, including planning and strategizing, which place executive-functioning demands on players, such as selective attention and inhibition [51].

As early as in 1987, it was observed, for the first time, that famous commercial video games such as *Donkey Kong* (Nintendo) and *Pac-Man* (Atari) can have a positive effect on cognitive skills, improving the reaction times of older adults [52]. A few years later, in 1989, *Space Fortress*, the first of a series of non-commercial computer games designed by cognitive psychologists as a training and research tool [53] - often defined in literature as "serious games" as they use gaming features as the primary medium for serious purposes [54] - became so successful that it was added to the training program of the Israeli Air Force. From that moment on, efficacy has been demonstrated not only for non-commercial video games developed ad hoc, but also for commercial video games, in the training of several cognitive abilities [55], especially perceptual attentional skills [51], mental spatial rotation abilities [56], and executive functions, such as task-switching [57] and working memory [58].

Regarding the use of video games in the assessment of executive functions, especially in these last years, different serious games and game-based versions of neuropsychological tests have been tested. For example, the time to complete an ad hoc low-cost computer game, the *Stroop Stepping Test* (SST), developed using dance pads adapted from exercise-based video games, has been reported to correlate significantly with traditional measure of executive functions (i.e., TMT, Stroop task) among older people [59]. Another serious game has been developed to assess the cognitive status of the elderly through a touch-based device, showing a strong relationship between the video game performance and inhibitory ability as measured in the Stroop task [60]. In a subsequent study, the same authors tested the effectiveness of a serious game developed with a low-cost game engine to assess executive functions skills, reporting strong correlations between classic test (i.e., WCST, Stroop task, N-Back task) and the game performance [61]. Furthermore, a computational model based on game data obtained by playing an ad hoc video game has been showed to be effective in the estimation of TMT performance [62]. Finally, two recent studies have shown a correlation between performance on serious games developed for the assessment of executive functions (i.e., *EXPANSE*, *Kitchen and Cooking*) and scores on TMT [63, 64].

1.3 Aim of the Study

As described in the above sections, several previous studies have provided evidence on the efficacy of both virtual reality content as well as video games in the assessment of executive functions [22]. In particular, they have demonstrated the effectiveness of using virtual reality contents or video games created specifically for the measurement of these cognitive functions in the neuropsychological assessment. However, studies conducted thus far have some limitations, specifically:

- They have focused only on the testing of non-commercial contents: These types of solutions are often expensive and time-consuming in their development [65], as well as focused on a single specific population, which makes it difficult to replicate the study and to standardize the data.
- They have investigated separately virtual reality contents and video games: Thus far, only non-gamified virtual contents [35], or non-immersive video games (i.e., played on a desktop) [62], have been tested.

A new type of interactive content, in particular, virtual reality video games, could address such limitations. Immersive video games, which entered the market in 2016 thanks to the commercial diffusion of products, such as HTC Vive (Valve) and PlayStation VR (Sony Corp.), could be particularly interesting not only for entertainment purposes, but also in many other fields, including psychology and neuropsychology [66, 67]. Thanks to unique features of immersion (i.e., the perception of being physically present in a non-physical world) [68] and interactivity (i.e., the ability of the digital content to respond to the users' actions and be modified by them), virtual reality video games lead in fact to incomparable opportunities, significantly different from those provided by non-immersive gaming, such as traditional console or PC games [67, 69] Interestingly, a previous study [70] has reported the efficacy of a virtual reality video game, in particular, an exercise-based dance game (i.e., *DANCE*), for the training of executive functions in older people. However, this study tested only a non-commercial virtual reality game, and no indications regarding the possible effectiveness of this type of content even for the assessment of executive functions were given.

Within the context described above, this study aimed to compare the effectiveness of a commercial virtual reality video game (i.e., a dance game) in the assessment of executive functions of healthy young adults with the effectiveness of traditional neuropsychological measures. Furthermore, the specific objective of this experiment was to implement a structural equation model to estimate the association of the traditional neuropsychological measures (as a latent variable) with the virtual reality game performance (see the conceptual model in Fig. 1). The main hypotheses explored by the research were:

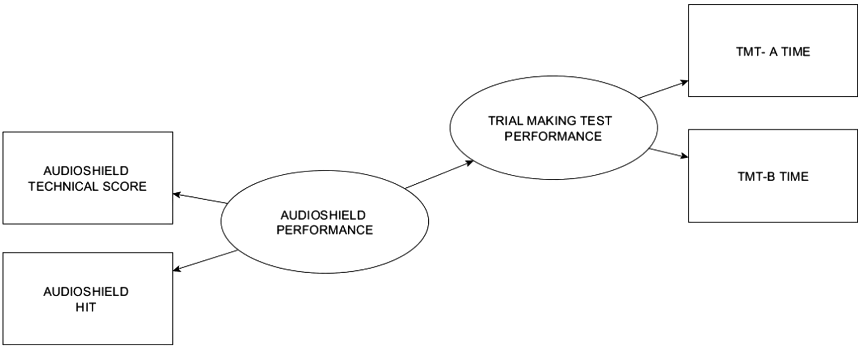

Fig. 1. Conceptual model of relationship among variables.

Hypothesis 1. Performance on the traditional neuropsychological test will correlate significantly with the performance on the virtual reality video game.

Hypothesis 2. The direct path between scores on the virtual reality video game and scores on the traditional neuropsychological test (i.e., time to complete the tasks) will be negative.

2 Materials and Methods

2.1 Participants

38 participants – 13 females (34.2%) and 25 males (65.8%) - were recruited among students and personnel of the University of Milano-Bicocca and of other universities in Milan. Participants' average age was 25.8 ± 4.14 years (min–max = 18–35), average level of education 16 ± 3.05 years (min–max = 11–21) No credits (ECTS) or economic rewards were provided during the research. To be included in the study, individuals had to meet the following criteria: (1) age between 18 and 35 years old; (2) no major medical disorders (heart disease or high blood pressure, neurological disorders, epilepsy); (3) no left-handed; (4) no presence of pharmacotherapy (psychoactive drugs, anti-hypertensive, anti-depressants); (5) no significant visual impairment (all with normal or corrected-to-normal visual acuity); and (6) not having previous experience with the *Audioshield* game. Before participating, all participants were provided with written information about the study and were required to give written consent in order to be included. The Ethical Committee of the University of Milano-Bicocca approved the study. The research was conducted in accordance with the American Psychological Association ethical principles and code of conduct.

2.2 Measures

Psychological Assessment. The following self-report questionnaires were administered to the participants at the start of the experimental session.

Demographics. The participants were asked to indicate their gender (female/male), age (years old), and years of education.

Gaming Habits and Virtual Reality Knowledge. Individuals were asked to indicate their gaming habits (mean hours spent gaming per week), their previous experience with games included in the study (yes/no), and their previous experience with virtual reality (yes/no); furthermore, they were asked to assess their knowledge of virtual reality on a 7-point Likert scale (from "not at all" to "very much").

Cognitive Assessment. To test the participant's executive functions the following traditional neuropsychological tests was used:

Trial Making Test (TMT) [10]. The TMT is one of the most widely used instruments to measure executive functioning in neuropsychological assessment [71, 72]. In this study, we adopted the Italian version included in the ENB-2 battery [73]. The test consists of two parts (A and B) that clinically assess psychomotor speed (Part A) and mental flexibility (Part B) [74]. The individual performance on TMT is given in the form of a score on TMT-A and on TMT-B, which represents the time (in seconds) taken to complete the task.

2.3 Virtual Reality Video Game

In this study, we adopted *Audioshield*, a virtual reality dance game created by Dylan Fitterer for the HTC Vive (Valve) and launched in April 2016. We selected this particular game among other commercial virtual reality games for two main reasons. First, this genre of video games (i.e., dance game), which incorporates both cognitive engagement and physical activity, may have particular benefits for executive functions [75, 76]. Second, this title has some characteristics that could be closely related to executive functions, such as the inhibition of responses and working memory.

Audioshield, in particular, is a dance game in which orbs come flying towards the player, who needs to follow the beat of the music to successfully hit them. The player uses the HTC Vive's handheld motion-sensing controls to operate two shields, blue and red. Red balls must be deflected with the red shield controlled by the right hand while blue balls must be deflected with the left hand. Purple orbs require a combination of both arms. During the game, the color and the direction of the orbs that the player has to hit changes continuously, for example, the blue balls can come from the right (or red from the left) and require the user to respond correctly very quickly.

As detailed later in the procedure, after a brief explanation of the video game by the experimenter, the participants were asked to practice for about 2 min using the song "Engage" (difficulty: normal; shield: gladiator; environment: horizon). Subsequently, the individuals were asked to complete the song "I drop gems" (difficult: normal; shield: gladiator; environment: horizon) while their performance was evaluated. It took about 5 min to complete the gameplay. The dependent variables used to evaluate the video game performance were: (a) the technical score (i.e., a measure of how many orbs players hit, with 10.00 being a perfect technical score with no misses and 0.00 being all missed), (b) the number of balls the player missed, and (c) the numbers of orbs the player hit.

Regarding the hardware, the virtual reality setting adopted in this study included the following units:

- *HTC Vive system (Valve)*, a consumer-grade virtual reality system designed for the use in video games, consists of an HMD, two controllers, and two infrared laser emitter units. The headset covers a nominal field of view of about 110 (approximately 90 per eye) through two 1080×1200 pixel displays that are updated at 90 Hz. Games played with the Vive allow physical movement within a play area that is limited to 4×4 m.
- *A portable computer* (MSI GT73 VR, Intel® Core™ i7 processor, GeForce® GTX 1070 8 GB, 17.3" Full HD 1920×1080 pixel displays).

2.4 Procedure

After individuals gave written informed consent to participate, they completed the self-report questionnaire assessing demographics, gaming habits, and virtual reality knowledge. Subsequently, the participants underwent the neuropsychological evaluation. They were required to complete the paper-and-pencil version of the TMT (Part A and Part B).

Once this phase was completed, the Vive was connected to the PC through a 5 m cable with an HMDI connection, a USB 2 connection, and power. Participants were asked to wear the HMD and were given the HTC Vive controllers. They were then provided with a training period of about 2 min on *Audioshield* using the song "Engage" to familiarize themselves with the video game tasks and controllers. In the next step, individuals completed the song "I drop gems" while their performance was evaluated. The audio level was set to 45 for all participants.

Subjectively, tracking appeared stable when using this configuration, and the video game was playable with no visible tracker artifacts. All measurements were taken in an 8 × 5 m room with a 3.2 m high ceiling lighted by fluorescent lighting, with no reflective surfaces and no exposure to natural lighting. In the center of this room, a 4 × 4 m grid (i.e., the play area) was drawn on the floor using string and chalk, with grid lines drawn 1 m apart. The entire procedure took about 40 min to complete.

2.5 Analytic Strategy and Data Modelling

This research adopted the structural equation modeling (SEM) approach to analyze the cumulative network of relationships among the variables under study. SEM is a widely used quantitative technique for testing models by allowing a simultaneous estimation of a system of regression equations involving multiple endogenous variables as well as their measurement errors [77]. The conceptual model (Fig. 1) was designed by considering game performance (as measured by two observed variables: *Audioshield* technical score and number of hits) and executive functioning [as measured by two observed variables: psychomotor speed (TMT - Part A) and mental flexibility (TMT - Part B)] as an endogenous latent variable.

The Maximum Likelihood method [78] was adopted to determine the parameters for the SEM analysis. The practical and statistical significance of the model was evaluated using the following goodness-of-fit indices: Root Mean Square Error of Approximation (RMSEA, RMSEA < 0.05) [79]; Standardized Root Mean Square Residual (SRMR, SRMR < 0.05) [80]; Normed fit Index (NFI, NFI > 0.95) [81]; Tucker-Lewis Index (TLI, TLI > 0.95) [81]; and Comparative Fit Index (CFI, CFI > 0.95) [81]. All models were tested using Amos software. In keeping with the current literature (e.g., [82]), we estimated confidence limits using both Monte Carlo simulation and bootstrapping methods with a set of random samples (k = 500) and 95% confidence intervals for unstandardized effects.

3 Results

Main descriptive statistics and zero-order correlations of considered variables are summarized in Table 1.

Table 1. Main descriptive statistics and zero-order correlations for TMT scores and *Audioshield* performance (N = 38). Note: * p < .05, ** p < .01 *** p < .001

Variable	1	2	3	4
1. TMT - task A	–			
2. TMt - task B	.390*	–		
3. Audioshield - Score	−.477***	−.417**	–	
4. Audioshield - Hit	−0.241	−.375**	.582***	–
Mean	12.58	59.87	9.71	555.4
Standard Deviation	7.26	20.26	0.211	449.5
Skewness	0.279	1.25	0.135	1.96
Min - Max	18–35	29–117	9–10	97–2,338

The results of the correlational analysis revealed that, in general, game performance was negatively associated with TMT scores (i.e., time to complete the TMT-A and TMT-B). In particular, statistically significant negative correlations were found between *Audioshield* scores and both psychomotor speed (TMT-part A, r = −.477, p < .001) and mental flexibility (TMT-part B, r = −.417, p < .001). Interestingly, the number of *Audioshield* total hits correlated negatively with mental flexibility (TMT-part B, r = −.375, p < .001) whereas it did not correlate with psychomotor speed (TMT-part A, r = −.241, p = n.s.). A more detailed picture of the set of relationships among variables under study can be obtained by estimating the structural equation model (See Fig. 2).

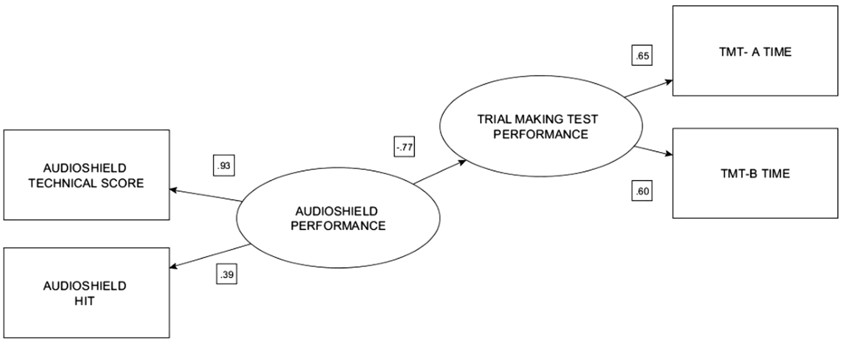

Fig. 2. Results of the structural model estimation. Standardized direct effects were reported, bootstrap estimation (k = 1,500) was conducted along with 95% confidence intervals

The results of the structural equation model revealed an excellent general fit of the conceptual structure with the observed data. In fact, all goodness of fit indexes suggested that the data fit the measurement model well: $\chi^2(1) = 1.41$, p = .235, NC = 1.41, RMSEA = .050, pclose = .257, NFI = .968, NNFI = .988, and CFI = .988, supporting the robust effects among the variables under study both conceptually and statistically. In particular, performance on *Audioshield* was negatively associated with the time to complete the TMT part A and B (β = −.77, $p < .01$, 95% CI [−1.82; −.099]), with higher scores on the virtual reality game indicating less time to complete the tasks. The results support the H2, suggesting that scores on *Audioshield* can be used to reliably estimate the results of TMT.

4 Discussion

Starting from the first hypothesis of the study, the proposed significant correlation between performance on the virtual reality game and scores on the TMT was supported. A traditional paper-and-pencil task was used to assess various cognitive processes, including psychomotor speed and executive functions (i.e., task-switching, flexibility) [74, 83], especially the part B (TMT-B) [71, 84]. In particular, both the time to complete the TMT-A and the TMT-B correlated significantly with the performance on the virtual reality game (i.e., number of orbs hit by the players and technical score on *Audioshield*), suggesting that the virtual reality game may measure the same aspect of cognition as the traditional paper-and-pencil TMT.

Regarding the second hypothesis, as expected, the structural equation model revealed that performance on the virtual reality game can be used to estimate scores on the TMT. In particular, the results showed that performance on *Audioshield* was negatively associated with the time to complete the TMT part A and B, with higher scores on the virtual reality game indicating less the time to complete the tasks.

Our findings are consistent with previous results showing a correlation between the performance on some video games and the scores on the TMT [63, 64] as well as the possibility of predicting scores on this traditional test using individuals' game performance data [62]. For example, it has been recently observed that individuals' game performance data (i.e., latency times, correct answers) on a serious game called *EXPANSE*, which aims to assesses executive functions, correlated strongly with standard task aimed at assessing these functions, including the TMT-A and TMT-B, in a sample of 354 healthy adults [63]. Similarly, another study conducted with 12 patients with Alzheimer's disease, focusing on 9 mild cognitive impairments, reported a correlation between the performance on a serious game board-game named *Kitchen and Cooking* and the TMT [64]. Furthermore, a computational model based on the performance on *Scavenger Haunt*, a computer game designed heavenly inspired by the TMT, was reported to be effective in predicting performance both at the TMT-A and the TMT-B [62].

However, it is important to underline that our study found important differences compared to the research carried out in the past. First, in this study, we tested a commercial video game instead of a serious game [63, 64] or a game-based version of the TMT [62]. In addition, we tested an immersive video game in this experiment while other studies used non-immersive video games (i.e., played on desktops) [62–64]. Compared to previous literature, in fact, the current research is the first to provide preliminary evidence on the effectiveness of immersive commercial video games, a type of video game that combines the advantages of commercial games with those offered by virtual reality [67].

Although the results of the present study could be interesting for their possible applications, this research has some important limitations that could affect the generalizability of the findings. First, it is important to underline that the obtained results refer to a specific virtual reality video game, in this case, *Audioshield*, a dance game launched in April 2016. In the future, it would be interesting to explore this particular genre of games more in-depth, as these games are deemed interesting with respect to the functioning of executive functions [70]. For example, it would be interesting to test *Beat Saber* (Beat Games), since it has very similar characteristics as the game tested in this study, and it was the most successful virtual reality game title of 2018 on the Steam platform (Vive) [85]. In addition, future studies should investigate the effectiveness of other genres of video games played in virtual reality, for example, puzzle and casual games appear to be particularly interesting, since they are genres with potentially very relevant features in relation to the functioning of executive functions [86, 87].

Second, the results emerging from this study refer to the specific virtual reality system in use [i.e., HTC Vive (Steam)]. Future research should investigate the adoption of other commercial virtual reality systems with different characteristics in terms of immersion and interactions with the game, including other off-the-shelf virtual reality systems, such as Oculus Go (Oculus) or Playstation VR (Sony corp.), for instance. These products are very appealing, since unlike the system tested in this study, they could be used very easily even by non-expert operators and are more budget-friendly.

Third, in this study, a specific test was used to measure executive functions (i.e., the TMT). It would be interesting to adopt also other traditional tests, such as the WCST or the Stroop test, that assess different aspects of executive functions. Finally, another limitation of this study is related to the small sample size and the specific sample included in the study comprising young adults who played often (more than 12 h per week) and had a low knowledge of virtual reality (68% of the sample did not try it before the experiment).

5 Conclusion

Our findings show the feasibility of using a commercial virtual reality game to assess executive functions and cognitive abilities, as measured by the TMT. In particular, user performance on *Audioshield* correlated significantly with time to complete the TMT-A as well as the TMT-B. Furthermore, the most important finding of this exploratory study was that the performance on the immersive video game predicted the TMT scores (i.e., time to complete the TMT-A and TMT-B).

Future studies should investigate the possibility of using commercial virtual reality games to assess executive functions as well as of other cognitive skills, since they represent a very promising tool for the neuropsychological assessment. Using off-the-shelves video games rather than video games created ad hoc could have several advantages for the assessment of executive functions as well as other cognitive processes, in particular:

- Low-cost and ready-to-use: Commercial video games are low-cost and their production time is short, unlike serious games and game-based version of the neuropsychological test.
- Advanced graphic quality and gameplay mechanics: Commercial video games are more costly and a greater number of people is involved in its development compared to ad hoc computer games, for example, *Red Dead Redemption 2* and *GTA V* (Rockstar Games) costed about 944 million and 256 million dollars, respectively, involving each one a team of at least 1.000-person [88, 89] to ensure an astonishing quality of the game, which has a significant effect on the overall gaming experience. This would not be achievable with smaller budgets and staff.
- Possibility to collect the data on very large populations: Commercial video games are played in different cultural and social contexts, which makes it easy to replicate studies in different contexts and to collect the data from large samples.

In addition, virtual reality games have also several advantages compared to video games played on desktops, including:

- Require high physical and cognitive involvement: Immersive video games require a high involvement of the player at both the motor and cognitive level, with a greater engagement compared to non-immersive games [67, 69].
- Elicit more intense positive emotions in the players: Video games played in virtual reality elicit a higher self-reported sense of happiness in users in comparison to non-immersive games [67];
- Allow to gather multiple types of data of the user: Video games played in virtual reality allow the collection of a wide variety of data compared to non-immersive games (e.g., scores, time to complete the task, etc.), such as those about the player's movement within the virtual environment.

Acknowledgments. This work was supported by the "Virtual Video Games" project, which is a part of the research plan of the first author, a post-doctoral research fellow at the University of Milano-Bicocca. The authors would like to thank MSI Italy for supporting the study and providing PCs and hardware for data acquisition. In addition, we thank our Ph.D. Ambra Ferrari and post-graduate students Erica Ilari and Gabriele Barone for helping with data acquisition. Last, we would like to thank all the participants for their willingness to participate and enthusiasm.

Author contributions.

FP proposed the study, supervised the scientific asset, and wrote the first draft of the paper. AP analyzed the data. MEM carried out experiments. All three authors were involved in the drafting, revising, and completing the manuscript.

References

1. Müller, U., Kerns, K.: The development of executive function. In: Handbook of Child Psychology and Developmental Science, pp. 1–53. Wiley, Hoboken (2015)
2. Carlson, S.M.: developmentally sensitive measures of executive function in preschool children. Dev. Neuropsychol. **28**(2), 595–616 (2005)

3. Diamond, A.: Executive functions. Annu. Rev. Psychol. **64**(1), 135–168 (2013)
4. Baddeley, A.D., Hitch, G.J.: Developments in the concept of working memory. Neuropsychology **8**(4), 485–493 (1994)
5. Kiesel, A., et al.: Control and interference in task switching—a review. Psychol. Bull. **136** (5), 849–874 (2010)
6. Miyake, A., Friedman, N.P.: The nature and organization of individual differences in executive functions: four general conclusions. Curr. Dir. Psychol. Sci. **21**(1), 8–14 (2012)
7. Bailey, C.E.: Cognitive accuracy and intelligent executive function in the brain and in business. Ann. N. Y. Acad. Sci. **1118**(1), 122–141 (2007)
8. Moffitt, T.E., et al.: A gradient of childhood self-control predicts health, wealth, and public safety. Proc. Natl. Acad. Sci. U.S.A. **108**(7), 2693–2698 (2011)
9. Martínez-Pernía, D., et al.: Using game authoring platforms to develop screen-based simulated functional assessments in persons with executive dysfunction following traumatic brain injury. J. Biomed. Inform. **74**, 71–84 (2017)
10. Reitan, R.M.: Validity of the trail making test as an indicator of organic brain damage. Percept. Mot. Skills **8**(3), 271–276 (1958)
11. Stroop, J.R.: Studies of interference in serial verbal reactions. J. Exp. Psychol. **18**(6), 643–662 (1935)
12. Grant, D.A., Berg, E.: A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. J. Exp. Psychol. **38**(4), 404–411 (1948)
13. Berg, E.A.: A simple objective technique for measuring flexibility in thinking. J. Gen. Psychol. **39**(1), 15–22 (1948)
14. Holtzman, D.M., Morris, J.C., Goate, A.M.: Alzheimer's disease: the challenge of the second century. Sci. Trans. Med. **3**(77), 77sr1 (2011)
15. Ardila, A., Ostrosky-Solis, F., Rosselli, M., Gómez, C.: Age-related cognitive decline during normal aging: the complex effect of education. Arch. Clin. Neuropsychol. **15**(6), 495–513 (2000)
16. Chan, R., Shum, D., Toulopoulou, T., Chen, E.: Assessment of executive functions: review of instruments and identification of critical issues. Arch. Clin. Neuropsychol. **23**(2), 201–216 (2008)
17. Bottari, C., Dassa, C., Rainville, C., Dutil, É.: The criterion-related validity of the IADL profile with measures of executive functions, indices of trauma severity and sociodemographic characteristics. Brain Inj. **23**(4), 322–335 (2009)
18. Burgess, P.W., et al.: The case for the development and use of 'ecologically valid' measures of executive function in experimental and clinical neuropsychology. J. Int. Neuropsychol. Soc. **12**(02), 194–209 (2006)
19. Elkind, J.S., Rubin, E., Rosenthal, S., Skoff, B., Prather, P.: A simulated reality scenario compared with the computerized wisconsin card sorting test: an analysis of preliminary results. CyberPsychol. Behav. **4**(4), 489–496 (2001)
20. DeRight, J., Jorgensen, R.S.: I just want my research credit: frequency of suboptimal effort in a non-clinical healthy undergraduate sample. Clin. Neuropsychol. **29**(1), 101–117 (2015)
21. Jurado, M.B., Rosselli, M.: The elusive nature of executive functions: a review of our current understanding. Neuropsychol. Rev. **17**(3), 213–233 (2007)
22. Parsons, T.D., Carlew, A.R., Magtoto, J., Stonecipher, K.: The potential of function-led virtual environments for ecologically valid measures of executive function in experimental and clinical neuropsychology. Neuropsychol. Rehabil. **27**(5), 777–807 (2017)
23. Alderman, N., Burgess, P.W., Knight, C., Henman, C.: Ecological validity of a simplified version of the multiple errands shopping test. J. Int. Neuropsychol. Soc. **9**(1), 31–44 (2003)

24. Wilson, B.A., Alderman, N., Burgess, P.W., Emslie, H., Evans, J.J.: BADS : behavioural assessment of the dysexecutive syndrome. Harcourt Assessment (1997)
25. Chevignard, M., et al.: An ecological approach to planning dysfunction: script execution. Cortex **36**(5), 649–669 (2000)
26. Valladares-Rodríguez, S., Pérez-Rodríguez, R., Anido-Rifón, L., Fernández-Iglesias, M.: Trends on the application of serious games to neuropsychological evaluation: a scoping review. J. Biomed. Inform. **64**, 296–319 (2016)
27. Burdea, G., Coiffet, P.: Virtual reality technology. Presence Teleoperators Virtual Environ. **12**(6), 663–664 (2003)
28. Schultheis, M.T., Himelstein, J., Rizzo, A.A.: Virtual reality and neuropsychology: upgrading the current tools. J. Head Trauma Rehabil. **17**(5), 378–394 (2002)
29. Rizzo, A.A., Buckwalter, J.G., Neumann, U., Kesselman, C., Thiebaux, M.: Basic issues in the application of virtual reality for the assessment and rehabilitation of cognitive impairments and functional disabilities. CyberPsychol. Behav. **1**(1), 59–78 (1998)
30. Parsons, T.D., Silva, T.M., Pair, J., Rizzo, A.A.: Virtual environment for assessment of neurocognitive functioning: virtual reality cognitive performance assessment test. Stud. Health Technol. Inform. **132**, 351–356 (2008)
31. Korečko, S., et al.: Assessment and training of visuospatial cognitive functions in virtual reality: proposal and perspectivě perspectivě. In: 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2018), pp. 49–43 (2018)
32. Parsey, C.M., Schmitter-Edgecombe, M.: Applications of technology in neuropsychological assessment. Clin. Neuropsychol. **27**(8), 1328–1361 (2013)
33. Weiss, P.L., Kizony, R., Feintuch, U., Rand, D., Katz, N.: Virtual reality applications in neurorehabilitation. In: Selzer, M., Clarke, S., Cohen, L.G., Kwakkel, G., Miller, R. (eds.) Textbook of Neural Repair and Rehabilitation, pp. 198–218. Cambridge University Press, Cambridge (2014)
34. Bohil, C.J., Alicea, B., Biocca, F.A.: Virtual reality in neuroscience research and therapy. Nat. Rev. Neurosci. **12**(12), 752–762 (2011)
35. Raspelli, S., et al.: Validating the neuro VR-based virtual version of the multiple errands test: preliminary results. Presence Teleoperators Virtual Environ. **21**(1), 31–42 (2012)
36. Cipresso, P., et al.: Virtual multiple errands test (VMET): a virtual reality-based tool to detect early executive functions deficit in Parkinson's disease. Front. Behav. Neurosci. **8**, 405 (2014)
37. Nir-Hadad, S.Y., Weiss, P.L., Waizman, A., Schwartz, N., Kizony, R.: A virtual shopping task for the assessment of executive functions: validity for people with stroke. Neuropsychol. Rehabil. **27**(5), 808–833 (2017)
38. Zygouris, S., et al.: Can a virtual reality cognitive training application fulfill a dual role? Using the virtual supermarket cognitive training application as a screening tool for mild cognitive impairment. J. Alzheimer's Dis. **44**(4), 1333–1347 (2015)
39. Jansari, A.S., Froggatt, D., Edginton, T., Dawkins, L.: Investigating the impact of nicotine on executive functions using a novel virtual reality assessment. Addiction **108**(5), 977–984 (2013)
40. Renison, B., Ponsford, J., Testa, R., Richardson, B., Brownfield, K.: The ecological and construct validity of a newly developed measure of executive function: the virtual library task. J. Int. Neuropsychol. Soc. **18**(03), 440–450 (2012)
41. Vourvopoulos, A., Faria, A.L., Ponnam, K., Bermudez i Badia, S.: RehabCity. In: Proceedings of the 11th Conference on Advances in Computer Entertainment Technology - ACE 2014, pp. 1–8 (2014)

42. Davison, S.M.C., Deeprose, C., Terbeck, S.: A comparison of immersive virtual reality with traditional neuropsychological measures in the assessment of executive functions. Acta Neuropsychiatry **30**(02), 79–89 (2018)
43. Entertainment Software Assotiation: Essential facts about the computer and video game industry (2018). http://www.theesa.com/about-esa/essential-facts-computer-video-game-industry/
44. Lorenz, R.C., Gleich, T., Gallinat, J., Kühn, S.: Video game training and the reward system. Front. Hum. Neurosci. **9**, 40 (2015)
45. Green, C.S., Bavelier, D.: Learning, attentional control, and action video games. Curr. Biol. **22**(6), R197–R206 (2012)
46. Gee, J.P., Paul, J.: What video games have to teach us about learning and literacy. Comput. Entertain. **1**(1), 20 (2003)
47. Russoniello, C.V., O'Brien, K., Parks, J.M.: The effectiveness of casual video games in improving mood and decreasing stress. J. Cyber Ther. Rehabil. **2**(1), 53–66 (2009)
48. Przybylski, A.K., Rigby, C.S., Ryan, R.M.: A motivational model of video game engagement. Rev. Gen. Psychol. **14**(2), 154–166 (2010)
49. Granic, I., Lobel, A., Engels, R.C.: The benefits of playing video games. Am. Psychol. **69** (1), 66–78 (2014)
50. Pallavicini, F., Ferrari, A., Mantovani, F.: Video games for well-being: a systematic review on the application of computer games for cognitive and emotional training in the adult population. Front. Psychol. **9**, 2127 (2018)
51. Green, C.S., Bavelier, D.: Action video game modifies visual selective attention. Nature **423** (6939), 534–537 (2003)
52. Clark, J.E., Lanphear, A.K., Riddick, C.C.: The effects of videogame playing on the response selection processing of elderly adults. J. Gerontol. **42**(1), 82–85 (1987)
53. Donchin, E.: The space fortress game. Acta Psychol. (Amst) **71**(1–3), 17–22 (1989)
54. Fleming, T.M., et al.: Serious games and gamification for mental health: current status and promising directions. Front. Psychiatry **7**, 215 (2016)
55. Boyle, et al.: An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. Comput. Educ. **94**(C) 178–192 (2016)
56. Cherney, I.D., Bersted, K., Smetter, J.: Training spatial skills in men and women. Percept. Mot. Skills **119**(1), 82–99 (2014)
57. Parong, J., Mayer, R.E., Fiorella, L., MacNamara, A., Homer, B.D., Plass, J.L.: Learning executive function skills by playing focused video games. Contemp. Educ. Psychol. **51**, 141–151 (2017)
58. Toril, P., Reales, J.M., Mayas, J., Ballesteros, S.: Video game training enhances visuospatial working memory and episodic memory in older adults. Front. Hum. Neurosci. **10**, 206 (2016)
59. Schoene, D., Smith, S.T., Davies, T.A., Delbaere, K., Lord, S.R.: A stroop stepping test (SST) using low-cost computer game technology discriminates between older fallers and non-fallers. Age Ageing **43**(2), 285–289 (2014)
60. Tong, T., Chignell, M.: Developing a serious game for cognitive assessment. In: Proceedings of the Second International Symposium of Chinese CHI on - Chinese CHI 2014, pp. 70–79 (2014)
61. Tong, X., Gromala, D., Amin, A., Choo, A.: The Design of an Immersive Mobile Virtual Reality Serious Game in Cardboard Head-Mounted Display for Pain Management. In: Serino, S., Matic, A., Giakoumis, D., Lopez, G., Cipresso, P. (eds.) MindCare 2015. CCIS, vol. 604, pp. 284–293. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-32270-4_29

62. Hagler, S., Jimison, H.B., Pavel, M.: Assessing executive function using a computer game: computational modeling of cognitive processes. IEEE J. Biomed. Heal. Inf. **18**(4), 1442–1452 (2014)
63. Chicchi Giglioli, I.A., de Juan Ripoll, C.., Parra, E., Alcañiz Raya, M.: EXPANSE: a novel narrative serious game for the behavioral assessment of cognitive abilities. PLoS One, **13** (11), e0206925 (2018)
64. Manera, V., et al.: Kitchen and cooking' a serious game for mild cognitive impairment and Alzheimer's disease: a pilot study. Front. Aging Neurosci. **7**, 24 (2015)
65. Martínez-Pernía, D., González-Castán, Ó., Huepe, D.: From ancient Greece to the cognitive revolution: a comprehensive view of physical rehabilitation sciences. Physiother. Theor. Pract. **33**(2), 89–102 (2017)
66. Lin, J.-H.T.: Fear in virtual reality (VR): fear elements, coping reactions, immediate and next-day fright responses toward a survival horror zombie virtual reality game. Comput. Human Behav. **72**, 350–361 (2017)
67. Pallavicini, F., Ferrari, A., Pepe, A., Garcea, G., Zanacchi, A., Mantovani, F.: effectiveness of virtual reality survival horror games for the emotional elicitation: preliminary insights using resident evil 7: biohazard. In: Antona, M., Stephanidis, C. (eds.) UAHCI 2018. LNCS, vol. 10908, pp. 87–101. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-92052-8_8
68. Riva, G.: Is presence a technology issue? Some insights from cognitive sciences. Virtual Reality **13**(3), 159–169 (2009)
69. Pallavicini, F., et al.: What distinguishes a traditional gaming experience from one in virtual reality? An exploratory study. In: Ahram, T., Falcão, C. (eds.) AHFE 2017. AISC, vol. 608, pp. 225–231. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-60639-2_23
70. Eggenberger, P., Schumacher, V., Angst, M., Theill, N., de Bruin, E.: Does multicomponent physical exercise with simultaneous cognitive training boost cognitive performance in older adults? A 6-month randomized controlled trial with a 1-year follow-up. Clin. Interv. Aging **10**, 1335 (2015)
71. Arbuthnott, K., Frank, J.: Trail making test, part B as a measure of executive control: validation using a set-switching paradigm. J. Clin. Exp. Neuropsychol. **22**(4), 518–528 (2000)
72. Sanchez-Cubillo, I., et al.: Construct validity of the trail making test: role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. J. Int. Neuropsychol. Soc. **15**(03), 438 (2009)
73. Mondini, S.: Esame neuropsicologico breve 2 (ENB-2) : una batteria di test per lo screening neuropsicologico. Raffaello Cortina (2011)
74. Lezak, M.D.: Neuropsychological Assessment. Oxford University Press, Oxford (2004)
75. Anderson-Hanley, C., Maloney, M., Barcelos, N., Striegnitz, K., Kramer, A.: Neuropsychological benefits of neuro-exergaming for older adults: a pilot study of an interactive physical and cognitive exercise system (iPACES). J. Aging Phys. Act. **25**(1), 73–83 (2017)
76. Best, J.R.: Effects of physical activity on children's executive function: contributions of experimental research on aerobic exercise. Dev. Rev. **30**(4), 331–551 (2010)
77. Kline, R.B.: Principles and Practice of Structural Equation Modeling, 5th edn. The Guilford Press, New York (2011)
78. Gath, E.G., Hayes, K.: Bounds for the largest Mahalanobis distance. Linear Algebra Appl. **419**(1), 93–106 (2006)
79. Hu, L., Bentler, P.M.: Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equ. Model. Multidiscip. J. **6**(1), 1–55 (1999)
80. Marsh, H.W., Hau, K.-T.: Assessing goodness of fit. J. Exp. Educ. **64**(4), 364–390 (1996)

81. Morin, A.J.S., Marsh, H.W., Nagengast, B.: Exploratory structural equation modeling. In: Hancock, G.R., Mueller, R.O., Charlotte, N.C. (eds.) Structural Equation Modeling: A Second Course, 2nd edn., pp. 395–436, Information Age Publishing, Inc (2013)
82. Mackinnon, D.P., Lockwood, C.M., Williams, J.: Confidence limits for the indirect effect: distribution of the product and resampling methods. Multivar. Behav. Res. **39**(1), 99 (2004)
83. Salthouse, T.A.: What cognitive abilities are involved in trail-making performance? (2011)
84. Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., Wager, T.D.: The unity and diversity of executive functions and their contributions to complex 'frontal lobe' tasks: a latent variable analysis. Cogn. Psychol. **41**(1), 49–100 (2000)
85. Lang, B.: Valve reveals top selling VR games on steam in 2018, Road to VR (2018). https://www.roadtovr.com/valve-reveals-top-selling-vr-games-on-steam-2018/?platform=hootsuite. Accessed 11 Jan 2019
86. Baniqued, P.L., et al.: Cognitive training with casual video games: points to consider. Front. Psychol. **4**, 1010 (2014)
87. Oei, A.C., Patterson, M.D.: Playing a puzzle video game with changing requirements improves executive functions. Comput. Hum. Behav. **37**, 216–228 (2014)
88. Takahashi, D.: The DeanBeat: How much did Red Dead Redemption 2 cost to make? (updated)—VentureBeat, Venture Beat (2018). https://venturebeat.com/2018/10/26/the-deanbeat-how-much-did-red-dead-redemption-2-cost-to-make/
89. Danielle, T.: GTA V: How Much Did It Cost to Make, Twinfinite (2017). https://twinfinite.net/2017/12/gta-v-cost-make-how-much/