Universidad Politécnica de Valencia

Programa de Doctorado: Tecnologías para la Salud y el Bienestar

Ph.D. Thesis

Novel supports to the assessment of cognitive functions through the combined use of technologies and subjective and objective measurements

Irene Alice Margherita Chicchi Giglioli

<u>Ph.D. Supervisor:</u>

Mariano Alcañiz Raya

<u>Co-Supervisor:</u>

Gabriella Pravettoni

Valencia, September 2019

Abstract

Cognitive functions include all the processes through which an individual perceives, records, maintains, recovers, manipulates, uses and expresses information that are involved in any everyday activity. The traditional cognitive functioning assessment belongs to the applied neuropsychological science that focuses on the evaluation of the specific cognitive components through the recollection of observable behaviors during specific activities. The main standardized tools can be divided in three main groups: short scales of cognitive tracking tests – questionnaires, general neuropsychological batteries, and specific tests. These tools are well-validated and riable but, in the last decade, several research have shown that some patients can perform these neuropsychological tests well, even when they have significant difficulties in adapting their behaviors to daily life activities. This could depend on the fact that neuropsychological tests are too abstract, decontextualized and do not reflect daily life activities.

According to this, more recently, a new approach has substantially increased, potentially providing a higher ecological validity in functional cognitive abilities assessment than standardized approach: the use of advanced technological systems for neuropsychological assessment (ATSNA). ATSNA refer to a set of devices and software applications such as computerized tests, fun and interactive fantasy serious games (SG), and/or simulated virtual (VR) and/or augmented (AR) reality systems that go beyond traditional assessment tests and that supply the possibility to deliver controlled and dynamic stimuli, in ecologically valid, and secure environments. Technologically, ATSNA can be rendered through a non-immersive (2D) screen in which interaction is possible thanks to a keyboard or a mouse or through more immersive head mounted display systems (3D) in which the human eye-gaze and (virtual) hands allow subjects to interact with the synthetic elements. Hence, by developing an ATSNA system that allows subject to become protagonist in an ATS environment, cognitive functions can be assessed as if he/she was evaluated in a real-simulated environment. In fact, although to date several 2D and 3D

ASTNA systems have been developed, there are few studies that have compared 2D vs 3D (virtual systems) and no studies have compared virtual and augmented systems.

Starting from these premises, the main objective of the thesis was to design, develop, and validate a non-immersive 2D SG versus an immersive 3D SG and a dalily life activity in an immersive 3D VR environment versus an AR for the assessment of cognitive functions, comparing the efficacy and effectiveness of them.

In the first two experimental studies, the ATSNA environment was developed applying evidence-center design as model of reference and stealth assessment as method of evaluation. According to this, a narrative storytelling SG, settled in a spaceship, which aim was to discover a new planet where to live because the Earth had become uninhabitable, was created for leading participants in the play. The game focuses on the participant that was the protagonist in the interactive story, driving him/her in the situations and activities. In addition, the participant could explore and navigate in the environment, manipulate and interact with objects. The narrative nature gave the context to the activities to be solved by the participant, and for solving them he/she needed to concentrate, evaluate and decide strategies.

For the validation and the comparison analyses between the two systems, two studies were carried out to determine the efficacy and utility about the performance outcomes and individual variability. The relationships between the variables collected during the performance were also analyzed. The performance in the game was compared with traditional methods of evaluation. The first 2D study involved 354 healthy subjects and correlations were found between the game and traditional methods, suggesting that the game could be a valid tool for assessing cognitive functions in adults. The second study, compared 2D versus a 3D ATSNA version, it involved 94 healthy subjects and showed that 3D version was able to generate lower times and higher correct answers that the 2D, suggesting initial evidence of efficacy of a more immersive system compared to a non-immersive system. Although this result highlights a potential limitation on using different technological systems due to the differences on the two interaction methods (the 2D system applied mouse and keyboard; the 3D two virtual hands' controllers) and hardware and software latency data recording. Regarding individual variability on age, gender, and education, the findings showed consistency with the reference literature. Specifically, younger showed higher performance that older; higher educational levels reflected on a better score and about gender, results showed a more composite panorama.

Furthermore, to enhance the ecological validity of assessment, the last study of this thesis compared the behavioral performance and physiological responses, during an ecological cooking task, between a virtual and an augmented system on 50 healthy subjects. The cooking task consisted of 4 levels that increased in difficulty. As the level increased, additional activities appeared. The behavioral results showed that times are always lower in VR than in AR, increasing constantly in accordance with the difficulty of the tasks. Regarding physiological responses, the findings showed that AR condition produced more individual excitement and activation than VR.

Finally, previously to the studies, a network and cluster analysis on the development of ATSNA applications and the evolution from the beginning to the present has been conducted. The main finding concerned the last five years in which multidisciplinary research, such as in psychology, education, and medicine represent the main fields of ATSNA applications along with the traditional fields of engineering and computer science.

To conclude, ATSNA are proving to be reliable and effective tools for the assessment of cognitive functions in adults, providing more ecological validity and objectivity than traditional methods of assessment. Further researches, including also clinical populations, are research, including also clinical populations, is needed to enhance reliability validity of these tools.

Resumen

Las funciones cognitivas incluyen todos los procesos a través de los cuales un individuo percibe, registra, mantiene, recupera, manipula, usa y expresa informaciones que están involucradas en cualquier actividad cotidiana. La evaluación tradicional del funcionamiento cognitivo pertenece a la ciencia neuropsicológica aplicada que se centra en la evaluación de los componentes cognitivos específicos a través de la recolección de comportamientos observables durante actividades específicas. Las principales herramientas estandarizadas se pueden dividir en tres grupos principales: escalas cortas de pruebas de seguimiento cognitivo (cuestionarios), baterías neuropsicológicas generales y pruebas específicas. Estas herramientas están bien validadas y son confiables, pero, en la última década, varias investigaciones han demostrado que algunos pacientes pueden realizar bien estas pruebas neuropsicológicas, incluso cuando tienen dificultades significativas para adaptar sus comportamientos a las actividades de la vida diaria. Esto podría depender del hecho de que las pruebas neuropsicológicas son abstractas, descontextualizadas y no reflejan las actividades de la vida diaria. Según esto, más recientemente, un nuevo enfoque ha aumentado sustancialmente, lo que podría proporcionar una mayor validez ecológica en la evaluación de las capacidades cognitivas funcionales que el enfoque estandarizado: el uso de sistemas tecnológicos avanzados para la evaluación neuropsicológica (STAEN). STAEN se refiere a un conjunto de dispositivos y aplicaciones de software como pruebas computarizadas, juegos serios (JS) de fantasia divertidos e interactivos, y/o sistemas simulados virtuales (RV) y/o aumentados (RA) que van más allá de las pruebas de evaluación tradicionales y que brindan la posibilidad de proporcionar estímulos controlados y dinamicos en entornos ecológicamente validos y seguros. Tecnológicamente, los STAEN se pueden representar a través de una pantalla no inmersiva (2D) en la que la interacción es posible gracias a un teclado o un mouse o mediante sistemas de visualización montados en la cabeza (3D) más inmersivos en los que la mirada del ojo humano y las manos (virtuales) permiten que los sujetos interactúen con los elementos sintéticos. Por lo tanto, al desarrollar un sistema STAEN que permita al sujeto convertirse en

protagonista en un entorno STAEN, las funciones cognitivas pueden evaluarse como si él / ella fuera evaluado en un entorno real simulado. De hecho, aunque hasta la fecha se han desarrollado varios sistemas STAEN 2D y 3D, hay pocos estudios que hayan comparado 2D vs 3D (sistemas virtuales) y ningún estudio ha comparado sistemas virtuales y aumentados.

A partir de estas premisas, el objetivo principal de la tesis fue diseñar, desarrollar y validar un JS 2D no inmersivo versus un JS 3D inmersivo y una actividad de la vida diaria en un entorno de RV inmersiva versus RA para la evaluación de funciones cognitivas, comparando el eficacia y efectividad de los mismos.

En los primeros dos estudios experimentales, el entorno STAEN se desarrolló aplicando el diseño del centro de evidencia como modelo de referencia y la evaluación oculta como método de evaluación. De acuerdo con esto, se creó un JS narrativo, establecido en una nave espacial, cuyo objetivo era descubrir un nuevo planeta donde vivir porque la Tierra se había vuelto inhabitable, para los principales participantes en la obra. El juego se enfoca en el participante que fue el protagonista en la historia interactiva, conduciéndolo a él en las situaciones y actividades. Además, el participante podría explorar y navegar en el entorno, manipular e interactuar con objetos. La naturaleza narrativa dio el contexto a las actividades a ser resueltas por el participante, y para resolverlas necesitaba concentrarse, evaluar y decidir estrategias. Para la validación y los análisis de comparación entre los dos sistemas, se llevaron a cabo dos estudios para determinar la eficacia y la utilidad sobre los resultados de rendimiento y la variabilidad individual. También se analizaron las relaciones entre las variables recopiladas durante el desempeño de las tareas asignadas a los usuarios. El rendimiento en el juego se comparó con los métodos tradicionales de evaluación. El primer estudio 2D incluyó 354 sujetos sanos y se encontraron correlaciones entre el juego y los métodos tradicionales, lo que sugiere que el juego podría ser una herramienta válida para evaluar las funciones cognitivas en adultos. El segundo estudio, comparó la versión 2D versus una versión 3D STAEN, involucró a 94 sujetos sanos y mostró que la versión 3D fue capaz de generar tiempos más bajos y respuestas correctas más altas que la 2D, lo que sugiere evidencia inicial de la eficacia de un sistema más inmersivo en comparación con un sistema no inmersivo.

Aunque este resultado resalta una posible limitación en el uso de diferentes sistemas tecnológicos debido a las diferencias en los dos métodos de interacción (el sistema 2D aplicó el mouse y el teclado; el sistema 3D, los controladores de dos manos virtuales) y el registro de datos de latencia de hardware y software. Con respecto a la variabilidad individual en edad, género y educación, los hallazgos mostraron consistencia con la literatura de referencia. Específicamente, los más jóvenes mostraron un mayor rendimiento que los mayores; niveles educativos más altos se vieron reflejados en una mejor puntuación y sobre género, los resultados mostraron un panorama más compuesto.

Además, para mejorar la validez ecológica de la evaluación, el último estudio de esta tesis comparó el rendimiento conductual y las respuestas fisiológicas, durante una tarea de cocina ecológica, entre un sistema virtual y un sistema aumentado en 50 sujetos sanos. La tarea de cocinar consistió en 4 niveles que aumentaron en dificultad. A medida que el nivel aumentaba, aparecían actividades adicionales. Los resultados de comportamiento mostraron que los tiempos siempre son más bajos en RV que en RA, aumentando constantemente de acuerdo con la dificultad de las tareas. Con respecto a las respuestas fisiológicas, los hallazgos mostraron que la condición AR produjo más excitación y activación individual que la realidad virtual.

Finalmente, previamente a los estudios, se realizó un análisis de redes y clústeres sobre el desarrollo de aplicaciones STAEN y la evolución desde el principio hasta el presente. El hallazgo principal se refería a los últimos cinco años en los que la investigación multidisciplinaria, como en psicología, educación y medicina, representa los principales campos de las aplicaciones de STAEN junto con los campos tradicionales de la ingeniería y las ciencias de la computación.

Para concluir, los STAEN están demostrando ser herramientas confiables y efectivas para la evaluación de las funciones cognitivas en adultos, proporcionando más validez ecológica y objetividad que los métodos tradicionales de evaluación. Se necesitan más investigaciones, incluidas también las poblaciones clínicas, investigación, incluidas también las poblaciones clínicas, para mejorar la validez de fiabilidad de estas herramientas.

Resum

Les funcions cognitives inclouen tots els processos a través dels quals un individu percep, registra, manté, recupera, manipula, usa i expressa informacions que estan involucrades en qualsevol activitat quotidiana. L'avaluació tradicional del funcionament cognitiu pertany a la ciència neuropsicológica aplicada que se centra en l'avaluació dels components cognitius específics a través de la recol·lecció de comportaments observables durant activitats específiques. Les principals ferramentes estandarditzades es poden dividir en tres grups principals: escales curtes de proves de seguiment cognitiu (qüestionaris), bateries neuropsicológiques generals i proves específiques. Estes ferramentes estan ben validades i són confiables, però, en l'última dècada, diverses investigacions han demostrat que alguns pacients poden realitzar bé estes proves neuropsicológiques, inclús quan tenen dificultats significatives per a adaptar els seus comportaments a les activitats de la vida diària. Açò podria dependre del fet de que les proves neuropsicológiques són abstractes, descontextualitzades i no reflectixen les activitats de la vida diària. Segons açò, més recentment, un nou enfocament ha augmentat substancialment, la qual cosa podria proporcionar una major validesa ecològica en l'avaluació de les capacitats cognitives funcionals que l'enfocament estandarditzat: l'ús de sistemes tecnològics avançats per a l'avaluació neuropsicológica (STAAN). STAAN fa referència a un conjunt de dispositius i aplicacions software com ara proves computeritzades, jocs seriosos de fantasia (SG) divertits i i interactius, y/o sistemes de realitat virtual (VR) y/o augmentada (AR) simulats que van més enllà que les proves tradicionals d'avaluació i que proporcionen la possibilitat de mostrar estímuls controlats i dinàmics, en entorns ecològicament vàlids i segurs. Tecnològicament, els STAAN es poden representar a través d'una pantalla no inmersiva (2D) en la que la interacció és possible gràcies a un teclat o un ratolí o per mitjà de sistemes de visualització muntats en el cap (3D) més inmersivos en els que la mirada de l'ull humà i les mans (virtuals) permeten que els subjectes interactuen amb els elements sintètics. Per tant, al desenrotllar un sistema STAAN que permeta al subjecte convertir-se en protagonista en un entorn STAAN, les funcions cognitives poden avaluar-se com

si ell / ella fora avaluat en un entorn real-simulat. De fet, encara que fins a la data s'han desenrotllat diversos sistemes STAAN 2D i 3D, hi ha pocs estudis que hagen comparat 2D vs 3D (sistemes virtuals) i cap estudi ha comparat sistemes virtuals i augmentats.

Partint d'estes premisses, l'objectiu principal de la tesi era ser dissenyar, desenrotllar i validar un 2D SG no inmersiu enfront d'un 3D inmersiu i una activitat diària en un entorn 3D VR inmersiu enfront d'una en AR per a l'avaluació de funcions executives, comparant l'eficiència i eficàcia de cada un d'ells. En els primers dos estudis experimentals, l'entorn STAAN es va desenrotllar aplicant el disseny del centre d'evidència com a model de referència i l'avaluació oculta com a mètode d'avaluació. D'acord amb açò, es va crear una JS narrativo, establido en una nau espacial, l'objectiu de la qual era descobrir un nou planeta on viure perquè la Terra s'havia tornat inhabitable, per als principals participants en l'obra. El joc s'enfoca en el participant que va ser el protagonista en la història interactiva, conduint-li en les situacions i activitats. A més, el participant podria explorar i navegar en l'entorn, manipular i interactuar amb objectes. La naturalesa narrativa va donar el context a les activitats a ser resoltes pel participant, i per a resoldre-les necessitava concentrar-se, avaluar i decidir estratègies. Per a la validació i les anàlisis de comparació entre els dos sistemes, es van dur a terme dos estudis per a determinar l'eficàcia i la utilitat sobre els resultats de rendiment i la variabilitat individual. També es van analitzar les relacions entre les variables recopilades durant léxercici de les tasques assignades als usuaris. El rendiment en el joc es va comparar amb els mètodes tradicionals d'avaluació. El primer estudi 2D va incloure 354 subjectes sans i es van trobar correlacions entre el joc i els mètodes tradicionals, la qual cosa suggerix que el joc podria ser una ferramenta vàlida per a avaluar les funcions cognitives en adults. El segon estudi, va comparar la versió 2D versus una versió 3D STAAN, va involucrar a 94 subjectes sans i va mostrar que la versió 3D va ser capaç de generar temps més baixos i respostes correctes més altes que la 2D, la qual cosa suggerix evidència inicial de l'eficàcia d'un sistema més inmersiu en comparació amb un sistema no inmersiu. Encara que este resultat ressalta una possible limitació en l'ús de diferents sistemes tecnològics a causa de les diferències en els dos mètodes d'interacció (el sistema 2D va aplicar el ratolí i el teclat; el sistema

3D, els controladors de dos mans virtuals) i el registre de dades de latència de hardware i software. Respecte a la variabilitat individual en edat, gènere i educació, les troballes van mostrar consistència amb la literatura de referència. Específicament, els més jóvens van mostrar un major rendiment que els majors; nivells educatius més alts es van veure reflectits en una millor puntuació i sobre gènere, els resultats van mostrar un panorama més compost.

A més, per a millorar la validesa ecològica de l'avaluació, l'últim estudi d'esta tesi va comparar el rendiment conductual i les respostes fisiològiques, durant una tasca de cuina ecològica, entre un sistema virtual i un sistema augmentat en 50 subjectes sans. La tasca de cuinar va consistir en 4 nivells que van augmentar en dificultat. A mesura que el nivell augmentava, apareixien activitats addicionals. Els resultats de comportament van mostrar que els temps sempre són més baixos en RV que en RA, augmentant constantment d'acord amb la dificultat de les tasques. Respecte a les respostes fisiològiques, les troballes van mostrar que la condició ARA va produir més excitació i activació individual que la realitat virtual.

Finalment, prèviament als estudis, es va realitzar una anàlisi de xarxes i clusters sobre el desenrotllament d'aplicacions STAAN i l'evolució des del principi fins al present. La troballa principal es referia als últims cinc anys en què la investigació multidisciplinària, com en psicologia, educació i medicina, representa els principals camps de les aplicacions de STAAN junt amb els camps tradicionals de l'enginyeria i les ciències de la computació.

Per a concloure, els STAAN estan demostrant ser ferramentes confiables i efectives per a l'avaluació de les funcions cognitives en adults, proporcionant més validesa ecològica i objectivitat que els mètodes tradicionals d'avaluació. Es necessiten més investigacions, incloses també les poblacions clíniques, investigació, incloses també les poblacions clíniques, per a millorar la validesa de fiabilitat d'estes ferramentes.

Thesis structure

The thesis manuscript is structured as follows:

Part I introduces main topics of the thesis and includes a state of the art of the literature and the scientific objectives of the research carried out and reported in the Part II.

Part II encloses a selection of the most illustrative papers supporting the thesis objectives, which were published in journals in JCR and conferences. More in detail, it includes three journal papers and one conference paper:

- Paper I. The Past, Present, and Future of Virtual and Augmented Reality Research: a network and cluster analysis of the literature. It describes the state of the art in the development and efficacy testing of ATS from its first application until now, the field and research evolution in accordance with the journal and conference publications through a network and cluster analysis.
- Paper II. EXPANSE: a novel narrative serious game for the behavioural assessment of cognitive abilities. It describes the first 2D serious game development and the first experimental study of the thesis for assessing cognitive functions. The performance in the serious game and traditional methods were compared.
- Paper III. Are 3D virtual environments better than 2D interfaces in serious games performance? An explorative study for the assessment of executive functions. It describes the second experimental study of the thesis for assessing cognitive functions. The 2D serious games were compared to a 3D version of the serious game and the performance in the two serious games and traditional methods were compared.
- Paper IV. Individuals' variables in cognitive abilities using a narrative serious game. It describes the third experimental study of the thesis for assessing cognitive functions and includes the analyses of the individual variables on the performance in the serious game.
- Paper V. A virtual versus an augmented reality cooking task based-tools: a behavioral and physiological study on the assessment of executive functions. It is the last experimental study

of this thesis and it compares the behavioural performance on two ASTNA interfaces (VR and AR) using an ecological cooking task.

Part 3 discusses the results of the thesis, summarizes the work with the general conclusions and future works, and enumerates the publications derived from this thesis.

Contents

Part I

Introduction

Introduction

Cognitive functions

Cognitive functions include all the processes through which an individual perceives, records, maintains, recovers, manipulates, uses and expresses information and are involved in any task we face, from the simplest to the most complex (Lezak, 2004; Luria, 1966; 1973; Robert et al., 1998; Stuss & Knight, 2013). For example, the simple daily action of "answering the phone" involves different cognitive abilities: perception (hearing the phone ringing), attention (isolating the ring from all the other sounds), decision-making (answering or not), motor skills (picking up the phone), language skills (speaking and understanding what we are told), and social skills (interpreting the tone of voice and interacting appropriately).

More in detail:

- Perception refers to the ways in which information acquired from the environment through the sense organs (sight, touch, hearing, etc.) is transformed into an experience of objects, events, sounds, tastes, flavors, etc.
- Recognition is the ability to categorize an object (for example, when we see a blackbird we recognize it as belonging to the category "birds"), but also need to associate the object with representations that describe its use (when I recognize an object as a "pen" the knowledge that I have on pens, for example, I know that "need to write").
- Attention is the function that allows the individual to filter and process information or input from the external environment. It can be decomposed into several components: a) divided attention when the subject performs multiple tasks simultaneously; b) sustained attention (vigilance) when we must pay attention to a source of information for a prolonged time; c) selective attention when different stimuli arriving simultaneously, select those to pay attention to without being distracted by interfering stimuli; think for example when we are at a party and we are talking with a friend, meanwhile there are

many auditory stimuli (music, other people talking), but we are able to select and analyze only those from the person with whom we are conversing (cocktail party phenomenon).

- Memory is the ability to store new memories and retrieve them at a distance of time. There are several types of memory: a) long-term memory has unlimited capabilities and can store information for extremely long periods. It is the declarative memory that concerns all the explicit knowledge we have about the world and is subdivided into semantic memory (everything I know about lions) and episodic, that is linked to episodes of our life; procedural memory, on the other hand, consists of knowledge that cannot be expressed in words (such as riding a bike, driving a car, etc.); b) Working memory is a system for temporary retention and active information manipulation (for example, when we make a mental calculation, when we keep a telephone number in mind while typing it, etc.).
- Motor skills, that is the ability to move muscles and body and the ability to manipulate objects
- Language, therefore the ability to translate sounds or words into words (comprehension) and the ability to generate verbal or written output (the production of language).
- Executive functions. They are superior cognitive processes that supervise, direct and control the basic functions (perception, motor control, language, etc.) in the guide of behavior directed to a purpose. They include a series of processes and skills, including: the ability to plan a complex action in view of a goal, following precise and orderly steps and monitoring its execution; the cognitive flexibility, through which we are able to pass quickly from one activity / task to another, as we are able to change the strategies implemented in the resolution of a task, based on information coming from the environment; decision-making processes; the regulation of emotions and behavior, thanks to which we are able to modify and inhibit certain behaviors based on the context in which we find ourselves and finally motivation (Denkla, 1996; Diamond, 2013).

The executive functions coordinate all basic cognitive skills and are involved in the complex activities that we do every day (cooking, planning the day), but they come into play especially when there is a new or unexpected situation to face, when we have to find new ones solutions and develop new strategies (for example, a new task entrusted to us at work).

Each cognitive skill can involve a specific area of the brain or be the result of a network of connections between different brain areas, mainly associated to the prefrontal and parietal regions (Fig. 1) (Levin et al., 1991; Stuss & Alexander, 2000). Brain injuries or dysfunctions due to pathologies, traumas or other neurological and psychological conditions may involve very limited or larger areas and consequently affect specific individual cognitive functions or multiple functions (Diamond, 2005; Goldstein, 1944; Luria, 1969).

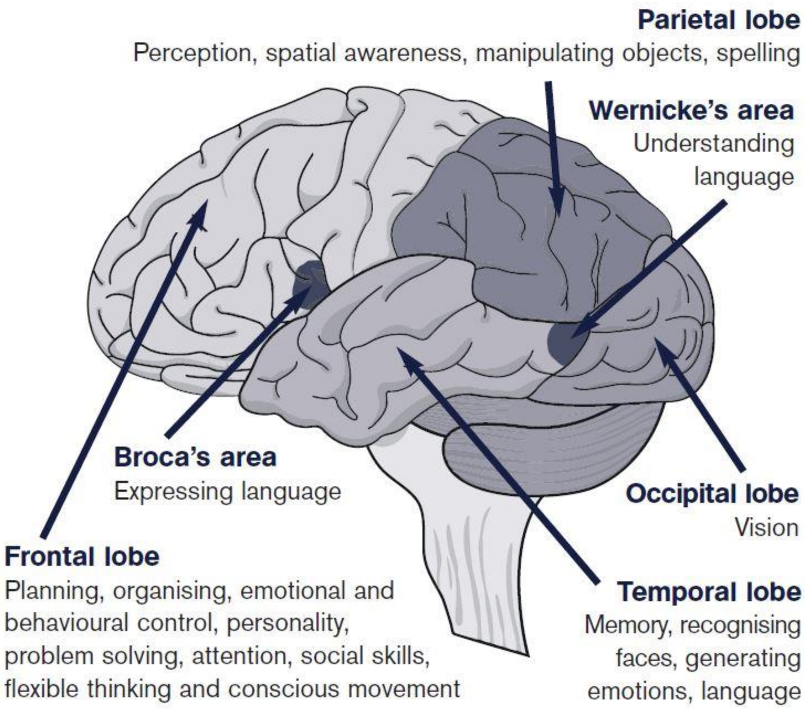


Figure 1. Cognitive brain area.

(Source of Andrew Taylor from https://www.headway.org.uk/about-brain-injury/individuals/effects-of-brain-injury/executive-dysfunction/)

Cognitive functioning assessment

Cognitive functioning assessment belongs to the applied neuropsychological science that focuses on the evaluation of specific cognitive components through the recollection of observable behaviors during specific activities (Denkla, 1994; Lezak, 2004). More in detail, the assessment of cognitive functioning includes the study of the general intellectual performance, attention, the processing speed of the information, the ability to learn and memory, perceptual abilities and motor, language and communication, the reasoning, the ability to solve problems and other functions associated, as mentioned in the previous paragraph, with the frontal brain areas.

Cognitive functioning assessment is performed through reliable and valid standardized tools, which they can be divided in three main groups:

a) Short scales of cognitive tracking tests – questionnaires;

b) General neuropsychological batteries;

c) Specific tests.

Short scales of cognitive tracking tests - questionnaires

These tests are easy to apply requiring little time for its application (from five to twenty minutes). They consist of a set of questions related to a certain number of cognitive areas (temporal and spatial orientation, attention and concentration, learning and memory, language, visuospatial abilities, etc.). Most of them have been originally designed for the evaluation of cognitive deficits in elderly patients, although they have been applied in a general way to patients with all types of cognitive deficits, both acute and chronic. The overall score obtained allows get a "cut-off point" that is used as a dichotomous distinction between a normal and pathological functioning, pointing out those individuals that require a more detailed neuropsychological evaluation (Barratt, 1959; Derryberry & Reed, 2002; Martin & Rubin, 2002). For example, the Attentional Control Scale (Derryberry & Reed, 2002) is a questionnaire, including 20-items rated on a 4 point Likert scale (1 =almost never to 4 = always) that assesses attentional control and shifting. Questionnaire validation showed an internal consistency of α = 0.88 and items test-retest reliability from 0.45 to 0.73 and 0.61 for the total score (Derryberry & Reed, 2002; Fajkowska & Derryberry, 2010).

A relevant limitation of these tests is the lack of sensitivity to detect focal cognitive deficits. Indeed, patients with frontal lesions or with injuries in the right hemisphere can get maximum scores in the most of the items. In addition, also specificity is very low, even in cases where the profile of the different sections of the tests is assessed. Therefore, these tests can provide a general assessment useful, for example, in the patient follow-up to over time and establish correlations between this global score and other relevant variables.

General neuropsychological batteries

A battery of neuropsychological evaluation can be defined as a set of tests or elements that explore the main cognitive functions in a systematized way, in order to detect and typify the existence of brain damage (Folstein et al., 1975). For example, the Wechsler Adult Intelligence Scale – Fourth edition (WAIS) is a complete neuropsychological battery well validated and used. It includes four index scores divided in ten subtests: working memory index, verbal comprehension index, perceptual reasoning index, and the processing speed index. Specifically, the working memory index tests the ability to keep many information in mind at once, such as to repeat a series of numbers backwards after hearing the prompt forward (Wechsler, 2008). There is a great controversy in the specialized literature about the advantages and limitations of the application of this type of procedures in the evaluation of patients with injury or possible cerebral dysfunction (Burgess et al., 1998; 2006; Chaytor et al., 2006; Parsons et al., 2017; Tirapu-Ustárroz et al., 2002). The main advantages refer to the possibility of studying the main syndromes and neuropsychological alterations in a relatively short time, the opportunity to have a broad database that facilitates, on the one hand, obtaining profiles that characterize different injuries of the brain (for example, ischemic accidents or hemorrhages in different vascular areas), and on the other hand, greater control over a set of socio demographic variables (age, educational level, gender, etc.) that affect the performance of individuals in these tests. Regarding this last consideration, normative data on neuropsychological tests are adjusted according to socio demographic variables due to the relevance of these on cognitive performance. Indeed, several studies reported that age and educational level have a significant influence on cognitive functioning. The results on age studies are consistent showing that older age obtained lower scores than younger (Bravo & Hébert, 1997; Moraes et al., 2010; Rossetti et al., 2011). Regarding educational level the literature revealed a homogenous panorama in which lower educational levels is associated to a worst performance than highly educated individuals (Bravo & Hébert, 1997; Moraes et al., 2010; Rossetti et al., 2011). Regarding gender, the scientific panorama is more controversial and some studies suggested that gender has an influence in the cognitive performance and in accordance with the specific task the performance differ from women to men, and others that not support this hypothesis (Bagherpoor & Akbar, 2014; Morgado et al., 2019; Van Hooren et al., 2007; Zarghi & Zarindast, 2011

Finally, this type of global assessment of cognitive functioning allows identifying not only the main deficits but also the skills preserved in each patient. Although, the main limitation is the lack of substantiation theoretical, because in general they constitute a group of tests more or less sensitive to the effects of different brain injuries, but they lack a conceptual framework that explains this selection of instruments (Tirapu-Ustárroz et al., 2002). Furthermore, in the clinical field, its use allows for more comparison of results between individuals and groups than the specific analysis of the errors committed by each patient, which is essential for the establishment of a subsequent personalized neuropsychological rehabilitation program.

Specific Tests

Specific tests for the assessment of executive functions have proved sensibility and effectiveness in detecting patients' dysfunctions in the prefrontal brain areas. The main measures include the Trail making Test, form A and B, for assessing processing speed of the information and cognitive flexibility (Reitan, 1958), the Winsconsin Card Sorting Test (WCST) for cognitive flexibility evaluation (Berg, 1948; Grant & Berg, 1948; Heaton, 1981), the Tower of London able to assess planning ability (Culberston & Zillmet, 1999), the Stroop Test for the assessment of attention and inhibition, etc. (Stroop, 1935). The main limitation of these tests refers to the lack of normative data and their uses allow comparing performance between groups (patients versus cognitively healthy individuals). Table 1 resume the main instruments used for the cognitive functioning assessment.

Table 1

Main Cognitive Tests for neuropsychological evaluation.

Tests	**Cognitive functions**
• Mini Mental State Examination (MMSE)	Cognitive functions' screening including: - Orientation to time and to place - Attention - Short-term memory registration and recall - Language - Visuo-spatial skills
• Montreal Cognitive Assessment (MoCA)	Cognitive functions' screening including: - Short-term memory recall/working memory - Visuo-spatial skills - Executive functions - Attention - Language - Abstract reasoning - Orientatio to time and place
• Continuous Performance Test (CPT) • WAIS IV (Cancellation, Symbol Search, and Coding) • Trail Making Test (Form A and B) • Dot Probe Task	Attention/ Processing speed of information
• Go-No-Go	Inhibition
• Complex figure of Rey test	Visuospatial abilities
• WAIS IV (Arithmetic, Digit Span, and Letter-Number subtests)	Memory
• WAIS IV (Comprehension and Similarities subtests	Reasoning
• Trail Making Test (Form B) • Stroop Test • Winsconsin Card Sorting Test (WCST)	Executive functions

In general, it is difficult to establish a standardized protocol for neuropsychological exploration, since the selection of tests depends not only on their sensitivity to the detection of brain damage, but also by the cognitive status of each patient. Furthermore, in the last decade several research have shown that some patients with well-identified prefrontal lesions perform these neuropsychological tests well, even when they have significant difficulties in adapting their behaviors to daily life activities (Burgess et al., 1998; 2006; Chaytor et al., 2006; Chevignard et al., 2000; Manchester et al., 2004). This could depend on the fact that neuropsychological tests are abstract, decontextualized and do not reflect daily life activities. Furthermore, these tests are subjective performance measures that can affect the objectivity of the results mainly due to: a) the test scores depend on the subjective interpretation and inferences by expert and/or clinician; b) the social desirability bias in which individuals tend to answer to questionnaires in a way that it is viewed favorably by others.

According to these, there is a consensus on the need for procedures that have an ecological validity and orientation in the assessment of the impact that cognitive deficits could generate on the functional aspects of daily life. Before addressing the issue of ecological validity, the following paragraph will define to what extent a test is reliable and valid.

Reliability and validity of a psychological test

A psychological test aims to evaluate specific theoretical constructs, such as attention or planning abilities and to be accurately applied and interpreted, has to comply with two statistics and psychometrics proprieties: reliability and validity (Trochim & Donnelly, 2005).

Reliability

Reliability refers to the consistency of a measure to generate similar results under consistent conditions (Trochim & Donnelly, 2005; Trochim, 2006). Three types of consistency can be identified:

a) Test-retest reliability refers to the consistency of a measure across time and the assessment requires using it on a group of people at one time, and repeating it again on the same group of people at a later time, looking test-retest Pearson's *r* correlations between the two set of scores. If the test-retest correlations are high, the measure can be considered consistent over time (Carmines & Zeller, 1979).
b) Internal consistency refers to the consistency across the items on a multiple-item measure. In a psychological test, all the items assume to reflect the same construct, and for the internal consistency, the scores items should correlate with each other. Statistically, the

Cronbach's α is the most common measure to compute for internal consistency and a value of +.80 or higher shows a good internal consistency (Eisinga et al., 2013).

c) Inter-rater reliability refers to the consistency across researchers/experts judgments. For example, in the creation of a questionnaire for attentional abilities, seven experts have to rate each item in accordance with the theoretical construct, and the ratings should highly correlated with each other. Higher correlations between experts' ratings means a good inter-rater reliability (Durand & Barlow, 2012).

Validity

Validity refers to the propriety of a test to measure what it claims to measure (Carmines & Zeller, 1979). There are various types of validity that can be estimated including:

a) Content validity refers to the relationship between items and the construct of reference. Generally, it is estimated by gathering the responses of a group of experts that evaluate items in accordance to the related construct;

b) Criterion validity is the extent to which scores on a measure are correlated with other variables (that are the criteria) that one would expect them to be correlated with. When the criterion is measured at the same time as the construct and including other measures of the same construct, criterion validity is referred to as concurrent validity; if the criterion is measured in the future, after the construct has been measured, it is referred to as predictive validity; finally, if criteria also include other measures of the same construct, it is referred to the convergent validity;

c) Discriminant validity is the extent to which scores on a measure are not correlated with measures of variables that are theoretically different.

Finally, as mentioned above, another relevant propriety of a measure is represented by the ecological validity that refers to the test ability to predict real-life functional behaviors (Burgess et al., 2006; Chaytor et al. 2003).

Ecological validity

Ecological validity refers to the ability of a test results' to be generalized to situations and activities of daily living (Brunwick, 1955). Specifically in neuropsychological assessment, the term highlights the importance to generate measures that are similar to real activities, predicting the functional relation between individual performance on traditional tests and behaviors in real life.

According to ecological validity definition, Franzen and Wilheim (1996) emphasized that the developments of ecologically valid neuropsychological tests depend on two requirements: verisimilitude and veridicality. Verisimilitude regards the similarity between test and every day activities demand. According to this approach, some neuropsychological tests have been developed, such as the Test of everyday attention (TEA; Robertson, 1994), the Behavioral Assessment Dysexecutive Syndrome (BADS; Wilson et al., 1996), the Cambridge Test of Prospective Memory (CAMPROMPT; Wilson, 2005). These tests tests are able to identify individuals with limited functional skills rather than to discriminate between patients and healthy individuals. The limitations refer to the ability of these tests to assess performance during real-life tasks. For example, the BADS consists of six subtasks, which assess everyday cognitive, emotional and behavioral performance, and one questionnaire (DEX) that showed some limitations in measuring daily life performance.

Veridicality refers to the extent to which the performance on standardized assessment measures are statistically related to the scores on measures that assess and predict daily life performance.

Findings have showed relationships between the two measures but the theoretical relationship with cognitive abilities was not evident. Chaytor and Schmitter-Edgecombe (2003) reviewed and compared the effectiveness of the two approaches in predicting everyday functional performance, including a variety of clinical sample (brain injury, frontal lobe damage, stroke, etc.) and healthy participants. The results showed that verisimilitude was higher and more consistent related to everyday cognitive performance than standardized tests and veridicality.

Until now, traditional neuropsychological measures, like the WCST, TMT, Stroop, and the Tower of London, were initially developed for the cognitive assessment of healthy individuals and based on the theoretical construct and only later were used in clinical populations (Burgess et al., 2006). As mentioned above, these measures are too abstracts and for example, the WCST is a measure to assess cognitive flexibility in which to participants are showed a number of stimulus cards and they are instructed to match these cards to a target card that after a number of trials change the criteria. In clinical assessment the expert neuropsychologist should infer and predict that a patient's inability to correctly match cards to the target card means that the patient have an impairment on cognitive flexibility in daily living, such as on cooking. In accordance with verisimilitude and veridicality requirements, a more ecological design approach should take into considerations real life situations, which can be representative of real life functioning (Burgess et al., 2006). Starting from these considerations, more functional assessment tests have been developed. For example, The Multiple Errands Test (MET) developed by Shallice and Burgess (2011) is a multitasking settled in a hospital or in a shopping mall in which participants should complete various tasks, such as buying items, moving to specific places, following rules and predefined times. The evaluation (correct answers, number and type of error, omissions) is reported by the experts' observations. The MET showed to be more able to detect and predict behavioral impairments in everyday living situations than traditional assessment (Alderman et al., 2003). However, the real life settings, time-required, and high costs represent the most limitations of this test that don't allow to replicate and standardize the results (Logie et al., 2011; Rand et al., 2009).

More recently, a new approach has substantially increased, potentially providing a higher ecological validity in functional cognitive abilities assessment than standardized approach: the use of advanced technological systems for neuropsychological assessment (ATSNA). ATSNA refers to computerized devices that supply the possibility to deliver simulated real activities in ecologically valid, controlled, dynamic, and secure environments (Bohil et al., 2011). Furthermore, via ATSNA systems is possible to keep control on veridicality of measures and on the verisimilitude of naturalistic observation of real life situations (Matheis et al., 2007; Jovanovski et al., 2012a, b).

Advanced technological systems for neuropsychological assesment

ATSNA refer to a set of devices and software applications such as computerized tests, serious games, virtual and augmented systems that go beyond traditional assessment tests and that actually, are the leading edge of practice and research (Cipresso et al., 2018).

In the last two decades, the uses of ATSNA have experienced a significant increase in psychology, improving and facilitating both neuropsychological assessment and therapy (Chicchi et al., 2015; Negut et al., 2016; Ventura et al., 2018; Cipresso et al., 2018). In a recent network and cluster analysis of the literature on the research fields using ICT tools (Cipresso et al., 2018), psychology is at third place (8.21% of the total production from 1900 to 2016), immediately after computer science (42.15%) and engineering (28.66%). Looking at the last five years (2011-2016), neurosciences and neurology are in third place with 11.10% of the total production, and psychology at fourth place (9.32%) with an increment over the total period, as well as rehabilitation (5.54%) and clinical neurology (4.42%), which over the total period did not appeared. The next three sections describe in detail the main tools, from the least to the most ecologically valid (Figure 2) tool – computerized tests, serious games (SG), virtual reality (VR) and augmented reality (AR) and applications that have been developed and tested for the assessment of cognitive functioning, showing effectiveness and sensitivity results. Finally, the fourth section discuss a novel methodological approach, called “stealth assessment”, for the development of ecological virtual environments based on the evidence-center design as a model of reference (ECD) (Mislevy, 1994; Mislevy et al., 2003; Shute & Ventura, 2013).

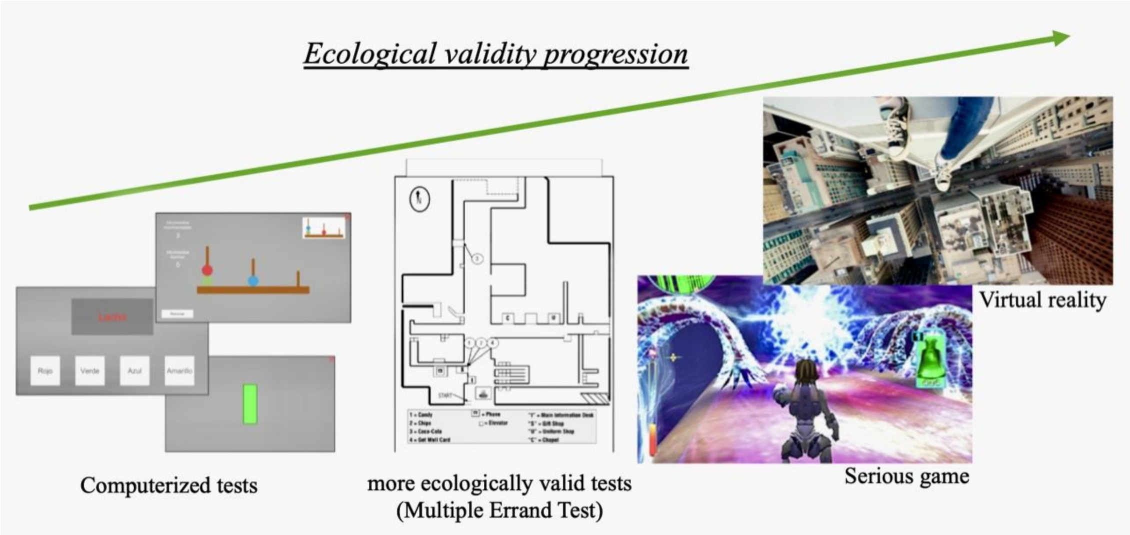


Figure 2. Ecological validity progression according to the ATSNA tools.

ATSNA: Computerized Tests

Since the 80s, several computer-based versions of traditional neuropsychological batteries and specific tests are been developed and tested. Some examples include the Wisconsin Card Sorting Test (Fortuny & Heaton 1996; Heaton 1999), the Raven's Matrices (Calvert & Waterfall 1982; French & Beaumont 1990; Knights et al. 1973; Rock & Nolen 1982; Waterfall 1979), Digit symbol subtests of the WAIS (Gilberstadt et al., 1976). Initial comparison studies on healthy subjects showed significant test-retest reliabilities and no significant difference between the two measures, suggesting equivalence of these measures to traditional tests (Eckerman et al., 1985; Mead & Drasgow, 1993). These first initial evidence highlighted some relevant advantages on a) stimuli presentation, b) scoring, c) normative data, and d) impact on participants (see Table 2).

Table 2

Computer-based neuropsychological assessment: advantages (adapted from Parsons, 2016)

Advantages	
Stimuli presentation	• Enhanced control over administration and scoring • Increased accuracy of timing presentation • Automatic randomization of test trials • Alternate forms and adaptive testing protocols

	• Ability to set accurate basal and ceiling levels and subsequently discontinue the test • Ease of administering tests in different languages • Capacity to rapidly test many persons • Reduced assessment times (adaptive testing) • Administer tests on portable devices (smartphones or handheld computers)
Scoring	• Increased accuracy of measurement/logging of response latency, strength, and variability • Ability to integrate and automate interpretive algorithms such as decision rules for determining impairment or statistically reliable • Ability to measure performance on time-sensitive tasks (e.g., reaction time)
Normative Data	• Enhanced normative data collection and comparison • Ease of exporting responses for data analytic purposes • Automated data exporting for research purposes
Impact on participant	• May increase openness and engagement of respondents • Decrease in examiner influence on responses • May increase accessibility and availability of neuropsychological services

However, testing them on different clinical populations raised some disadvantages involving, for example, lack of equivalence with traditional tests probably due to the different levels of technology familiarity, and lack of uniformity among the measures used due to the large amount of measures available. Specifically, computer-based neuropsychological assessment are been mainly proved on dementia (Cummings et al., 2012), attention deficit hyperactivity disorder (Nigg, 2005; Sonuga-Barke et al., 2008), obsessive-compulsive disorder (Martoni et al., 2015), showing strengths on control and standardization over administration and scoring, sensitivity and test-retest reliability (see Table 3 on aging studies) and disadvantages on stimuli presentation, scoring, normative data, and impact on participants (for more details see Table 4).

Table 3

Computerized assessment on aging (adapted from Parsons, 2016).

Authors	*Hardware*	*Cognitive functions*	*Strenghs*	*Weaknesses*
Doninger et al. (2009)	PC Web-based	• Attention • Memory • Executive function • Verbal Fluency • Visuospatial • Motor • Information processing	Significant differences in memory and executive functions between Mild Cognitive Impairment patients (MCI) and healthy controls	Length of time required for completion
Elwood (2001)	PC	• Attention • Memory • Reaction time • Spatial • Reasoning	Significant correlations with WAIS	Significant anxiety/ frustration in cognitively impaired subjects
Gualtieri (2005)	PC	• Attention • Memory • Psychomotor speed • Processing Speed • Cognitive Flexibility	Significant differences between healthy controls and MCI, and between MCI and mild AD	Lack of normative data
Li et al. (2013)	PC/Tablet – web-based	• Attention • Memory • Executive function • Language • Social cognition	High test–retest reliability and stability in all groups and adequate to detect AD-related cognitive	Difficult to distinguish early MCI from healthy controls without several

			impairment	rounds of administration
Saxton et al. *(2009)*	Tablet	• Attention • Memory • Executive function • Processing speed	High sensitivity to detect MCI	No normative data
Tornatore *(2005)*	PC	• Memory • Executive function • Language	High correlation with traditional tests	No normative data

Table 4

Computer-based neuropsychological assessment: disadvantages (adapted from Parsons, 2016)

Disavantages	
Stimuli presentation	• Errors can occur in test administration due to problematic hardware and software interactions • Do not allow for "testing the limits," • Do not allow for flexibility in evaluations • Do not provide structured encouragement
Scoring	• Behavioral responses from computerized tests may not provide identical (or even similar) results as paper-and-pencil counterpart • May mask deficits that would otherwise be apparent in some populations (e.g., persons with autism may perform better when faced with a computer) • Computerized assessments may tap into cognitive functions at a level that is rarely demanded in real-life settings
Impact on participant	• Computerized tests may not be experientially or psychometrically equivalent to paper-and-pencil counterparts (validity) • Negative attitudes (including anxiety) about computers persists, especially among individuals with limited exposure to technology, computerized administration may alter task performance • Understanding and assessing levels of effort and motivation can prove challenging

In addition, a general limitation of most of these initial attempts of computer-based neuropsychological tests refers to the fact that they basically replicate and automate the traditional tests. And, as mentioned above, the traditional assessment tests focused on construct-driven and are not able to predict everyday functioning. For example, the continuous performance test (CPT) is a neuropsychological test, assessing selective and sustained attention mainly for attention deficit disorder diagnosis. CPT gathers two data: commission errors that are behavioural responses given when is not required and omission errors that are no behavioural responses given when are required. Commission errors are presumed to reproduce impulsivity in everyday life, and omission errors replicate inattentional behaviours. Studies of Epstein et al. (2006) showed that commission and omission errors didn't predict the established predictions. Specifically, omission errors were not related to inattention behaviours but to hyperactivity behaviours. Commission errors were related to impulsivity and to hyperactivity and inattention behaviours. In other words, the degrees of ecological validity of these first attempts were underestimated and were not able to capture the real-life behaviours. For solving this drawback, in the last 10 years, Serious games (SG) and virtual reality (VR), as well as augmented reality (AR) in the last 10 years have been proposed as an attempt to increase the ecological validity of cognitive functioning assessment (Cipresso et al., 2018; Parsons, 2011; Schultheis et al. 2002; Renison et al. 2012).

ATSNA: serious games

A serious game (SG) is a computer game that includes specific and/or "serious" objectives with respect to an entertainment game. SG are varied and various, ranging from realistic everyday situations to those that take place in imaginary worlds (Wattanasoontorn et al., 2013). A realistic or imaginary SG promote engagement through "fun" and interactive game design approaches, allowing maintaining user attention and motivation on situations and activities and studies on narratives SG showed positive effects on maintaining engaged participants to SG tasks (Dickey, 2011; Park et al., 2010).

Despite the nature of the SG, its creation involves the specification of the technological tools, and the contents that should be implement:

a) Technological tools: include the design software tools, the database, and the game machine. Through the design software tools, it is possible to create two- (2D) or three-dimensional (3D) environments. 2D objects have only two dimensions, such as length and width, but no depth; 3D objects include also depth. The database involves all the data and information required by the system, including user information, scores, times, etc. Finally, the game machine consists of the specific code on how the system runs.

b) Contents: can be defined as significant information that will be supplied to the users when the SG is tested. According to the objective of the SG, contents are translated to useful information, such as activities/tasks and narratives. The narrative is an important part in SG, giving coherence and context to the participant, leading him/her inside the story and activities (Jenkins, 2004; McDaniel et al., 2010). According to this, narratives can enhance the participant immersion, engagement, and motivation. Specifically, immersion refers to the ability of narrative to getting lost and disconnecting from the real world, generating flow – a state of total absorption and engagement in a situation (Csikszentmihalyi, 1997) - and presence – a feel of "being there" in the SG (Slater, 2009). Studies on immersion using narrative SG showed significant effect of narrative with respect to non-narrative conditions (Green & Brock, 2000; Green et al., 2004; McQuiggan et al., 2008a). For example, McQuiggan et al. (2008a) developed a SG, named Crystal Island, and tested three versions-conditions: narrative vs. minimal-narrative vs. power point conditions. The narrative version contained a virtual scenario with virtual agents and narratives; the minimal-narrative version contained only the basic narrative without virtual agents and the power point used a slideshow presentation of the scenario. The results showed higher significant differences on presence questionnaire on narrative condition than minimal-narrative condition. These results suggest that narratives can positively influence the sense of immersion of the participant experience.

At present, the approach based on SG is in an initial phase and the first applications were developed for psychological treatments and education/learning (Burke et al., 2009; Chatham, 2007; Shute & Ventura, 2013). In psychological field, SG have been mainly developed and tested on the treatment and prevention of depression (see the reviews and meta-analysis of Fleming et al. 2014, 2017; Li et al., 2014, 2016). More in detail, Table 5 reports the main game applications and results obtained from the three reviews and meta-analysis mentioned above.

Table 5

Systematic reviews of SG for psychology (adapted from Fleming et al., 2017).

Authors	*Article type*	*Types of game*	*Game applications studies*	*Results and conclusions*
Fleming et al., 2014	Systematic Review	Serious games to treat and prevent depression	9 studies on six applications: -Think Feel Do: 6 sessions of computerized cognitive behavior therapy (cCBT) for children and adolescents with depression or emotional distress, delivered on a PC -SPARX: 7 cCBT modules for adolescents with depression, delivered on a PC. Uses a virtual therapist and fantasy world with narratives. -Rainbow SPARX: modified version of SPARX for sexual minority adolescents with depression. -The Journey: 7 cCBT-PC modules for youth with depression, delivered on a PC and using a 2D fantasy world and games with narratives. -gNAT island: CBT on 2/4 sessions delivered on a PC and with a therapist. -Journey to the Wild Divine: a game based on biofeedback approach setting in a fantasy world.	Significant effectiveness results on depression. Further studies are needed.

			- ReachOutCentral: a CBT program delivered on a PC focused on problem solving and role-playing developed in a new town.	
Li et al., 2014	Systematic Review and meta-analysis	Exergames to treat and prevent depression	9 studies using Wii Sports, Wii Fit and Wii Fit Plus exergames.	Significant effectiveness results on depression. Further studies are needed.
Li et al., 2016	Systematic Review and meta-analysis	Games to treat depression		Significant effectiveness results on depression. Further studies are needed.

The main features provided by SG for psychology include: a) the possibility to reach a larger number of people that suffer of mental disorders thanks to the acceptance and “appealing” of computer games in daily life (Andrade et al., 2014; Marchand & Hennig-Thurau, 2013); b) game rules and challenges to overpass for reaching a goal with positive feedbacks or suggestions, providing “engaging” and, as a consequence, reducing dropout rates (Batterham et al., 2008; Fleming et al., 2016); c) “effectiveness” because it can generate a sense of flow and presence, as if the participant were in the real world, allowing to try, and learn new abilities, as well as to repeat new learned behavior in a safe and interactive environment (Cheek et al., 2015; Fleming et al., 2014).

ATSNA: Virtual reality and Augmented reality

Virtual (VR) and augmented reality (AR) can be identified inside a *continuum* in which reality and VR represent the two extemes and AR and augmented reality lies in the middle of them (Milgram & Kishino, 1994).

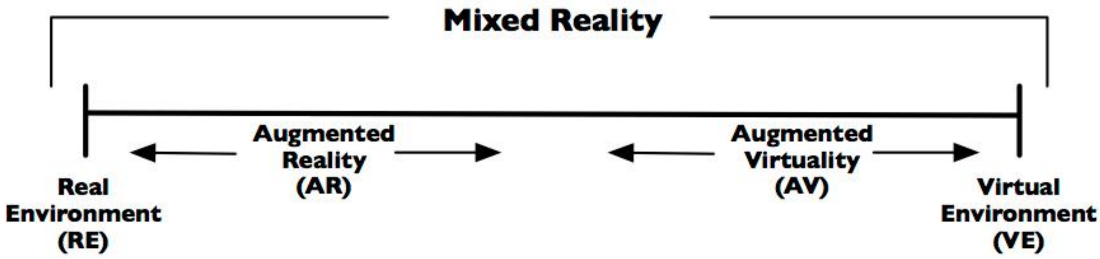


Figure 3. Virtual reality continuum (Source from Milgram et al., 1994 – http://etclab.mie.utoronto.ca/people/paul_dir/SPIE94/SPIE94.full.html)

VR, on one hand, refers to computer-simulated environments that can simulate physical and psychological sense of presence in places located in the real world (Slater et al., 2009). For a recent review on presence definitions, measures, and models see: Skarbez, R., Brooks Jr, F. P., & Whitton, M. C. (2018). A survey of presence and related concepts. *ACM Computing Surveys (CSUR)*, *50*(6), 96.

Factors as numbers of senses stimulated, and interactions, as well as the fidelity of the synthetic stimuli to reality, visual realism, and the ability of the system to isolate the user from external stimuli and reality contribute to VR level of immersion, influencing, in turn, the user experience. According to the level immersion, three VR systems can be identified: non-immersive, semi-immersive, and immersive. Non-immersive systems employ personal computer desktop with keyboard and mouse; semi-immersive display the virtual environments on large surface; and immersive systems use special equipment such as head mounted displays (HMD) to reproduce the visual scenes, controllers and haptic tactile devices for improving the senses stimulated and the interaction (Table 6). Furthermore, semi-immersive and immersive systems are more able to enhance the user perception of visual realism, fidelity of the stimuli and presence than non-immersive systems, depending on three factors: a) field of view (FOV); b) accuracy and complexity on graphics; and c) fidelity of interactions. Specifically, VR provides a FOV around 100 degrees and the possibility to recreate entire complex environments, gives to the user the perception of no-mediation by the tool (Slater, 2009). In other words, the tool disappeared from the user perception, allowing to him/her to experience a similar real-life situation.

On the other, augmented reality (AR) refers to the possibility to add synthetic information to the real world (Chicchi Giglioli et al., 2015; Ventura et al., 2018). Tecnologically, the FOV is between 35 and 45 degrees and the interaction depends on various headband camara sensors that through the human gaze to target, allow a real hands interaction. As can be seen, AR provides a narrower FOV than VR, a different interaction system, the no user isolation from physical world, and a madiated tool experience compared with VR. These four features present advantages and disavantages. First, a wider FOV allows the user to see the entire scenario at once and to use peripheral vision, while a narrower FOV can reduce distraction in the periphery and allow the user to focus on the interest area (McMahan et al., 2012; Ragan et al., 2010, 2012). Secondly, a more natural interaction, like in AR, supposes a higher experience fidelity (McMahan et al., 2012). Nevertheless, comparison studies on different hand controllers showed that the more familiar, like mouse, keyboards, and VR controllers, provided a best performance, although the participants appreciated a more natural interaction (McMahan et al., 2010). Third, the no user isolation from real-world in AR could allow a more real experience, allowing an immediate transferring of the behavioral performance in the real-life (Dunkin et al., 2007; Seymour, 2008; Saposnik et al., 2010). Forth, the AR experience is a mediated experience by the tool, while VR experience is an unitary unmediated experience by the tool.

Actually, the application development and the research studies on the neuropsychological assessment using VR has considerably increased in the last five years, showing an emerging interest in developing new cognitive assessment systems that overcome the limitations of traditional ones. Specifically, various applications, such as virtual shopping center, classroom, library, or office, have been developed and tested for the assessment of cognitive functioning on different clinical populations (see the reviews of Fleming et al., 2017; Kane & Parsons, 2019; Negut et al., 2014; Valladares-Rodríguez et al., 2016). For example, in the domain of executive functions, Raspelli et al., (2011) developed a virtual supermarket (VMET) displayed on a monitor, in which 9 post-stroke patients and 20 healthy subjects (10 younger adults and 10 older adults) should navigate for the supermarket, buying a list of products. The results showed significant correlations between the VMET and the Test of Attentional performance (Zimmerman & Fimm, 1992) and no correlations with the Stroop Test, for the assessment of inhibition control, and the Iowa Gambling Task on decision-making abilities. Cipresso et al. (2014) proved the VMET on Parkinson patients with mild cognitive impairments (n= 15), with normal cognitive functioning (n=15) and healthy subjects (n=20), and comparing the VMET performance with traditional neuropsychological tests. Results on criterion validity showed that the VMET is more able to detect differences in executive functions between groups than traditional assessment. Other researcher teams worked on a virtual classroom for the evaluation of the Attention Deficit Hyperactivity Disorder (ADHD) in children (Diaz-Orueta et al., 2012; Iriarte et al., 2016; Nolin et al., 2016; Pollak et al., 2009; Rizzo et al., 2000). To provide a realistic assessment environment, the virtual classroom included a typical rectangular classroom with desks, blackboard, a virtual teacher, and in which participants should perform various attentional and inhibition tasks based on the traditional Continuous Performance Task. The results of these studies showed, on one hand, positive usability data (Rizzo et al., 2000), and, on the other, good convergent and construct validity between the performance on the virtual classroom and the traditional tests (Diaz-Orueta et al., 2012; Iriarte et al., 2016; Nolin et al., 2016), as well as, sensibility and specificity to discriminate between children with and without ADHD (Pollak et al., 2009). The most of the VR studies presented in literature are in their first stage, including small groups of participants in which the main aim is to assess the feasibility of using new technologies for the assessment.

Regarding AR, several studies has been conducted in the treatment of certain disorders, such as phobias, allowing patients learning and repeating new behaviors to cope with fearful stimuli in safe and reactive environments generating effectiveness in behavioral changes in real contexts

(Chicchi et al, 2015; Ventura et al., 2018; Suso-Ribera et al., 2018) but, to our knowledge, no previous studies have assessed cognitive functions using an AR ecological task and comparing the behaviolral responses with VR.

Table 6

Features of the three VR systems.

Features	Non-Immersive	Semi-Immersive	Immersive
Immersion	×	×	✓
Real-time reactions	✓	✓	✓
Isolation from reality	×	×	✓
Sense stimulated (visuo, audio, tactile)	✓	✓	✓
Ecologically valid	✓	✓	✓

Stealth assessment and evidence-center design

A valid approach to develop contents is Stealth Assessment (SA) that refers to the use of Evidence-Centred Design (ECD), which consists of design valid assessment and provides real-time estimates of individuals' abilities across a range of skills (Mislevy, 1994; Mislevy et al., 2003; Shute & Ventura, 2013). More in detail, ECD defines a framework consisting of three conceptual and functional models that work all together. The first model, named competency model, allows identifying what set of constructs should be assessed. For example, in the assessment of executive function, this model allows researcher to expand extensively on the specific abilities that he/she want to assess, such as attention, planning, and cognitive flexibility. The second model, the evidence model, focuses on that behaviors and performance that should reveal those constructs identified in the competency model. The third model, the task model, allows creating situations, narratives, and tasks that should be developed to elicit those behaviors that should reflected the constructs that we want to assess (Figure 4). The main advantage on using this approach includes the possibility to dynamically assess a cluster of abilities at the same time (as mentioned above, traditional assessment allows evaluating one abilities at time) in engaging and situated environments. Furthermore, during the gameplay, individuals naturally produce various sequences of behaviors while performing tasks, that the system automatically gathers in a woven and indirect/implicit way, allowing to reduce the social desirability bias and subjective interpretations.

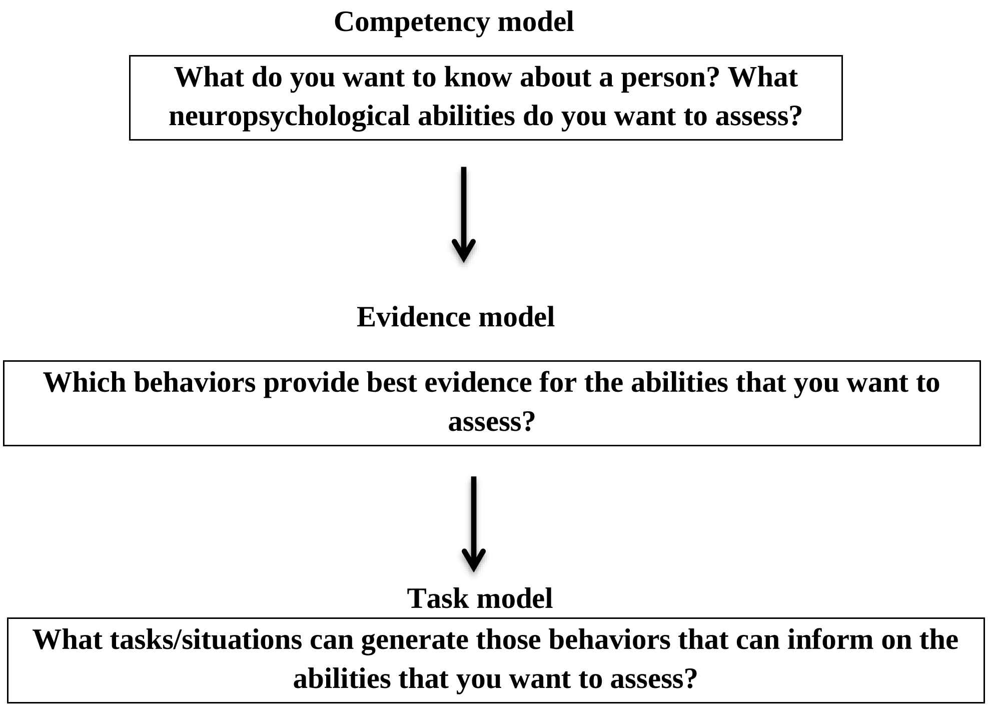


Figure 4. Evidence-centred design models (ECD)

ATSNA: advantages and limitations

Many advantages and some limitations have been found with regards to the use of ATSNA tools, along with providing more ecological environments than traditional approaches: a) ATSNA tests can support the standardization of stimuli presentation and the response gathering, providing more control during the assessment; b) the increasing use and commercialization of ATSNA tools allows to increase the size of the sample and as a consequence hinder studies on large-scales (Fleming et al., 2017); c) ATSNA tools allow to participants to continue exercising from home, following a virtual assistant that can guide their in the daily activities.

d) accuracy on latency and reaction time recording; e) higher accuracy in detecting cognitive abilities thanks to the accuracy in recording reaction times (milliseconds) and correct/incorrect answers; f) ATSNA tools minimize the examiner’s subjectivity interpretations and inferences, enabling automatic performance correction; and g) reduce user test anxiety, providing greater predictive validity compared to traditional tests. Regarding limitations: a) the use of ATSNA tools decreases face to face interaction between clinician and patients; b) the higher costs that ATSNA require than traditional assessment; c) the use of different input and output devices, such as desktop, HMD, or for the interaction, mouse, or keyboards or controllers, can change the individual performance (Germine et al., 2019); and d) the proliferation on the market of non-validated applications that jeopardize the work of researcher that clinically test their applications before making them public.

ATSNA challenges and thesis contribution

Despite the technological advances and the above mentioned advantages, the use of ATSNA for assessment is also related to several challenges. The main challenge encounters the arguments explicitly for the advantages of using games for assessment. There is a known need to check claims that such approaches to game development have more reliable evaluations. This goal is consistent with a wider need to demonstrate the value of their use as tests in the field of serious games for improved scientific rigor (Levy et al., 2016; Lumsden et al., 2016; Wiloth et al., 2016). A recent systematic review (Lumsden et al., 2016) demanded improved rigor in the analysis of serious game evaluation methods through research designs that increase causal inferences that can be made about the validity of serious games as assessment instruments and increased sample sizes to improve statistical power for increased confidence in determining that the tests were actually working. A few studies are already engaged in this research by exploring the similarities between user success in game evaluations and traditional validated evaluations (Levy et al., 2016; Wiloth et al., 2016).

Another challenge in using ATSNA tools for assessment can raise from the technological devices implemented. On one site, ATSNA hardware and software devices can vary in the accurary to record user response time latencies confounding the relationship between population (health versus clinical) and cognitive performance. On the other, difference in the sample on rating mouse/controller movements or touch, or differences in screen size or in visual display resolution have the pontential to influence test performance and the assessment of behavior. Therefore, a person might appear to be impaired when the cognitive differences depend on the devices implemented (Germine et al., 2019). To our knowledge, we were unable to find published studies showing comparison between the different specific ATSNA tools and their effectiveness in the assessment. A related challenge can be identified in the conversion of the traditional tests to ATSNA format and/or in the development of daily life activities or fantasy world contents to assess the basic contructs that can involve changes in the perceptual, cognitive, and motor complexity of a task. Indeed, for example the time requirements to complete the trail making task (Reitan, 1958) in the traditional paper and pencil format are longer than those in smaller screens (e.g. smartphone) and longer than larger screens (e.g. tablet; Begnum & Begnum, 2012). These changes in complexity of a task can be majors when the activity required to user is more ecologically valid and reflects a daily life situation. The format and contents changes have an impact on the score interpretations that is not well recognized and there is a need of more normative data across ATSNA tools to increase validity in cognitive assessment.

Generally, the advantages as well as the challenges presented above suggest that there is still much work to be done to develop games for assessment especially regarding scientifically demonstrating their validity. The contribution of the thesis fits into this context of reference with the aim to compare effectiveness using different ASTNA tools. In order to minimize the influence of ASTNA variability a particular attention has been dedicated to norms and designing ATSNA cognitive activities and tasks. Specifically, activities and tasks designed and developed across the implemented tools maintained and obtained similar scores. From a test design perspective, this was reach reducing dependence on stimuli or measurement features that can particularly be sensitive to tool related mistakes. In this way stimuli and situations have been accommodated to a variety of display tools and, where possible, the scores mainly based on accuracy and error of the responses rather than response time. In addition, user ATSNA experience and tool familiarity can interfere with valid assessment using technological tools, and therefore, user variability was taken into consideration in the designing and during the experimentation (for example, adding an initial tutorial before the tasks to familiarize with the tools).

Starting from these premises, using the evidence-centered design approach, we developed, on one side, a narrative SG, which the aims were to evaluate criterion and convergent validity of the score in the serious game and traditional neuropsychological evaluations. Furthermore, we evaluated the feasibility of using a 2D serious game and a 3D serious game for the assessment of cognitive functions. On the other, we developed an ecological dailylife task, such as cooking, for VR and AR systems and we investigated differences in the performance and the subjective sense of presence.

Objectives

The main objective of the thesis is to study the influence of different variables on ATSNA. In concrete the following variables have been studied:

- The use of a narrative, and interactive fantasy SG environments;
- The behavioural performance between a 2D interface and a 3D interface;
- Socio-demographic variables, such as gender, education, and age;
- The behavioural performance on an ecological cooking task between a VR and an AR interface.

The specific objectives were the following:

O1. To analyse the state of the art on the use of advanced technological systems. (Paper I)

O2. To design, using the stealth assessment approach, a narrative contextualized serious game based on cognitive abilities using ATSNA systems.

O3. To study the influence of narratives on the SG performance. (Paper II)

O4: To study the influence of 3D interfaces versus 2D interfaces, testing in two controlled studies the differential efficacy of the two systems by comparing the evaluation results with traditional neuropsychological tests (Paper III).

O5. To verify internal consistency of the scales and the criterion and convergent validity of the two systems. (Paper III).

O6. To study the influence of socio-demographic variables, such as gender, education, and age in SG performance and comparing it with the performance using traditional tests. (Paper IV).

O7. To study the performance differences of an ecological cooking task between a VR interface and an AR interface (Paper V).

Part II

Selected Papers

The Past, Present, and Future of Virtual and Augmented Reality Research: a network and cluster analysis of the literature

Cipresso, P., **Chicchi Giglioli, I. A.,** Raya, M. A., & Riva, G. (2018). The past, present, and future of virtual and augmented reality research: a network and cluster analysis of the literature. *Frontiers in psychology*, *9*.

Abstract

The recent appearance of low-cost Virtual Reality (VR) technologies – like the Oculus Rift, the HTC Vive and the Sony PlayStation VR – and Mixed Reality Interfaces (MRITF) – like the Hololens – is attracting the attention of users and researchers suggesting it may be the next largest stepping stone in technological innovation. However, the history of VR technology is longer than it may seem: the concept of VR was formulated in the 1960s and the first commercial VR tools appeared in the late 1980s.For this reason, during the last twenty years, hundreds of researchers explored the processes, effects and applications of this technology producing thousands of scientific papers. What is the outcome of this significant research work? This paper wants to provide an answer to this question by exploring, using advanced scientometric techniques, the existing research corpus in the field. We collected all the existent articles about VR in the Web of Science Core Collection scientific database, and the resultant dataset contained 21,667 records for VR and 9,944 for AR. The bibliographic record contained various fields, such as author, title, abstract, country, and all the references (needed for the citation analysis). The network and cluster analysis of the literature showed a composite panorama characterized by changes and evolutions over the time. Indeed, whether until five years ago, the main publication media on VR concerned both conference proceeding and journals, more recently journals constitute the main medium of

communication. Similarly, if at first computer science was the leading research field, nowadays clinical areas have increased, as well as the number of countries involved in virtual reality research. The present work discusses the evolution and changes over the time of the use of virtual reality in the main areas of application with an emphasis on the future expected virtual reality's capacities, increases and challenges. We conclude considering the disruptive contribution that VR/AR/MRITF will be able to get in scientific fields, as well in human communication and interaction, as already happened with the advent of mobile phones by increasing the use and the development of scientific applications (e.g. in clinical areas) and by modifying the social communication and interaction among people.

Introduction

In the last five years, virtual reality (VR) and augmented reality (AR) have attracted the interest of investors and the general public, especially after Mark Zuckerberg bought Oculus for two billion dollars (Castelvecchi, 2016; Luckerson, 2014). Currently, many other companies, such as Sony, Samsung, HTC, and Google are making huge investments in VR and AR (Castelvecchi, 2016; Ebert, 2015; Korolov, 2014). However, if VR has been used in research for more than 25 years, and now there are thousands of papers and many researchers in the field, comprising a strong, interdisciplinary community, AR has a more recent application history (Bohil et al., 2011; Burdea & Coiffet, 2003; Cipresso & Serino, 2014; Kim, 2005; Wexelblat, 2014). The study of VR was initiated in the computer graphics field and has been extended to several disciplines (Choi et al., 2015; Mazuryk & Gervautz, 1996; Sutherland, 1968; 1965). Currently, videogames supported by VR tools are more popular than the past, and they represent valuables, work-related tools for neuroscientists, psychologists, biologists, and other researchers as well. Indeed, for example, one of the main research purposes lies from navigation studies that include complex experiments that could be done in a laboratory by using VR, whereas, without VR, the researchers would have to go directly into the field, possibly with limited use of intervention. The importance of navigation studies for the functional understanding of human memory in dementia has been a topic of significant interest for a long time, and, in 2014, the Nobel Prize in "Physiology or Medicine" was awarded to John M. O'Keefe, May-Britt Moser, and Edvard I. Moser for their discoveries of nerve cells in the brain that enable a sense of place and navigation. Journals and magazines have extended this knowledge by writing about "the brain GPS," which gives a clear idea of the mechanism. A huge number of studies have been conducted in clinical settings by using Virtual Reality (Bohil et al., 2011; Serino et al., 2014), and Nobel Prize winner, Edvard I. Moser commented about the use of Virtual Reality (Minderer et al., 2016), highlighting its importance for research and clinical practice. Moreover, the availability of free tools for VR experimental and computational use has made it easy to access any field (Brown & Green, 2016; Cipresso, 2015; Cipresso et al., 2016; Riva et al., 2011).

AR is a more recent technology than VR and shows an interdisciplinary application framework, in which, nowadays, education and learning seem to be the most field of research. Indeed, AR allows supporting learning, for example increasing-on content understanding and memory preservation, as well as on learning motivation. However, if VR benefits from clear and more definite fields of application and research areas, AR is still emerging in the scientific scenarios.

In this article, we present a systematic and computational analysis of the emerging interdisciplinary VR and AR fields in terms of various co-citation networks in order to explore the evolution of the intellectual structure of this knowledge domain over time.

Virtual Reality Concepts and Features

The concept of VR could be traced at the mid of 1960 when Ivan Sutherland in a pivotal manuscript attempted to describe VR as a window through which a user perceives the virtual world as if looked, felt, sounded real and in which the user could act realistically (Sutherland, 1965).

Since that time and in accordance with the application area, several definitions have been formulated: for example, Fuchs & Bishop in 1992 defined VR as "real-time interactive graphics with 3D models, combined with a display technology that gives the user the immersion in the model world and direct manipulation" (Fuchs & Bishop, 1992); Gigante in 1993 described VR as

"The illusion of participation in a synthetic environment rather than external observation of such an environment. VR relies on a 3D, stereoscopic head-tracker displays, hand/body tracking and binaural sound. VR is an immersive, multi-sensory experience" (Gigante, 1993); and Cruz-Neira in 1993 "Virtual reality refers to immersive, interactive, multi-sensory, viewer-centered, 3D computer generated environments and the combination of technologies required building environments" (Cruz-Neira, 1993).

As we can notice, these definitions, although different, highlight three common features of VR systems: immersion, perception to be present in an environment, and interaction with that environment (Andersen & Thorpe, 2009; Bailenson et al., 2006; Biocca, 1997; Biocca et al., 2001; Heeter, 2000; Lombard & Ditton, 1997; Loomis et al., 1999; Skalski & Tamborini, 2007; Slater, 2009; Sundar et al., 2010). Specifically, immersion concerns the amount of senses stimulated, interactions, and the reality's similarity of the stimuli used to simulate environments. This feature can depend on the properties of the technological system used to isolate user from reality (Slater, 2009).

Higher or lower degrees of immersion can depend by three types of VR systems provided to the user:

- Non-immersive systems are the simplest and cheapest type of VR applications that use desktops to reproduce images of the world.
- Immersive systems provide a complete simulated experience due to the support of several sensory outputs devices such as Head Mounted Displays (HMD) for enhancing the stereoscopic view of the environment through the movement of the user's head, as well as audio and haptic devices.
- Semi-immersive systems such as Fish Tank VR are between the two above. They provide a stereo image of a three dimensional (3D) scene viewed on a monitor using a perspective projection coupled to the head position of the observer (Ware et al., 1993).Higher technological immersive systems have showed a closest experience to reality, giving to

the user the illusion of technological non-mediation and feeling him or her of “being in” or present in the virtual environment (Lombard & Ditton, 1997). Furthermore, higher immersive systems, than the other two systems, can give the possibility to add several sensory outputs allowing that the interaction and actions were perceived as real (Biocca et al., 2001; Heeter, 2000; Loomis et al., 1999).

Finally, the user’s VR experience could be disclosed by measuring presence, realism, and reality’s levels. Presence is a complex psychological feeling of “being there” in VR that involves the sensation and perception of physical presence , as well as the possibility to interact and react as if the user was in the real world (Heeter, 1992). Similarly, the realism’s level corresponds to the degree of expectation that the user has about of the stimuli and experience (Baños et al., 2000; Baños et al., 2005). If the presented stimuli are similar to reality, VR user’s expectation will be congruent with reality expectation, enhancing VR experience. In the same way, higher is the degree of reality in interaction with the virtual stimuli, higher would be the level of realism of the user’s behaviours (Baños et al., 2000; Baños et al., 2005).

From Virtual to Augmented Reality

Looking chronologically on VR and AR developments, we can trace the first 3D immersive simulator in 1962, when Morton Heilig created Sensorama, a simulated experience of a motorcycle running through Brooklyn characterized by several sensory impressions, such as audio, olfactory, and haptic stimuli, including also wind to provide a realist experience (Heilig, 1962). In the same years, Ivan Sutherland developed The Ultimate Display that, more than sound, smell, and haptic feedback, included interactive graphics that Sensorama didn’t provide. Furthermore, Philco developed the first Head-Mounted Display (HMD) that together with The Sword of Damocles of Sutherland was able to update the virtual images by tracking user’s head position and orientation (Sutherland, 1965). In the 70s, the University of North Carolina realized GROPE, the first system of force-feedback and Myron Krueger created VIDEOPLACE an Artificial Reality in which the users’ body figures were captured by cameras and projected on a screen (Krueger et al., 1985). In this way two or more users could interact in the 2D-virtual space.

In 1982, the US' Air Force created the first flight simulator (Visually Coupled Airbone System Simulator – VCASS) in which the pilot through an HMD could control the pathway and the targets. Generally, the 80's were the years in which the first commercial devices began to emerge: for example, in 1985 the VPL company commercialized the DataGlove, glove sensors' equipped able to measure the flexion of fingers, orientation and position, and identify hand gestures. Another example is the Eyephone, created in 1988 by the VPL Company, an HMD system for completely immerging the user in a virtual world. At the end of 80's, Fake Space Labs created a Binocular-Omni-Orientational Monitor (BOOM), a complex system composed by a stereoscopic-displaying device, providing a moving and broad virtual environment, and a mechanical arm tracking. Furthermore, BOOM offered a more stable image and giving more quickly responses to movements than the HMD devices. Thanks to BOOM and DataGlove, the NASA Ames Research Center developed the Virtual Wind Tunnel in order to research and manipulate airflow in a virtual airplane or space ship. In 1992, the Electronic Visualization Laboratory of the University of Illinois created the CAVE Automatic Virtual Environment, an immersive virtual reality system composed by projectors directed on three or more walls of a room.

More recently, many videogames companies have improved the development and quality of VR devices, like Oculus Rift, or HTC Vive that provide a wider field of view and lower latency. In addition, the actual HMD's devices can be now combined with other tracker system as eye-tracking systems (FOVE), and motion and orientation sensors (e.g. Razer Hydra, Oculus Touch, or HTC Vive).

Simultaneously, at the beginning of 90', the Boing Corporation created the first prototype of AR system for showing to employees how set up a wiring tool (Carmigniani et al., 2011). At the same time, Rosenberg and Feiner developed an AR fixture for maintenance assistance, showing that the operator performance enhanced by added virtual information on the fixture to repair (Rosenberg, 1993). In 1993 Loomis and colleagues produced an AR GPS-based system for helping the blind in the assisted navigation through adding spatial audio information (Loomis et al., 1998). Always in the 1993 Julie Martin developed "Dancing in Cyberspace", an AR theater in which actors interacted with virtual object in real time (Cathy, 2011). Few years later, Feiner et al. (1997) developed the first Mobile AR System (MARS) able to add virtual information about touristic buildings (Feiner et al., 1997). Since then, several applications have been developed: in 2000, Thomas et al., created ARQuake, a mobile AR video game; in 2008 was created Wikitude that through the mobile camera, internet, and GPS could add information about the user's environments (Perry, 2008). In 2009 others AR applications, like AR Toolkit and SiteLens have been developed in order to add virtual information to the physical user's surroundings. In 2011, Total Immersion developed D'Fusion, and AR system for designing projects (Maurugeon, 2011). Finally, in 2013 and 2015, Google developed Google Glass and Google HoloLens, and their usability have begun to test in several field of application.

Virtual Reality Technologies

Technologically, the devices used in the virtual environments play an important role in the creation of successful virtual experiences. According to the literature, can be distinguished input and output devices (Burdea et al., 1996; 2003). Input devices are the ones that allow the user to communicate with the virtual environment, which can range from a simple joystick or keyboard to a glove allowing capturing finger movements or a tracker able to capture postures. More in detail, keyboard, mouse, trackball, and joystick represent the desktop input devices easy to use, which allow the user to launch continuous and discrete commands or movements to the environment. Other input devices can be represented by tracking devices as bend-sensing gloves that capture hand movements, postures and gestures, or pinch gloves that detect the fingers movements, and trackers able to follow the user's movements in the physical world and translate them in the virtual environment.

On the contrary, the output devices allow the user to see, hear, smell, or touch everything that happens in the virtual environment. As mentioned above, among the visual devices can be found a wide range of possibilities, from the simplest or least immersive (monitor of a computer) to the most immersive one such as virtual reality glasses or helmets or HMD or CAVE systems.

Furthermore, auditory, speakers, as well as haptic output devices are able to stimulate body senses providing a more real virtual experience. For example, haptic devices can stimulate the touch feeling and force models in the user.

Virtual Reality Applications

Since its appearance, VR has been used in different fields, as for gaming (Meldrum et al., 2012; Zyda, 2005), military training (Alexander et al., 2017), architectural design (Song et al., 2017), education (Englund et al., 2017), learning and social skills training (Schmidt et al., 2017), simulations of surgical procedures (Gallagher et al., 2005), assistance to the elderly or psychological treatments are other fields in which virtual reality is bursting strongly (Freeman et al., 2017; Neri et al., 2017). A recent and extensive review of Slater and Sanchez-Vives (Slater and Sanchez-Vives, 2016) reported the main VR application evidences, including weakness and advantages, in several research areas, such as science, education, training, physical training, as well as social phenomena, moral behaviors, and could be used in other fields, like travel, meetings, collaboration, industry, news, and entertainment. Furthermore, another review published this year by Freeman et al. (2017) focused on VR in mental health, showing the efficacy of VR in assessing and treating different psychological disorders as anxiety, schizophrenia, depression, and eating disorders.

There are many possibilities that allow the use of virtual reality as a stimulus, replacing real stimuli, recreating experiences, which in the real world would be impossible, with a high realism. This is why VR is widely used in research on new ways of applying psychological treatment or training, for example, to problems arising from phobias (agoraphobia, phobia to fly, etc.) (Botella et al., 2017). Or, simply, it is used like improvement of the traditional systems of motor rehabilitation (Borrego et al., 2016; Llorens et al., 2014), developing games that ameliorate the tasks. More in detail, in psychological treatment, Virtual Reality Exposure Therapy (VRET) has showed its efficacy, allowing to patients to gradually face fear stimuli or stressed situations in a safe environment where the psychological and physiological reactions can be controlled by the therapist (Botella et al., 2017).

Augmented Reality Concept

In 1994, Milgram and Kishino, conceptualized the Virtual-Reality Continuum that takes into consideration four systems: real environment, augmented reality (AR), augmented virtuality, and virtual environment (Milgram & Kishino, 1994). AR can be defined a newer technological system in which virtual objects are added to the real world in real-time during the user's experience. Per Azuma et al. (2001) an AR system should: (1) combine real and virtual objects in a real environment; (2) run interactively and in real-time; (3) register real and virtual objects with each other. Furthermore, even if the AR experiences could seem different from VRs, the quality of AR experience could be considered similarly. Indeed, like in VR, feeling of presence, level of realism, and the degree of reality represent the main features that can be considered the indicators of the quality of AR experiences. Higher the experience is perceived as realistic, and there is congruence between the user's expectation and the interaction inside the AR environments, higher would be the perception of "being there" physically, and at cognitive and emotional level. The feeling of presence, both in AR and VR environments, is important in acting behaviors like the real ones(Botella et al., 2005; Breton-Lopez et al., 2010; Juan et al., 2005; Wrzesien et al., 2013).

Augmented Reality Technologies

Technologically, the AR systems, however various, present three common components, such as a geospatial datum for the virtual object, like a visual marker, a surface to project virtual elements to the user, and an adequate processing power for graphics, animation, and merging of images, like a pc and a monitor (Carmigniani et al., 2011). To run, an AR system must also include a camera able to track the user movement for merging the virtual objects, and a visual display, like glasses through that the user can see the virtual objects overlaying to the physical world. To date, two-display systems exist, a video see-through (VST) and an optical see-though (OST) AR systems (Botella et al., 2005; Juan et al., 2005; Juan et al., 2007). The first one, disclosures virtual objects to the user by capturing the real objects/scenes with a camera and overlaying virtual objects, projecting them on a video or a monitor, while the second one, merges the virtual object on a transparent surface, like glasses, through the user see the added elements. The main difference between the two systems is the latency: an OST system could require more time to display the virtual objects than a VST system, generating a time lag between user's action and performance and the detection of them by the system.

Augmented Reality Applications

Although AR is a more recent technology than VR, it has been investigated and used in several research areas such as architecture (Lin & Hsu, 2017), maintenance (Schwald & De Laval, 2003), entertainment (Ozbek et al., 2004), education (Nincarean et al., 2013; Bacca et al., 2014; Akçayır & Akçayır, 2017), medicine (De Buck e tal., 2005), and psychological treatments (Botella et al., 2005; Botella et al., 2010;Breton-Lopez et al., 2010; Juan et al., 2005; Wrzesien et al., 2011a; 2011b; Wrzesien et al., 2013; see the review Chicchi Giglioli et al., 2015)). More in detail, in education several AR applications have been developed in the last few years showing the positive effects of this technology in supporting learning, such as an increased-on content understanding and memory preservation, as well as on learning motivation (Radu, 2012; 2014). For example, Ibanez et al., (2014) developed a AR application on electromagnetism concepts' learning, in which students could use AR batteries, magnets, cables on real superficies, and the system gave a real-time feedback to students about the correctness of the performance, improving in this way the academic success and motivation (Di Serio, 2012). Deeply, AR system allows the possibility to learn visualizing and acting on composite phenomena that traditionally students study theoretically, without the possibility to see and test in real world (Chien et al., 2010; Chen et al., 2011).

As well in psychological health, the number of research about AR is increasing, showing its efficacy above all in the treatment of psychological disorder (see the reviews Baus and Bouchard, 2015; Chicchi Giglioli et al., 2015). For example, in the treatment of anxiety disorders, like phobias, AR exposure therapy (ARET) showed its efficacy in one-session treatment, maintaining the positive impact in a follow-up at one or three month after. As VRET, ARET provides a safety and an ecological environment where any kind of stimulus is possible, allowing to keep control over the situation experienced by the patients, gradually generating situations of fear or stress.

Indeed, in situations of fear, like the phobias for small animals, AR applications allow, in accordance with the patient's anxiety, to gradually expose patient to fear animals, adding new animals during the session or enlarging their or increasing the speed. The various studies showed that AR is able, at the beginning of the session, to activate patient's anxiety, for reducing after one hour of exposition. After the session, patients even more than to better manage animal's fear and anxiety, ware able to approach, interact, and kill real feared animals.

Methods

Data collection

The input data for the analyses were retrieved from the scientific database Web of Science Core Collection and the search terms used were "Virtual Reality" and "Augmented Reality", regarding papers published during the whole timespan covered.

Web of Science Core Collection is composed of: Citation Indexes, Science Citation Index Expanded (SCI-EXPANDED) --1970-present, Social Sciences Citation Index (SSCI) --1970-present, Arts & Humanities Citation Index (A&HCI) --1975-present, Conference Proceedings Citation Index- Science (CPCI-S) --1990-present, Conference Proceedings Citation Index- Social Science & Humanities (CPCI-SSH) --1990-present, Book Citation Index– Science (BKCI-S) --2009-present, Book Citation Index– Social Sciences & Humanities (BKCI-SSH) --2009-present, Emerging Sources Citation Index (ESCI) --2015-present, Chemical Indexes, Current Chemical Reactions (CCR-EXPANDED) --2009-present (Includes Institut National de la Propriete Industrielle structure data back to 1840), Index Chemicus (IC) --2009-present.

The resultant dataset contained a total of 21,667 records for VR and 9,944 records for AR. The bibliographic record contained various fields, such as author, title, abstract, and all of the references (needed for the citation analysis). The research tool to visualize the networks was Cite space v.4.0.R5 SE (32 bit) (Chen, 2006) under Java Runtime v.8 update 91 (build 1.8.0_91-b15). Statistical analyses were conducted using Stata MP-Parallel Edition, Release 14.0, StataCorp LP.

The betweenness centrality of a node in a network measures the extent to which the node is part of paths that connect an arbitrary pair of nodes in the network (Brandes, 2001; Chen, 2006; Freeman, 1977).

Structural metrics include betweenness centrality, modularity, and silhouette. Temporal and hybrid metrics include citation burstness and novelty. All the algorithms are detailed (Chen, C., Ibekwe-SanJuan, F., & Hou, J., 2010).

Results

The analysis of the literature on VR shows a complex panorama. At first sight, according to the document-type statistics from the Web of Science (WoS), proceedings papers were used extensively as outcomes of research, comprising almost 48% of the total (10,392 proceedings), with a similar number of articles on the subject amounting to about 47% of the total of 10, 199 articles. However, if we consider only the last five years (7,755 articles representing about 36% of the total), the situation changes with about 57% for articles (4,445) and about 33% for proceedings (2,578). Thus, it is clear that VR field has changed in areas other than at the technological level.

About the subject category, nodes and edges are computed as co-occuring subject categories from the Web of Science "Category" field in all the articles.

According to the subject category statistics from the WoS, computer science is the leading category, followed by engineering, and, together, they account for 15,341 articles, which make up about 71% of the total production). However, if we consider just the last five years, these categories reach only about 55%, with a total of 4,284 articles (Table 7 and Figure 5).

The evidence is very interesting since it highlights that VR is doing very well as new technology with huge interest in hardware and software components. However, with respect to the past, we are witnessing increasing numbers of applications, especially in the medical area. In particular, note its inclusion in the top ten lists of rehabilitation and clinical neurology categories (about 10% of the total production in the last five years). It also is interesting that neuroscience and neurology, considered together, have shown an increase from about 12% to about 18.6% over the last five years. However, historic areas, such as automation and control systems, imaging science and photographic technology, and robotics, which had accounted for about 14.5% of the total articles ever produced were not even in the top ten for the last five years, with each one accounting for less than 4%.

Table 7

Category statistics from the WoS for the entire period and the last five years

%	Frequency	Subject category (for all the period)
42,15%	9131	Computer Science, 1990 - 2016
28,66%	6210	Engineering, 1990 - 2016
8,21%	1779	Psychology, 1990 - 2016
7,15%	1548	Neurosciences & Neurology, 1992 - 2016
6,55%	1418	Surgery, 1992 - 2016
5,85%	1267	Automation & Control Systems, 1993 - 2016
4,80%	1040	Neurosciences, 1992 - 2016
4,74%	1027	Imaging Science & Photographic Technology, 1992 - 2016
4,30%	931	Education & Educational Research, 1993 - 2016
3,92%	849	Robotics, 1992 - 2016

%	Frequency	Subject category (for the last five years)
29,80%	2311	Computer Science, 2011 - 2016
25,44%	1973	Engineering, 2011 - 2016
11,10%	861	Neurosciences & Neurology, 2011 - 2016
9,32%	723	Psychology, 2011 - 2016
7,70%	597	Surgery, 2011 - 2016
7,53%	584	Neurosciences, 2011 - 2016
6,02%	467	Education & Educational Research, 2011 - 2016
5,54%	430	Rehabilitation, 2011 - 2016
4,42%	343	Clinical Neurology, 2011 - 2016
3,92%	304	Materials Science, 2011 - 2016

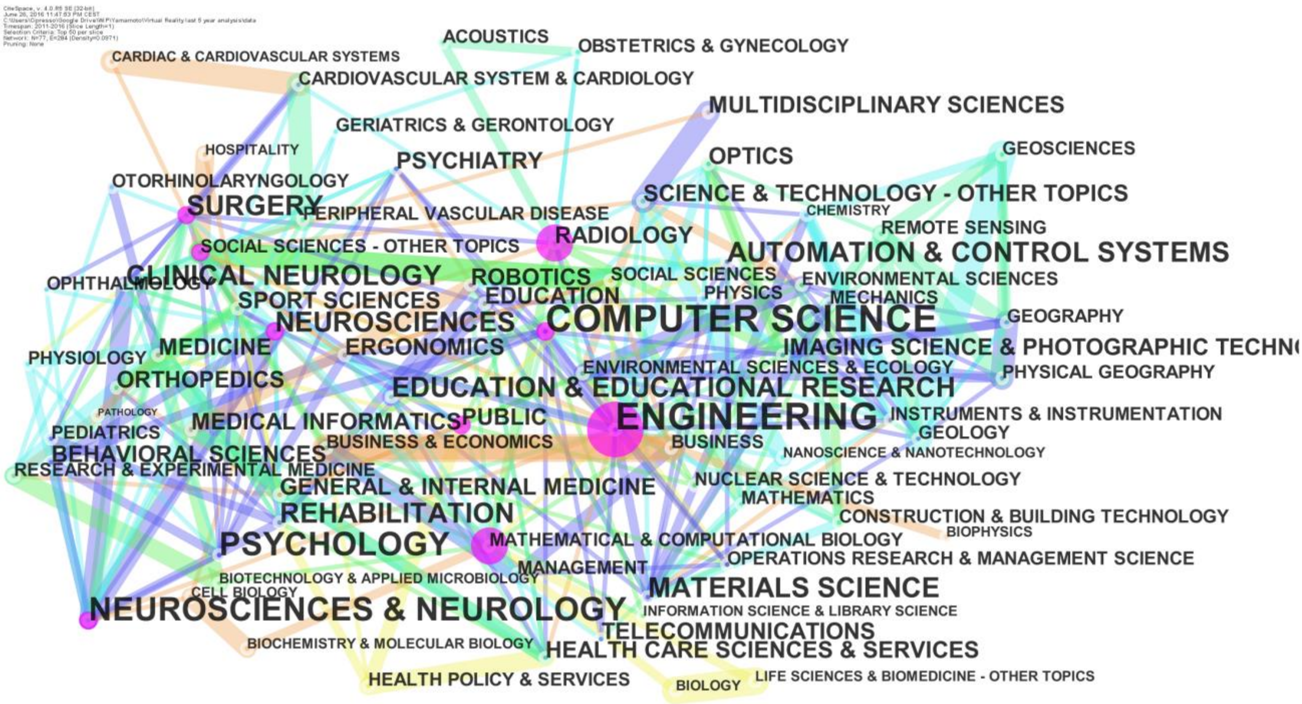


Figure 5. Category from the WoS: network for the last five years.

About the countries, nodes and edges are computed as networks of co-authors countries. Multiple occurrence of a country in the same paper is counted once.

The countries that were very involved in VR research have published for about 47% of the total (10,200 articles altogether). Of the 10,200 articles, the USA, China, England, and Germany published 4921, 2384, 1497, and 1398, respectively. The situation remains the same if we look at the articles published over the last five years. However, VR contributions also came from all over the globe, with Japan, Canada, Italy, France, Spain, South Korea, and The Netherlands taking positions of prominence, as shown in Figure 5.

Network analysis was conducted to calculate and to represent the centrality index (Brandes, 2001; Freeman, 1977), i.e., the dimension of the node in Figure 6. The top-ranked country, with a centrality index of 0.26, was the USA (2011), and England was second, with a centrality index of 0.25. The third, fourth, and fifth countries were Germany, Italy, and Australia, with centrality indices of 0.15, 0.15, and 0.14, respectively.

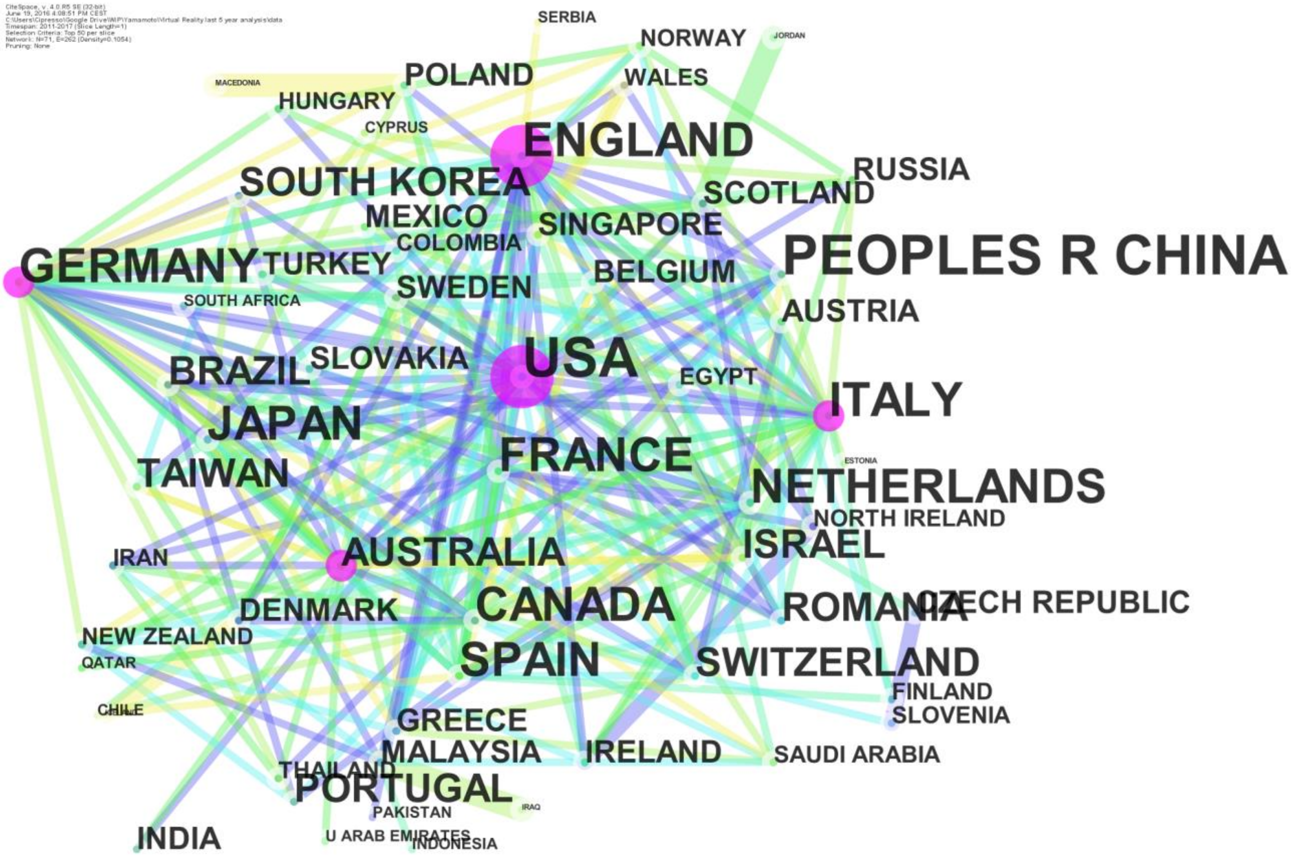


Figure 6. Country network (node dimension represents centrality).

About the Institutions, nodes and edges are computed as networks of co-authors Institutions (Figure 7).

The top-level institutions in VR were in the USA, where three universities were ranked as the top three in the world for published articles; these universities were the University of Illinois (159), the University of South California (147), and the University of Washington (146). The USA also had the eighth-ranked university, which was Iowa State University (116). The second country in the ranking was Canada, with the University of Toronto, which was ranked fifth with 125 articles and McGill University, ranked 10th with 103 articles.

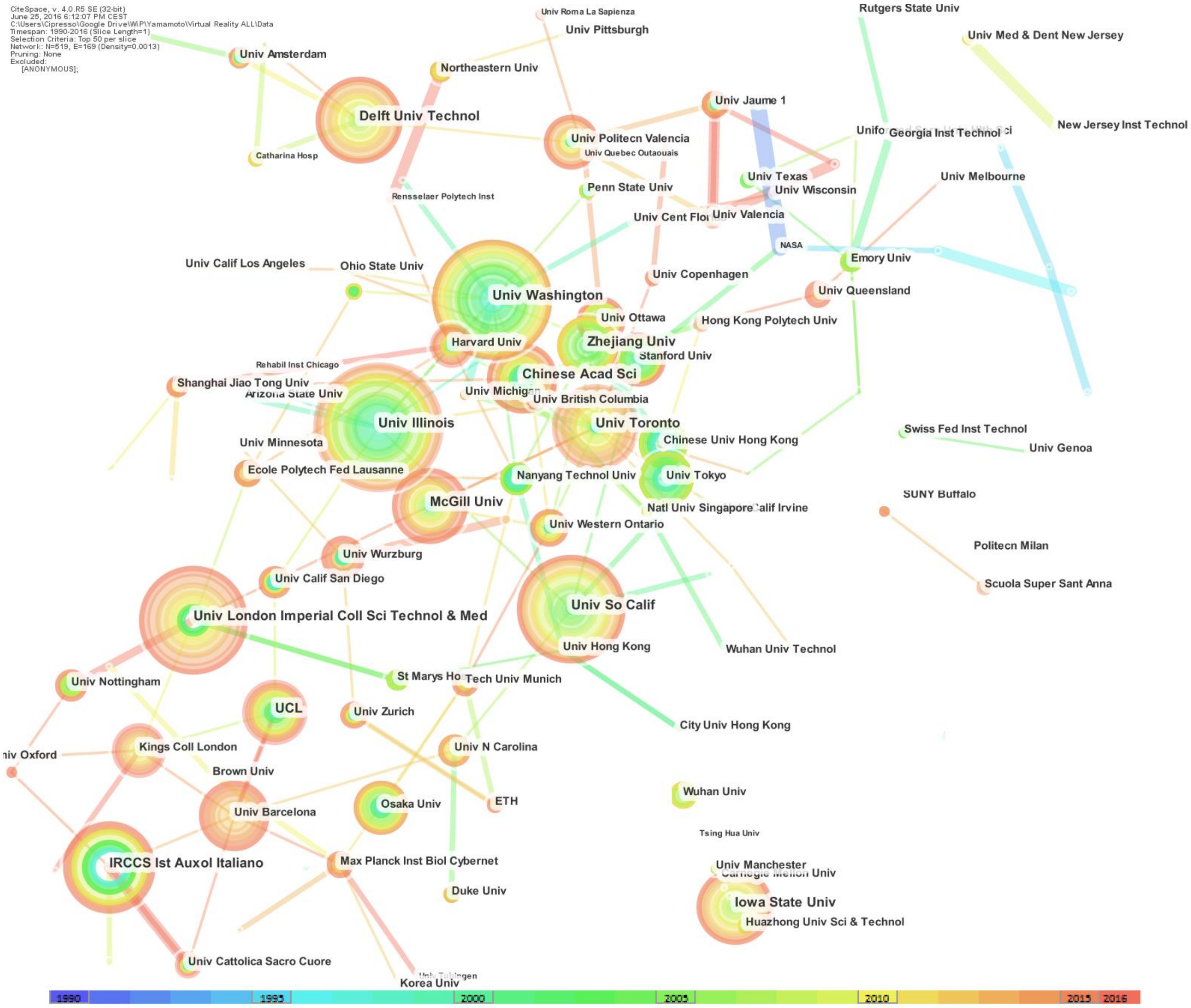


Figure 7. Network of institutions: the dimensions of the nodes represent centrality.

Other countries in the top-ten list were The Netherlands, with the Delft University of Technology ranked fourth with 129 articles; Italy, with IRCCS Istituto Auxologico Italiano, ranked sixth (with the same number of publication of the institution ranked fifth) with 125 published articles; England, which was ranked seventh with 125 articles from the University of London’s Imperial College of Science, Technology, and Medicine; and China with 104 publications, with the Chinese Academy of Science, ranked ninth. Italy’s Istituto Auxologico Italiano, which was ranked fifth, was the only non-university institution ranked in the top-ten list for VR research (Figure 7).

About the Journals, nodes and edges are computed as journal co-citation networks among each journal in the corresponding field.

The top-ranked Journals for citations in VR are *Presence: Teleoperators & Virtual Environments* with 2689 citations and *CyberPsychology & Behavior* (Cyberpsychol BEHAV) with 1884 citations; however, looking at the last five years, the former had increased the citations, but the latter had a far more significant increase, from about 70% to about 90%, i.e., an increase from 1029 to 1147.

Following the top two journals, IEEE Computer Graphics and Applications (*IEEE Comput Graph)* and Advanced Health Telematics and Telemedicine (*St HEAL T)* were both left out of the top-ten list based on the last five years. The data for the last five years also resulted in the inclusion of Experimental Brain Research (*Exp BRAIN RES)* (625 citations), Archives of Physical Medicine and Rehabilitation (*Arch PHYS MED REHAB)* (622 citations), and *Plos ONE* (619 citations) in the top-ten list of three journals, which highlighted the categories of rehabilitation and clinical neurology and neuroscience and neurology. Journal co-citation analysis is reported in Figure 8, which clearly shows four distinct clusters.

Network analysis was conducted to calculate and to represent the centrality index, i.e., the dimensions of the nodes in Figure 8. The top-ranked item by centrality was Cyberpsychol BEHAV, with a centrality index of 0.29. The second-ranked item was Arch PHYS MED REHAB, with a centrality index of 0.23. The third was Behaviour Research and Therapy (Behav RES THER), with a centrality index of 0.15. The fourth was BRAIN, with a centrality index of 0.14. The fifth was Exp BRAIN RES, with a centrality index of 0.11.

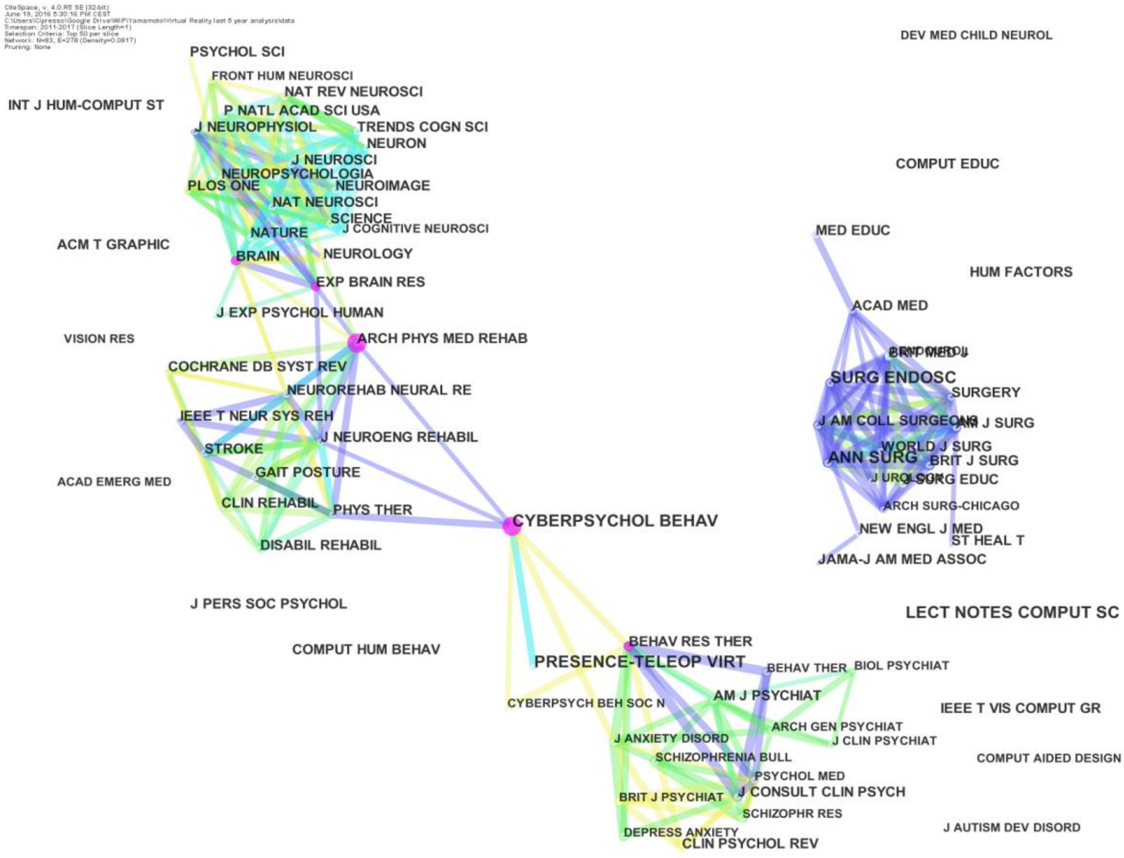


Figure 8. Co-citation network of journals: the dimensions of the nodes represent centrality. Full list of official abbreviations of WoS journals can be found here: https://images.webofknowledge.com/WOK46/help/WOS/A_abrvjt.html.

Who's Who in VR Research

Authors are the heart and brain of research, and their roles in a field are to define the past, present, and future of disciplines and to make significant breakthroughs to make new ideas arise (Figure 9).

VR research is very young and changing with time, but the top-ten authors in this field have made fundamentally significant contributions as pioneers in VR and taking it beyond a mere technological development. The purpose of the following highlights is not to rank researchers;

rather, the purpose is to identify the most active researchers in order to understand where the field is going and how they plan for it to get there.

The top-ranked author is Riva G, with 180 publications. The second-ranked author is Rizzo A, with 101 publications. The third is Darzi A, with 97 publications. The forth is Aggarwal R, with 94 publications. The six authors following these three are Slater M, Alcaniz M, Botella C, Wiederhold BK, Kim SI, and Gutierrez-Maldonado J with 90, 90, 85, 75, 59, and 54 publications, respectively (Figure 10).

Considering the last five years, the situation remains similar, with three new entries in the top-ten list, i.e., Muhlberger A, Cipresso P, and Ahmed K ranked seventh, eighth, and tenth, respectively.

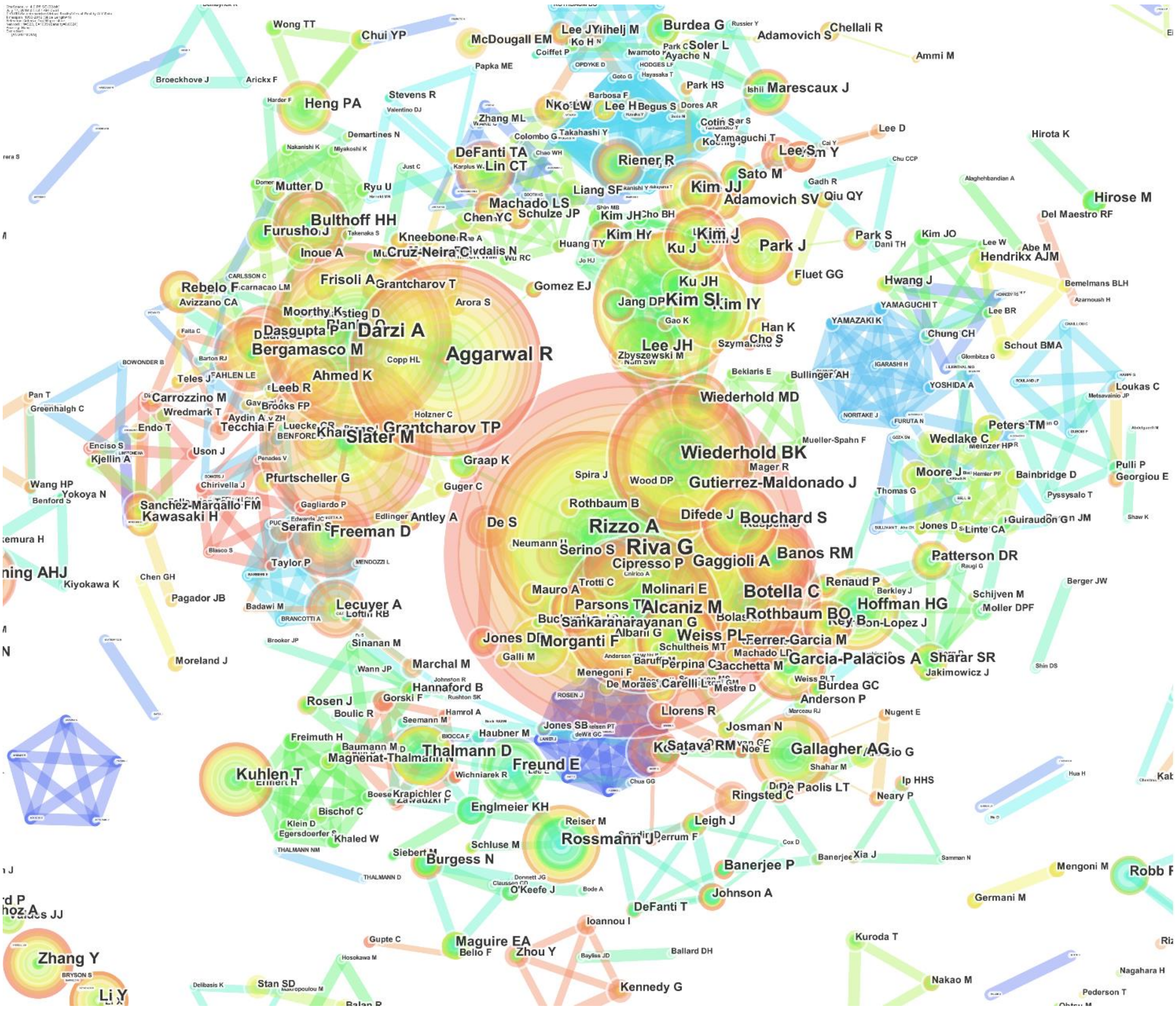


Figure 9. Network of authors' numbers of publications: the dimensions of the nodes represent the centrality index, and the dimensions of the characters represent the author's rank.

The authors’ publications number network shows the most active authors in VR research. Another relevant analysis for our focus on VR research is to identify the most cited authors in the field.

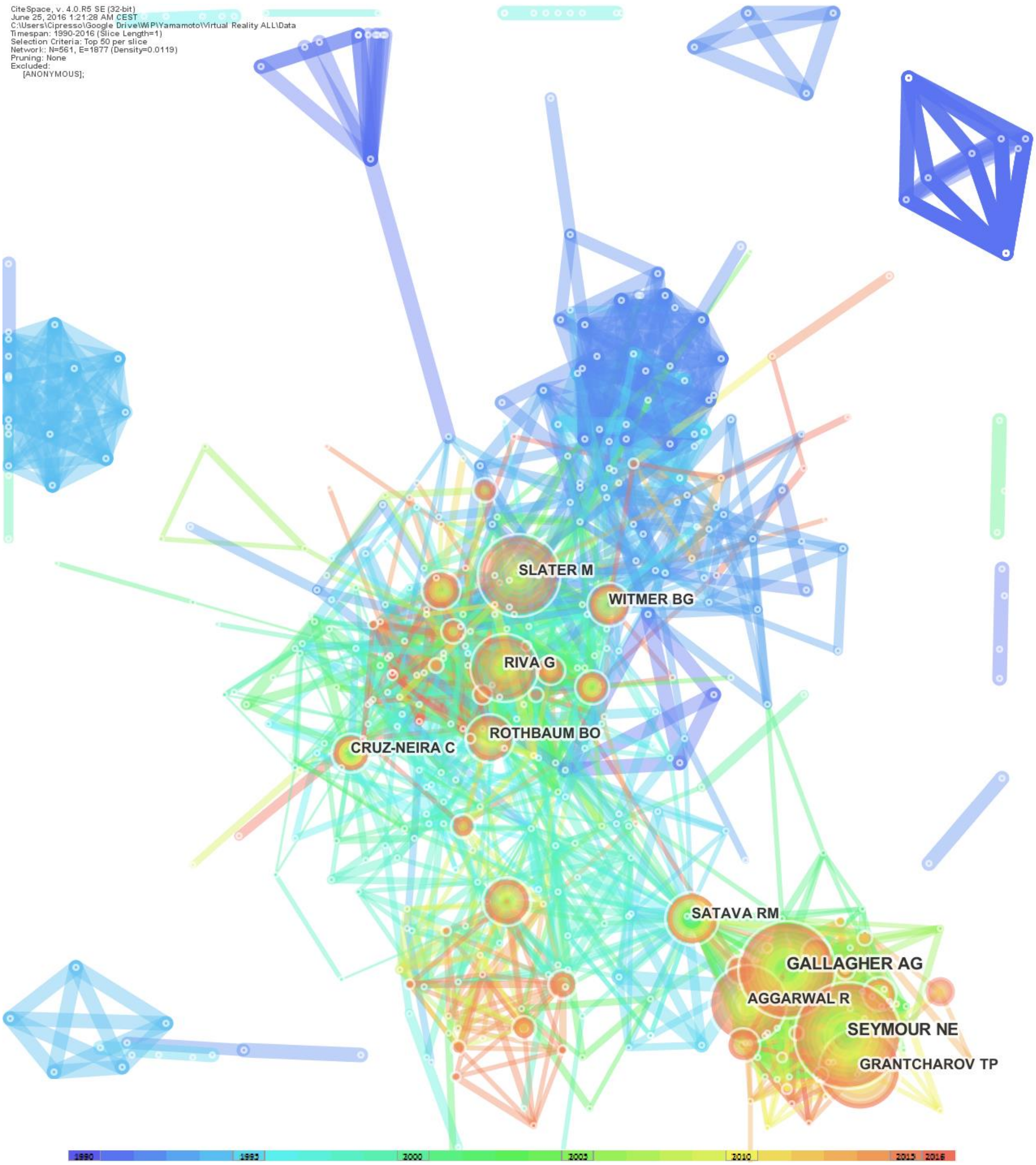


Figure 10. Authors' co-citation network: The dimensions of the nodes represent centrality index, and the dimensions of the characters represent the author's rank. The 10 authors that appear on the top-ten list are considered to be the pioneers of VR research.

For this purpose, the authors' co-citation analysis highlights the authors in term of their impact on the literature considering the entire time span of the field (Bu et al., 2016; González-Teruel et al., 2015; White & Griffith, 1981). The idea is to focus on the dynamic nature of the community of authors who contribute to the research.

Normally, authors with higher numbers of citations tend to be the scholars who drive the fundamental research and who make the most meaningful impacts on the evolution and development of the field. In the following, we identified the most-cited pioneers in the field of VR Research.

The top-ranked author by citation count is Gallagher AG (2001), with 694 citations. Second is Seymour NE (2004), with 668 citations. Third is Slater M (1999), with 649 citations. Fourth is Grantcharov TP (2003), with 563 citations. Fifth is Riva G (1999), with 546 citations. Sixth is Aggarwal R (2006), with 505 citations. Seventh is Satava RM (1994), with 477 citations. Eighth is Witmer BG (2002), with 454 citations. Ninth is Rothbaum BO (1996), with 448 citations. Tenth is Cruz-neira C (1995), with 416 citations.

Citation network and cluster analyses for VR

Another analysis that can be used is the analysis of document co-citation, which allows us to focus on the highly-cited documents that generally are also the most influential in the domain (González-Teruel et al., 2015; Orosz et al., 2016; Small, 1973).

The top-ranked article by citation counts is Seymour NE (2002) in Cluster #0, with 317 citations. The second article is Grantcharov TP (2004) in Cluster #0, with 286 citations. The third is Holden MK (2005) in Cluster #2, with 179 citations. The 4th is Gallagher AG (2005) in Cluster #0, with 171 citations. The 5th is Ahlberg G (2007) in Cluster #0, with 142 citations. The 6th is Parsons TD (2008) in Cluster #4, with 136 citations. The 7th is Powers MB (2008) in Cluster #4, with 134 citations. The 8th is Aggarwal R (2007) in Cluster #0, with 121 citations. The 9th is Reznick RK

(2006) in Cluster #0, with 121 citations. The 10th is Munz Y (2004) in Cluster #0, with 117 citations.

The network of document co-citations is visually complex (Figure 11) because it includes thousands of articles and the links among them. However, this analysis is very important because can be used to identify the possible conglomerate of knowledge in the area, and this is essential for a deep understanding of the area. Thus, for this purpose, a cluster analysis was conducted (Chen et al., 2010; González-Teruel et al., 2015; Klavans &Boyack, 2015). Figure 12 shows the clusters, which are identified with the two algorithms in Table 8.

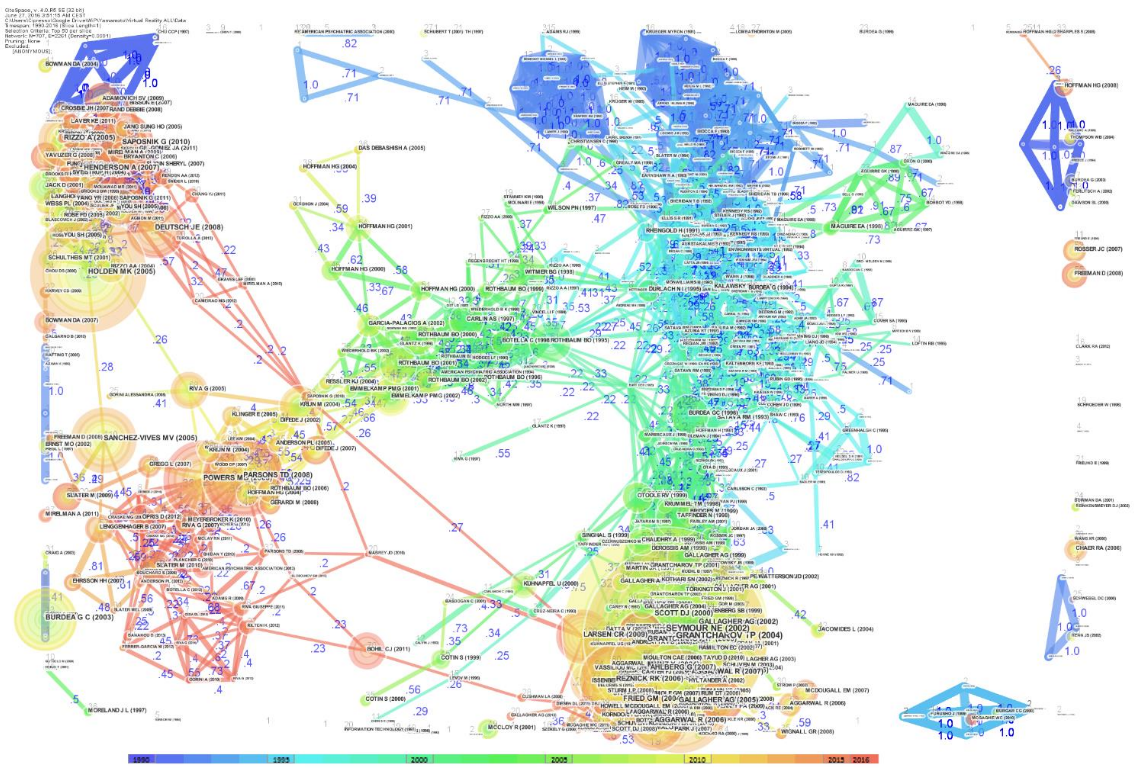


Figure 11. Network of document co-citations: The dimensions of the nodes represent centrality, the dimensions of the characters represent the rank of the article rank, and the numbers represent the strengths of the links. It is possible to identify four historical phases (colors: blue, green, yellow and red) from the past VR research to the current research.

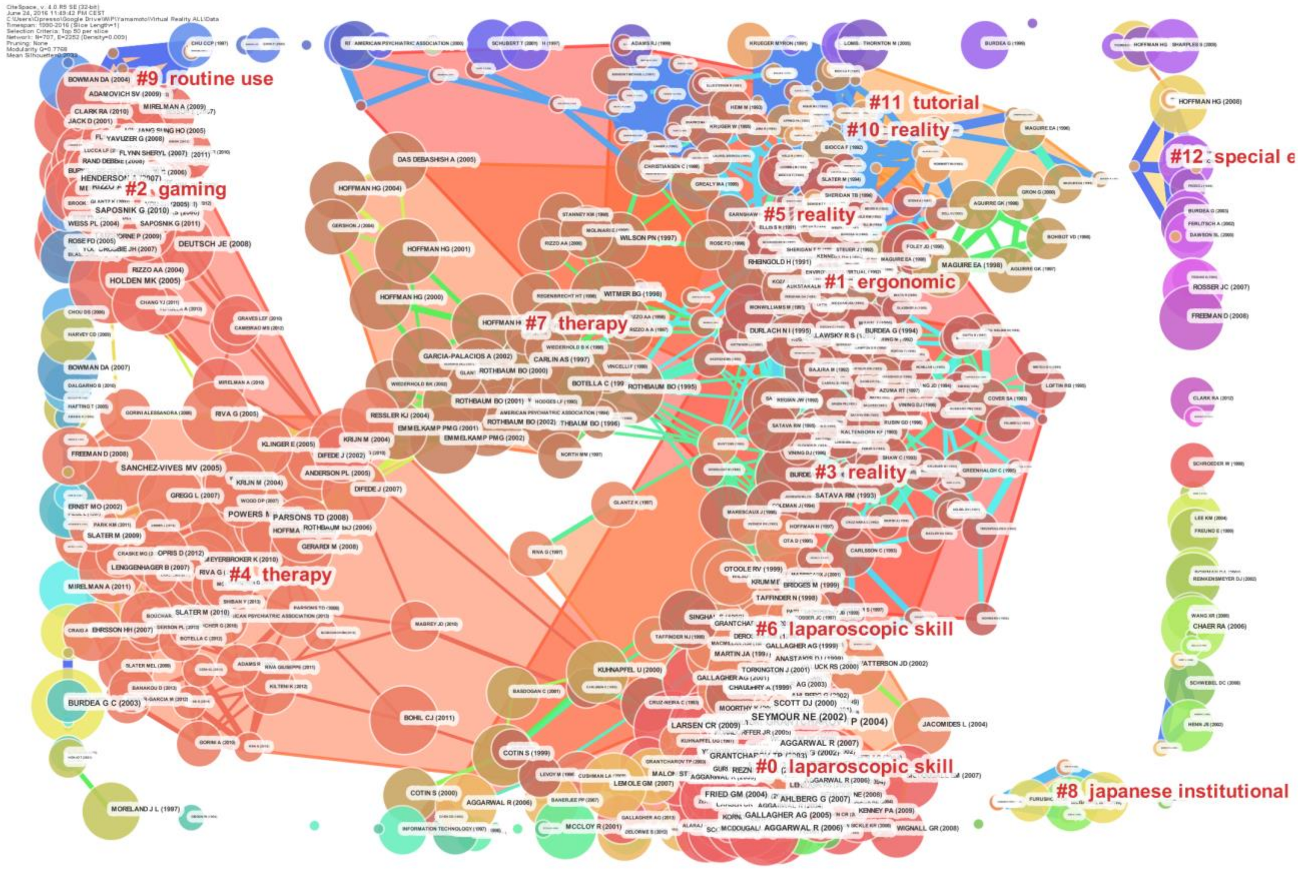


Figure 12. Document co-citation network by cluster: The dimensions of the nodes represent centrality, the dimensions of the characters represent the rank of the article rank and the red writing reports the name of the cluster with a short description that was produced with the mutual information algorithm; the clusters are identified with coloured polygons.

Table 8

Cluster ID and silhouettes as identified with two algorithms (Chen, C., Ibekwe-- SanJuan, F., & Hou, J. (2010)

ID	*Size*	*Silho-uette*	*Mean (Citee Year)*	*Label (TFIDF, tf*idf weighting algorithm)*	*Label (LLR, log-likelihood ratio, p-level)*
0	84	0.812	2005	(25.82) laparoscopic skill; (25.01) proficiency; (24.5) basic laparoscopic skill; (24.14) trainer; (23.79) establishing validity	training (143.21, 1.0E-4); performance (73.38, 1.0E-4); laparoscopic skill (72.93, 1.0E-4);
1	77	0.758	1992	(17.76) ergonomic; (17.66) reality; (16.83) virtual reality; (16.04) virtual environment; (15.76) assembly	ergonomic (54.1, 1.0E-4); virtual reality interface (34.63, 1.0E-4); developing virtual environment (34.48, 1.0E-4);

2	62	0.992	2007	(24.5) gaming; (24.5) wii; (24.47) stroke; (23.07) rehabilitation; (22.38) cerebral palsy	stroke (82.9, 1.0E-4); children (75.13, 1.0E-4); stroke rehabilitation (57.95, 1.0E-4);
3	61	0.758	1994	(15) reality; (14.66) virtual reality; (14.25) surgery; (14.1) telemedical information society; (13.73) chemistry	telemedical information society (34.85, 1.0E-4); gaining insight (23.21, 1.0E-4); next decade (18.32, 1.0E-4);
4	56	0.934	2008	(25.4) therapy; (23.55) exposure therapy; (22.41) disorder; (21.63) virtual reality exposure therapy; (20.99) posttraumatic stress	treatment (109.92, 1.0E-4); posttraumatic stress disorder (78.95, 1.0E-4); virtual reality exposure therapy (66.15, 1.0E-4);
5	49	0.885	1992	(16.03) reality; (15.31) virtual reality; (15.01) autistic children; (12.79) child; (12.79) children	autistic children (29.81, 1.0E-4); possibilities (23.84, 1.0E-4); communication (22.08, 1.0E-4);
6	41	0.855	1998	(17.6) laparoscopic skill; (16.95) direct observation; (16.95) measuring operative performance; (16.95) videotape; (16.15) measuring	laparoscopic skills training (52.73, 1.0E-4); measuring operative performance (40.97, 1.0E-4); videotape (40.97, 1.0E-4);
7	41	0.946	1998	(20.71) therapy; (18.76) exposure therapy; (17.85) exposure; (17.35) anxiety; (17.2) virtual reality exposure therapy	virtual reality exposure therapy (32.01, 1.0E-4); spider phobia (27.67, 1.0E-4); ptsd vietnam veteran (22.12, 1.0E-4);
8	38	1	1989	(30.67) japanese institutional mechanism; (30.67) systems perspective; (20.88) mechanism; (19.25) perspective; (17.97) system	japanese institutional mechanism (615.45, 1.0E-4); systems perspective (615.45, 1.0E-4); virtual reality (16.28, 1.0E-4);
9	21	1	1987	(23.27) routine use; (23.27) current application; (23.27) behavioral-assessment; (23.27) obstacle; (23.27) future possibilities	future possibilities (168.77, 1.0E-4); routine use (168.77, 1.0E-4); current application (168.77, 1.0E-4);
10	18	0.934	1991	(12.45) reality; (12.26) virtual-reality; (9.73) medicine; (9.07) virtual reality; (5.71) technology	virtual-reality (88.95, 1.0E-4); medicine (34.87, 1.0E-4); pretty interface (9.63, 0.005);
11	16	0.937	1990	(13.37) tutorial; (12.45) reality; (11.98) virtual reality; (11.12) virtual reality technology; (10.78) technology	tutorial (51.15, 1.0E-4); virtual reality technology (44.66, 1.0E-4); space (16.78, 1.0E-4);
12	12	1	1988	(20.05) special effect; (20.05) cyberspace; (13.65) space; (11.38) effect; (10.73) reality	special effect (128.6, 1.0E-4); cyberspace (128.6, 1.0E-4); virtual reality (27.79, 1.0E-4);
13	8	0.995	1997	(14.88) neural substrate; (14.88) human spatial navigation; (14.88) cognitive map; (11.56) navigation; (10.64) cognitive	neural substrate (72.6, 1.0E-4); human spatial navigation (66.58, 1.0E-4); cognitive map (66.58, 1.0E-4);
14	6	0.993	2008	(12.06) neurosurgery; (9.74) computer technology; (9.74) surgical application; (9.43) surgery; (8.55) teaching	neurosurgery (28.72, 1.0E-4); computer technology (18.1, 1.0E-4); surgical application (18.1, 1.0E-4);

The identified clusters highlight clear parts of the literature of VR research, making clear and visible the interdisciplinary nature of this field. However, the dynamics to identify the past, present, and future of VR research cannot be clear yet. We analysed the relationships between these clusters and the temporal dimensions of each article. The results are synthesized in Figure 9. It is clear that cluster #0 (laparoscopic skill), cluster #2 (gaming and rehabilitation), cluster #4 (therapy), and cluster #14 (surgery) are the most popular areas of VR research. (See Figure 12 and Table 8 to identify the clusters.) From Figure 13, it also is possible to identify the first phase of laparoscopic skill (cluster #6) and therapy (cluster #7). More generally, it is possible to identify four historical phases (colours: blue, green, yellow and red) from the past VR research to the current research.

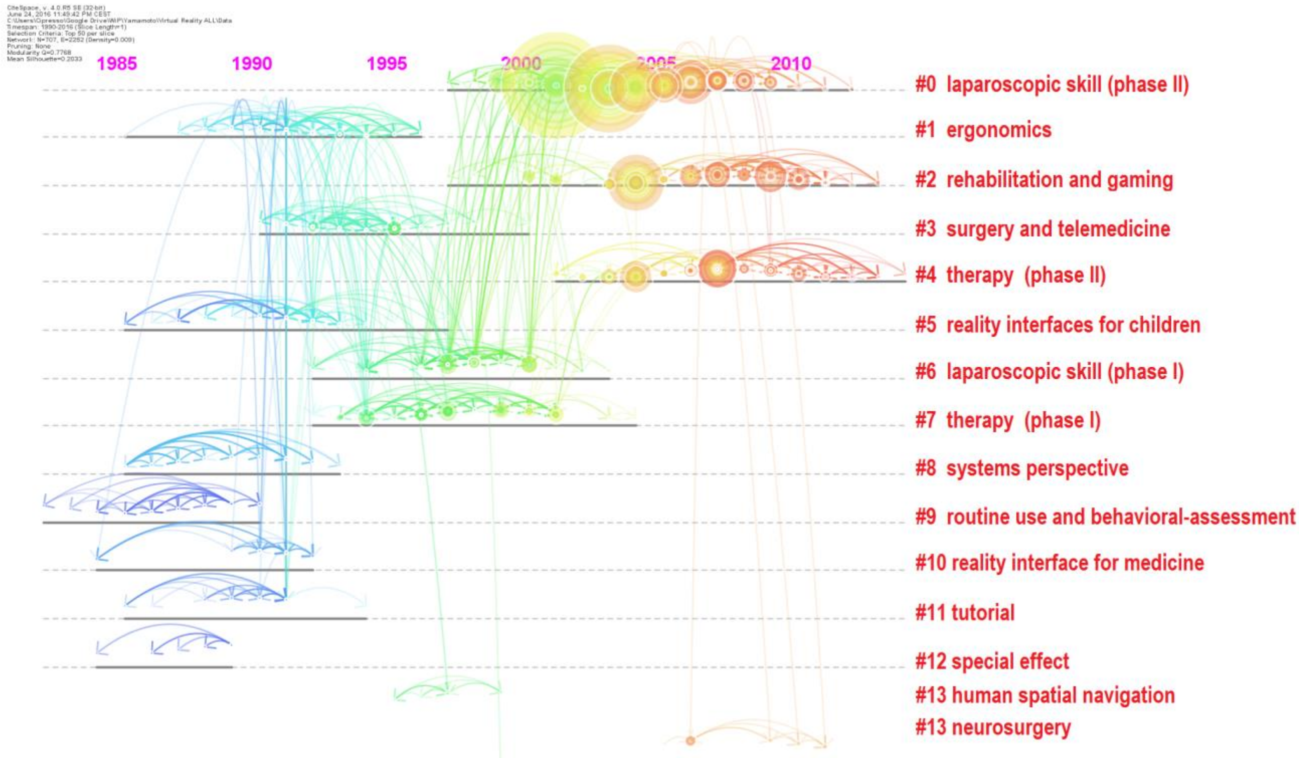


Figure 13. Network of document co-citation: The dimensions of the nodes represent centrality, the dimensions of the characters represent the rank of the article rank and the red writing on the right-hand side reports the number of the cluster, such as in table 2, with a short description that was extracted accordingly.

We were able to identify the top 486 references that had the most citations by using burst citations algorithm. Citation burst is an indicator of a most active area of research. Citation burst is a detection of a burst event, which can last for multiple years as well as a single year. A citation burst provides evidence that a particular publication is associated with a surge of citations. The burst detection was based on Kleinberg's algorithm (Kleinberg, 2002; 2003). The top-ranked document by bursts is Seymour NE (2002) in Cluster #0, with bursts of 88.93. The second is Grantcharov TP (2004) in Cluster #0, with bursts of 51.40. The third is Saposnik G (2010) in Cluster #2, with bursts of 40.84. The fourth is Rothbaum BO (1995) in Cluster #7, with bursts of

38.94. The fifth is Holden MK (2005) in Cluster #2, with bursts of 37.52. The sixth is Scott DJ (2000) in Cluster #0, with bursts of 33.39. The seventh is Saposnik G (2011) in Cluster #2, with bursts of 33.33. The eighth is Burdea GC (1996) in Cluster #3, with bursts of 32.42. The ninth is Burdea G C (2003) in Cluster #22, with bursts of 31.30. The tenth is Taffinder N (1998) in Cluster #6, with bursts of 30.96 (Table 9).

Table 9

Cluster ID and references of burst article

Cluster	***References***	***Year***	***Strength***	***Begin***	***End***	***1990 - 2016***
7	Rothbaum Bo, 1995, Am J Psychiat, V152, P626	1995	38,94	**1996**	2003	
3	Burdea Gc, 1996, Force Touch Feedback, V, P	1996	32,42	**1997**	2004	
6	Taffinder N, 1998, St Heal T, V50, P124	1998	30,96	**2000**	2006	
0	Scott Dj, 2000, J Am Coll Surgeons, V191, P272, Doi	2000	33,39	**2003**	2008	
0	Seymour Ne, 2002, Ann Surg, V236, P458, Doi	2002	88,93	**2004**	2010	
22	Burdea G C, 2003, Virtual Reality Tech, V, P	2003	31,30	**2004**	2010	
0	Grantcharov Tp, 2004, Brit J Surg, V91, P146, Doi	2004	51,40	**2005**	2012	
2	Holden Mk, 2005, Cyberpsychol Behav, V8, P187, Doi	2005	37,52	**2007**	2013	
2	Saposnik G, 2010, Stroke, V41, P1477, Doi	2010	40,84	**2012**	2016	
2	Saposnik G, 2011, Stroke, V42, P1380, Doi	2011	33,33	**2012**	2016	

Citation network and cluster analyses for AR

Looking at Augmented Reality scenario, the top ranked item by citation counts is Azuma RT (1997) in Cluster #0, with citation counts of 231. The second one is Azuma R (2001) in Cluster #0, with citation counts of 220. The third is Van Krevelen D W F (2010) in Cluster #5, with citation counts of 207. The 4th is Lowe DG (2004) in Cluster #1, with citation counts of 157. The 5th is Wu HK (2013) in Cluster #4, with citation counts of 144. The 6th is Dunleavy M (2009) in Cluster #4, with citation counts of 122. The 7th is Zhou F (2008) in Cluster #5, with citation counts of 118. The 8th is Bay H (2008) in Cluster #1, with citation counts of 117. The 9th is Newcombe RA (2011) in Cluster #1, with citation counts of 109. The 10th is Carmigniani J (2011) in Cluster #5, with citation counts of 104.

Figure 14 Network of document co-citations: The dimensions of the nodes represent centrality, the dimensions of the characters represent the rank of the article rank, and the numbers represent the strengths of the links. It is possible to identify four historical phases (colors: blue, green, yellow and red) from the past AR research to the current research.

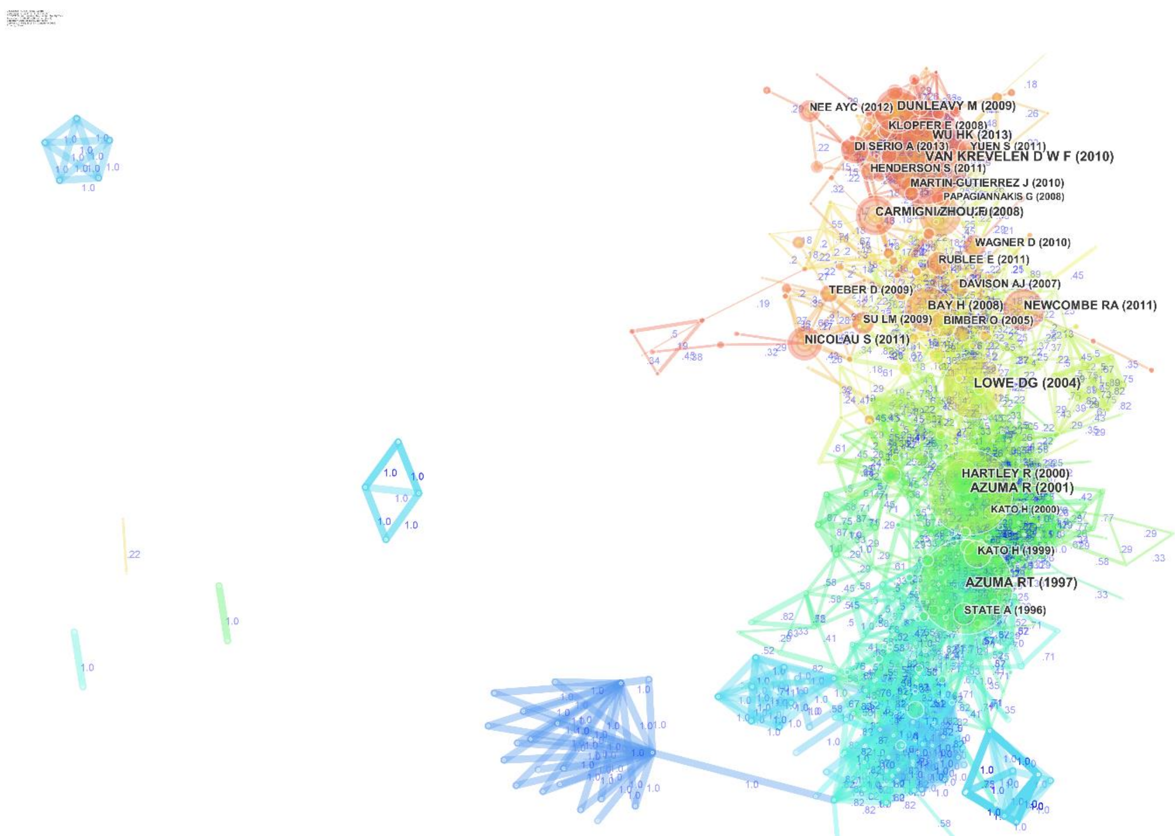


Figure 15. Network of document co-citations: The dimensions of the nodes represent centrality, the dimensions of the characters represent the rank of the article rank, and the numbers represent the strengths of the links. It is possible to identify four historical phases (colors: blue, green, yellow and red) from the past AR research to the current research.

The network of document co-citations is visually complex (Figure 14) because it includes thousands of articles and the links among them. However, this analysis is very important because can be used to identify the possible conglomerate of knowledge in the area, and this is essential for a deep understanding of the area. Thus, for this purpose, a cluster analysis was conducted (González-Teruel et al., 2015; Klavans &Boyack, 2015; Chen et al., 2010). Figure 15 shows the clusters, which are identified with the two algorithms in Table 10.

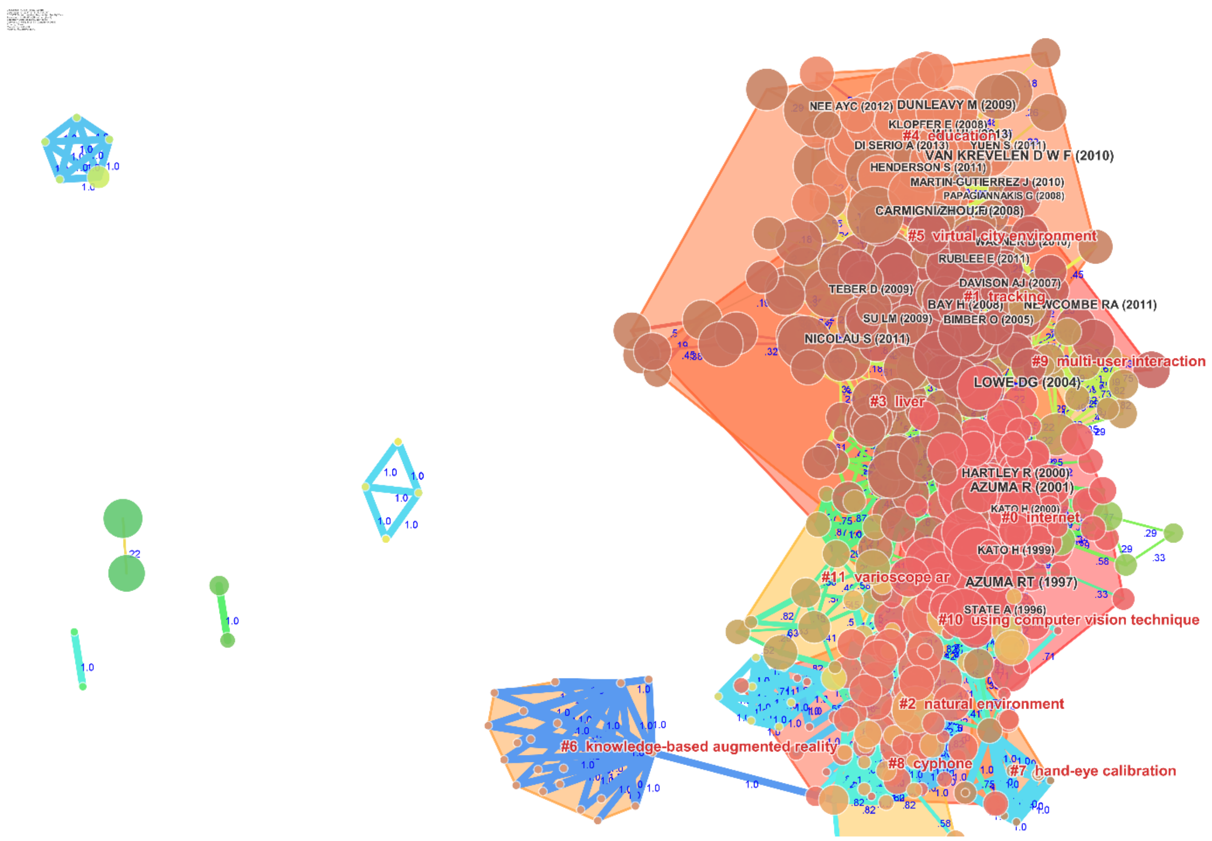


Figure 16. Document co-citation network by cluster: The dimensions of the nodes represent centrality, the dimensions of the characters represent the rank of the article rank and the red writing reports the name of the cluster with a short description that was produced with the mutual information algorithm; the clusters are identified with coloured polygons.

Table 10

Cluster ID and silhouettes as identified with two algorithms (Chen, C., Ibekwe--SanJuan, F., & Hou, J., 2010).

ID	***Size***	***Silho-uette***	***Mean (Citee Year)***	***Label (TFIDF, tf*idf weighting algorithm)***	***Label (LLR, log-likelihood ratio, p-level)***
0	122	0.669	1999	(18.41) internet	internet (39.96, 1.0E-4)
1	66	0.806	2007	(16.67) tracking	mobile phone (47.52, 1.0E-4)
2	65	0.827	1994	(17.48) natural environment	natural feature tracking (57.72, 1.0E-4)
3	56	0.89	2004	(17.33) liver	laparoscopic surgery (30.43, 1.0E-4)
4	50	0.943	2011	(19.32) education	education (64.26, 1.0E-4)
5	48	0.86	2007	(15.96) virtual city environment	virtual city environment (32.68, 1.0E-4)
6	20	0.997	1989	(21.65) knowledge-based augmented reality	knowledge-based augmented reality (250.67, 1.0E-4)
7	19	0.926	1992	(19.32) hand-eye calibration	hand-eye calibration (104.98, 1.0E-4)

The identified clusters highlight clear parts of the literature of AR research, making clear and visible the interdisciplinary nature of this field. However, the dynamics to identify the past, present, and future of AR research cannot be clear yet. We analysed the relationships between these clusters and the temporal dimensions of each article. The results are synthesized in Figure 12. It is clear that cluster #1 (tracking), cluster #4 (education), and cluster #5 (virtual city environment) are the current areas of AR research. (See Figure 16 and Table 10 to identify the clusters.) It is possible to identify four historical phases (colors: blue, green, yellow and red) from the past AR research to the current research.

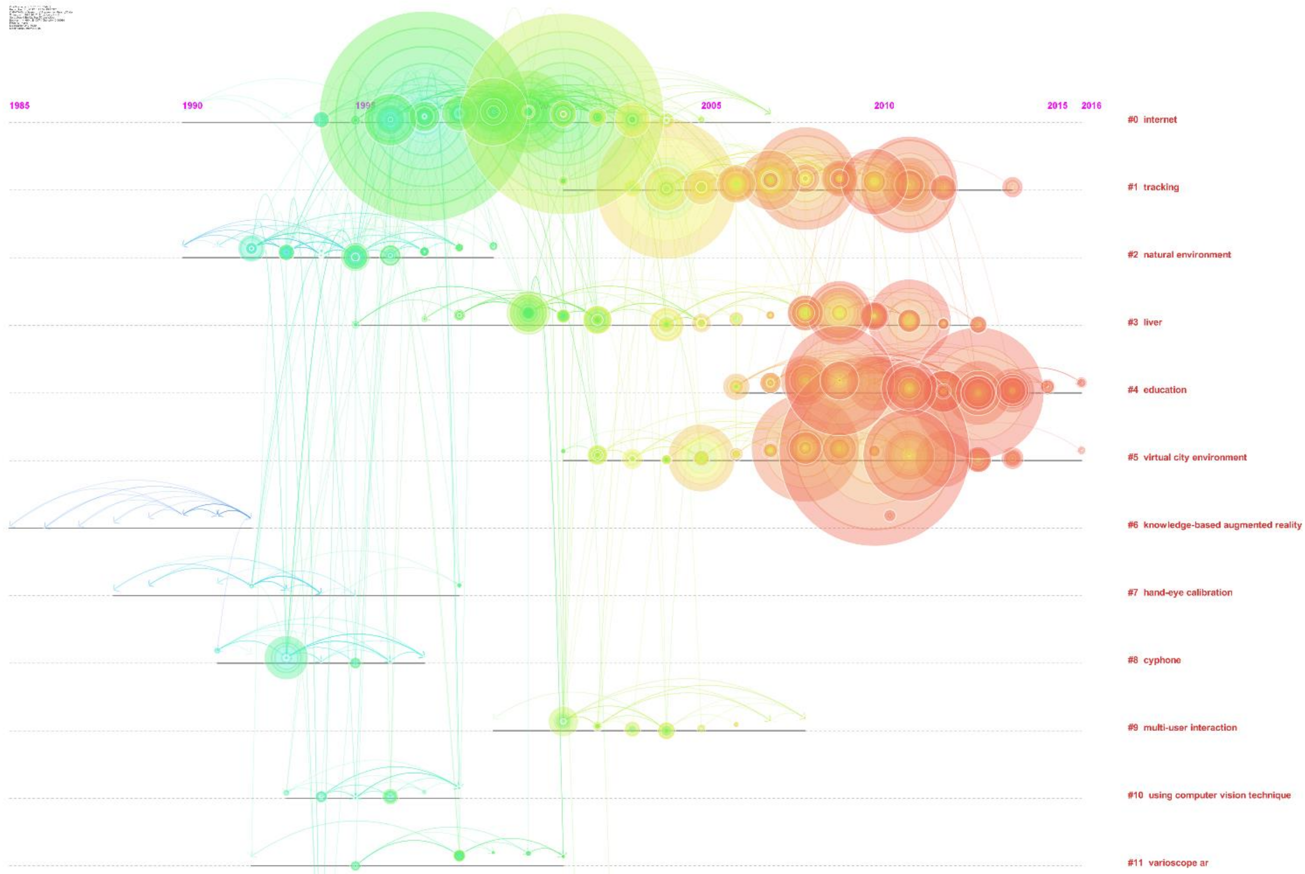


Figure 17. Network of document co-citation: The dimensions of the nodes represent centrality, the dimensions of the characters represent the rank of the article rank and the red writing on the right-hand side reports the number of the cluster, such as in table 8, with a short description that was extracted accordingly.

We were able to identify the top 394 references that had the most citations by using burst citations algorithm. Citation burst is an indicator of a most active area of research. Citation burst is a detection of a burst event, which can last for multiple years as well as a single year. A citation burst provides evidence that a particular publication is associated with a surge of citations. The burst detection was based on Kleinberg's algorithm (Kleinberg, 2002; 2003). The top ranked document by bursts is Azuma RT (1997) in Cluster #0, with bursts of 101.64. The second one is Azuma R (2001) in Cluster #0, with bursts of 84.23. The third is Lowe DG (2004) in Cluster #1, with bursts of 64.07. The 4th is Van Krevelen D W F (2010) in Cluster #5, with bursts of 50.99. The 5th is Wu HK (2013) in Cluster #4, with bursts of 47.23. The 6th is Hartley R (2000) in Cluster #0, with bursts of 37.71. The 7th is Dunleavy M (2009) in Cluster #4, with bursts of 33.22.

The 8th is Kato H (1999) in Cluster #0, with bursts of 32.16. The 9th is Newcombe RA (2011) in Cluster #1, with bursts of 29.72. The 10th is Feiner S (1993) in Cluster #8, with bursts of 29.46.

Discussion

Our findings have profound implications for two reasons. At first the present work highlighted the evolution and development of VR and AR research and provided a clear perspective based on solid data and computational analyses. Secondly our findings on VR made it profoundly clear that the clinical dimension is one of the most investigated ever and seems to increase in quantitative and qualitative aspects, but also include technological development and article in computer science, engineer, and allied sciences.

Figure 13 clarifies the past, present, and future of VR research. The outset of VR research brought a clearly-identifiable development in interfaces for children and medicine, routine use and behavioural-assessment, special effects, systems perspectives, and tutorials. This pioneering era evolved in the period that we can identify as the development era, because it was the period in which VR was used in experiments associated with new technological impulses. Not surprisingly, this was exactly concomitant with the new economy era in which significant investments were made in information technology, and it also was the era of the so-called 'dot-com bubble' in the late 1990s. The confluence of pioneering techniques into ergonomic studies within this development era was used to develop the first effective clinical systems for surgery, telemedicine, human spatial navigation, and the first phase of the development of therapy and laparoscopic skills. With the new millennium, VR research switched strongly toward what we can call the clinical-VR era, with its strong emphasis on rehabilitation, neurosurgery, and a new phase of therapy and laparoscopic skills. The number of applications and articles that have been published in the last five years are in line with the new technological development that we are experiencing at the hardware level, for example, with so many new, head-mounted displays, and at the software level with an increasing number of independent programmers and VR communities.

Finally, Figure 16 identifies clusters of the literature of AR research, making clear and visible the interdisciplinary nature of this field. The dynamics to identify the past, present and future of AR research cannot be clear yet, but analysing the relationships between these clusters and the temporal dimensions of each article tracking, education, and virtual city environment are the current areas of AR research. AR is a new technology that is showing its efficacy in different research fields, and providing a novel way to gather behavioural data and support learning, training, and clinical treatments.

Looking at scientific literature conducted in the last few years, it might appear that most developments in VR and AR studies have focused on clinical aspects. However, the reality is more complex; thus, this perception should be clarified. Although researchers publish studies on the use of VR in clinical settings, each study depends on the technologies available. Industrial development in VR and AR changed a lot in the last 10 years. In the past, the development involved mainly hardware solutions while nowadays, the main efforts pertain to the software when developing virtual solutions. Hardware became a commodity that is often available at low cost. On the other hand, software needs to be customized each time, per each experiment, and this requires huge efforts in term of development. Researchers in AR and VR today need to be able to adapt software in their labs.

VR and AR developments in this new clinical era rely on computer science and vice versa. The future of VR and AR is becoming more technological than before, and each day, new solutions and products are coming to the market. Both from software and hardware perspectives, the future of AR and VR depend on huge innovations in all fields. The gap between the past and the future of AR and VR research is about the "realism" that was the key aspect in the past versus the "interaction" that is the key aspect now. First 30 years of VR and AR consisted of a continuous research on better resolution and improved perception. Now, researchers already achieved a great resolution and need to focus on making the VR as realistic as possible, which is not simple. In fact, a real experience implies a realistic interaction and not just great resolution. Interactions can be improved in infinite ways through new developments at hardware and software levels.

Interaction in AR and VR is going to be “embodied,” with implication for neuroscientists that are thinking about new solutions to be implemented into the current systems. For example, the use of hands with contactless device (i.e., without gloves) makes the interaction in virtual environments more natural. The Leap Motion device (https://www.leapmotion.com/) allows one to use of hands in VR without the use of gloves or markers. This simple and low-cost device allows the VR users to interact with virtual objects and related environments in a naturalistic way. When technology is able to be transparent, users can experience increased sense of being in the virtual environments (the so-called sense of presence).

Other forms of interactions are possible and have been developing continuously. For example, tactile and haptic device able to provide a continuous feedback to the users, intensifying their experience also by adding components, such as the feeling of touch and the physical weight of virtual objects, by using force feedback. Another technology available at low cost that facilitates interaction is the motion tracking system, such as Microsoft Kinect, for example. Such technology allows one to track the users’ bodies, allowing them to interact with the virtual environments using body movements, gestures, and interactions. Most HMDs use an embedded system to track HMD position and rotation as well as controllers that are generally placed into the user’s hands. This tracking allows a great degree of interaction and improves the overall virtual experience.

In this scenario, it is clear that the future of VR and AR research is not just in clinical applications, although the implications for the patients are huge. The continuous development of VR and AR technologies is the result of research in computer science, engineering, and allied sciences. The reasons for which from our analyses emerged a “clinical era” are threefold. First, all clinical research on VR and AR includes also technological developments, and new technological discoveries are being published in clinical or technological journals but with clinical samples as main subject. As noted in our research, main journals that publish numerous articles on technological developments tested with both healthy and patients include *Presence: Teleoperators & Virtual Environments*, *Cyberpsychology & Behavior* (Cyberpsychol BEHAV), and *IEEE Computer Graphics and Applications* (IEEE Comput Graph). It is clear that researchers in psychology, neuroscience, medicine, and behavioural sciences in general have been investigating whether the technological developments of VR and AR are effective for users, indicating that clinical behavioural research has been incorporating large parts of computer science and engineering. A second aspect to consider is the industrial development. In fact, once a new technology is envisioned and created it goes for a patent application. Once the patent is sent for registration the new technology may be made available for the market, and eventually for journal submission and publication. Moreover, most VR and AR research that that proposes the development of a technology moves directly from the presenting prototype to receiving the patent and introducing it to the market without publishing the findings in scientific paper. Hence, it is clear that if a new technology has been developed for industrial market or consumer, but not for clinical purpose, the research conducted to develop such technology may never be published in a scientific paper. Although our manuscript considered published researches, we have to acknowledge the existence of several researches that have not been published at all. The third reason for which our analyses highlighted a "clinical era" is that several articles on VR and AR have been considered within the Web of Knowledge database that is our source of references. In this article, we referred to "research" as the one in the database considered. Of course, this is a limitation of our study, since there are several other databases that are of big value in the scientific community, such as IEEE Xplore Digital Library, ACM Digital Library, and many others. Generally, the most important articles in journals published in these databases are also included in the Web of Knowledge database; hence, we are convinced that our study considered the top-level publications in computer science or engineering. Accordingly, we believe that this limitation can be overcome by considering the large number of articles referenced in our research. Considering all these aspects, it is clear that clinical applications, behavioural aspects, and technological developments in VR and AR research are parts of a more complex situation compared to the old platforms used before the huge diffusion of HMD and solutions. We think that this work might provide a clearer vision for stakeholders, providing evidence of the current research frontiers and the challenges that are expected in the future, highlighting all the connections and implications of the research in several fields, such as clinical, behavioural, industrial, entertainment, educational, and many others.

EXPANSE: a novel narrative serious game for the behavioral assessment of cognitive abilities.

Chicchi Giglioli, I. A., 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.

Abstract

EFs are a set of processes that supports many cognitive domains as goal setting, monitoring, planning, and cognitive-behavioural flexible control. Currently, many standardized paper-and-pencil tests or scales are used to assess EFs. These tests are easy to administer, score, and interpret but present some limitations in terms of generalizability of behaviours in real life. More recently, Information and Communication Technology has provided a higher ecological validity in the EFs assessment. In order to increase the ecological validity, we have developed a serious game (SG), named EXPANSE, which aim was to compare the participants' game performance (latency times, and correct answers) with the results obtained in the traditional tasks and scales. 354 healthy subjects participated to the study and the findings showed significant correlations among standard tasks and the serious game. The exploratory nature of the present study, on one hand, highlighted that SG could be an additional behavioral tool to assess EFs and, on the other, we need further investigations, including clinical populations, for better defining the game sensitivity toward EF components. Finally, the results show that serious games are a promising technology for the evaluation of real cognitive behavior along with traditional evaluation.

Introduction

Cognitive functions are critical human aspects involving the nature and heterogeneity of everyday experience. Cognitive functioning is a broad term referring to higher-level mental processes of processing information involving the ability to sustain attention, inhibiting responses, planning behaviors, maintaining goals and information in mind, and solving problems (Chan et al., 2008; Lezak, 1982). These processes include a subset of functions of the brain, called executive functions (EF), such as attention, inhibitory control, cognitive flexibility, planning, and problem solving, governing goal-directed behaviors and adapting responses in accordance with specific situations (Diamond, 2013; Stuss & Alexander, 2000). EFs are essential abilities in every aspect of life and in the past two decades, EFs have become a major focus of research in psychology (Baler & Volkow, 2006; Diamond, 2005; Lui & Tannock, 2007; Tavares et al., 2007). Indeed, EFs dysfunctions have been found in many mental disorders including addictions (Baler & Volkow, 2006), depression (Tavares et al., 2007), and attention deficit disorder (Diamond, 2005; Lui & Tannock, 2007), producing inabilities in many everyday living activities, such as study, recreational activities, social relationships and work (Chevignard et al., 2000).

Currently, many standardized paper-and-pencil tests or scales are used to assess EFs, such as the Dot Probe Task, Stroop Test, Go/NoGo, the TMTA-B, the Tower of London and the Wisconsin Card Sorting Test (WCST) (Burgess et al., 2009). These tests are easy to administer, score, and interpret but present some limitations in terms of ecological validity, limiting the generalizability of behaviors in real life (Spooner & Pachana, 2006). Ecological validity refers to the capacity of a test to predict individual's real-world performance and the current EF measures are abstract, decontextualized and not able to capture the real dynamic and complex performance in daily life activities (Burgess et al., 2009; Chaytor & Schmitter-Edgecombe, 2006). Indeed, various studies showed that low scores on traditional measures, do not inevitably entail poor executive behaviors in real life and vice-versa (Barker et al., 2004; Chevignard et al., 2000; Manchester, 2004). In accordance with these findings, new instruments for evaluating the complexity of EFs in situations similar to daily ones, such as the Behavioral Assessment of Dysexecutive Syndrome (BADS), has been developed and tested (Wilson et al., 1996). The BADS consists of six subtests and a Disexecutive Questionnaire (DEX) that assess everyday cognitive, emotional, and behavioral changes together with a self-report and an independent rater questionnaire (Wilson et al., 1996). The BADS has been successfully tested in different clinical populations (Perfetti et al., 2010; Roy et al., 2015; Spitoni et al., 2018)showing good validity (Wilson et al., 1998), the DEX showed some limitations in measuring performance during real-life tasks (Barker et al., 2011).

More recently, the use of Information and Communication Technology (ICT) has substantially increased, potentially providing a higher ecological validity in the EFs assessment. On one hand, various applications of virtual reality (VR) have been developing for the assessment of cognitive functions (Knight & Titov, 2009; Martínez-Pernía et al., 2017; Parsons & Rizzo, 2008). Currently, most of the studies are in experimental phase, attempting to verify the construct and/or ecological validity comparing healthy and clinical populations to support rehabilitation interventions (Lo Priore et al., 2003 ; Marié et al, 2003 ; McGeorge et al., 2001 ; Rand et al., 2005, 2009). For example, Lo Priore et al. (2003) compared a 3D-VR and a 2D store in which patients vs. healthy subjects had to explore the environments and solve six sequences of tasks, ordered for complexity and created to stimulate executive functions, programming, categorical abstraction, short-term memory and attention. For the ecological validity, the authors used physiological (skin conductance response: SCR), neuropsychological, and questionnaire measures, showing a significantly higher SCR during tasks in the 3D-VR condition. Rand et al. (2005, 2009) developed a virtual mall in which participants (healthy and post-stroke) have been involved in a shopping task EFs-based. The results showed that the 3D task is sensitive to differentiate between the two groups and that it positively correlated with traditional measures, providing support to construct and ecological validity. However, VR approach shows some limitations. First, only few studies have established a significant construct and ecological validity of VR environment (Besnard et al., 2016). Indeed, if, theoretically, VR could enhance the ecological validity on cognitive assessment, predicting more realistic functional behaviors in real-life activities, another relevant issue to take into consideration in the conceptual creation of simulated daily activity could be represented by the heuristic bias of past experience. The success or failing of past experiences can have an influence on future behaviors (Skinner, 1953). More in detail, subjective perceptions of past behaviors influenced behaviors in future experiences. This phenomenon depends on cognitive activity, such as the memory of experiences, personality and motivational aims that influence the judgment of memory of past experience and the future. Second, VR can create cyber sickness, including discomfort, fatigue, headache, and nausea (Johnson, 2005). Third, at technological level, the development of VR is high-cost and implicates maintenance (Parsey et al., 2013), as well as VR systems require specific settings, such as lighting and large space that limit their application in clinical or educational fields (Rizzo et al., 2004; Werner et al., 2009).

Serious games (SG) constitute another promising approach that is showing to be positively related to a variety of cognitive abilities such as concentration, attention and working memory (Andrews, 2011; Connolly et al., 2012). SG can be defined as games developed for specific purposes that vary in accordance with the aims, the technology involved, and the interaction (Burke et al., 2009; Chatham, 2007). Furthermore, SG allows creating multiple tasks that incorporate several executive functions to carry out behaviors as if the participant was in the real world (Lamberts et al., 2010; Poulin et al., 2013). Serious games have been proposed for psychological interventions (Fleming et al., 2017), due to three advantages: appealing, engagement, and effectiveness. As mentioned above, computer games are widespread and the amount of time expended playing game is increasing (Andrade et al., 2014; Marchand & Hennig-Thurau, 2013), theoretically allowing reaching people that no receive treatments (Andrade et al., 2014). Second, games are fun and able to create engagement enhancing the motivation and reducing drop-out (Batterham et al., 2008; Fleming et al., 2016). Third, SG can provide safe and responsive environments in which users can test, shape, change, and learn new behaviors (Cheek et al., 2015; Fleming et al., 2014). So far, the use of SGs has been well addressed for clinical treatments and poorly addressed for behavioral assessment (Andrews, 2011; Connolly et al., 2012; Poulin et al., 2013). The present study, propose the use of a SG, named EXPANSE, for EF assessment with high ecological validity. EXPANSE has been created starting from the cognitive constructs related to EFs and the traditional assessment tests. A narrative storytelling game, settled in a spaceship, which aim is to discover a new planet where to live because the Earth has become uninhabitable, has been created for leading participants in the play. EXPANSE is focused on the participant that is the protagonist in the interactive story that drives him/her in the EF's situations and activities (for more details see the material and methods section). In addition, the participant can explore and navigate in the environment, manipulate and interact with objects. The narrative nature gives the context to the activities to be solved by the participant, and for solving them he/she needs to concentrate, evaluate and decide strategies.

Material and Methods

Subjects

A total of 354 healthy subjects (MMSE > 24; 177=women and 177=men; Mean Age=39.72; SD=8.90) participated in this study (Folstein et al., 1975). Subjects were provided by a panel company, which operates with an incentive system and managed the survey and task responses of this research. Before participating in the study, each participant received written information about the study and was required to give written consent for inclusion in the study. The study obtained ethical approval by the Ethical Committee of the Polytechnic University of Valencia.

Questionnaires

Four surveys were administrated before participants completed the tasks: demographic questionnaire (e.g. gender, age, level of computer games expertise), Cognitive Flexibility Scale (Martin & Ruben, 1995), Attentional Control Scale (Derryberry & Reed, 2002) and Barratt Impulsiveness Scale (Barratt, 1959).

Tasks

The tasks were developed using Unity 5.5.1f1 software and completed on a personal computer. The c# programming language was applied using the Visual Studio tool. Participants completed a total of 14 tasks (6 standard tasks; 8 self-designed games) randomly presented. Computerized versions of these standard tasks were administered: Dot Probe Task neutral version published by Miller and Fillmore (2010), the neutral pictures (20 in total) were selected from the International Affective Picture System (IAPS) (Lang, 2005); Go/NoGo Task (Fillmore et al., 2006); Stroop Test (Stroop, 1935); Trail Making Task, paper-and-pencil-based version published by Reitan (1955); Wisconsin Card Sorting Test (Grant & Berg, 1948) and Tower of London - Drexler (TOL^{DX}) (Culberston & Zillmer, 1999)

Eight games were designed, each one according to one of the standard tasks mentioned, and presented in one overarching game-scenario. Four of these tasks were aimed to measure attention (Dot Probe Task: AT1 – AT2; Go/NoGo task: AT3; Stroop Test: AT4), while three of the games were thought to measure cognitive flexibility (Trail Making Task: CF1; Wisconsin Card Sorting Test: CF2 – CF3), and one game was intended to assess planning (Tower of London –Drexler: PL1).

- AT1- “The takeoff: As a pilot, you will have to take off the spaceship” - A cross in the middle of the screen was followed by a pair of pictures that were presented together. After that, participants were ordered to press the E or I keys if the following X appeared on the left or on the right respectively.
- AT2: “Resources: You have reached a new planet! There are a lot of resources that you must organize”- This task followed the same dynamic than AT1, but in this case, we used different pictures according with another context.
- AT3: “The aliens: In space there are many elements that you must deal with. The most dangerous are the aliens that want to attack the ship” - Four kinds of objects could appear in front of the spacecraft that the participant was driving: a petrol tank, a meteorite, a spacecraft or an alien. These objects appeared in pairs, first the users saw the meteorite or the petrol tank (cue); and then this object was followed by the spacecraft or the alien (target). The subject was instructed to shoot the alien (go) and they had to do nothing if they saw the spacecraft (no-go) (Figure 17).

Figure 18. Screenshot of the SG-AT3. Two objects appeared in pairs (a) and participants were instructed to shoot the alien (b).

- AT4: “The strongbox: The oxygen valve has broken! You must fix it, but it´s behind a coded strongbox” - A strongbox with a screen, a group of rotatory letters and two switches was shown. A colored word was presented in the vault screen, and participants were asked to use the switches to rotate the letters until writing the color of the word in the screen, ignoring its meaning (Figure 18).

Figure 19. Screenshot of the SG-AT4. Participants were asked to use the switches to rotate the letters until writing the color of the word in the screen.

- CF1: “Water and food (Part A): You have run out of water. To obtain it you will have to pump up the level of the water little by little. (Part B) This time you have run out of water and food. To obtain them you will have to pump and cultivate at the same time” - This game was inspired in Gonzalez and colleagues´ work (2013). As in the standard task, this was a two-part puzzle task. The first part of the task (A) consisted in a 10x10 squares puzzle, that subjects were asked to go through, making only horizontal or vertical movements by matching consecutive bottles from the empty one until the full one. The second part of the task (B) consisted in a 10 x 10 squares puzzle. In this case, participants were asked to match two kinds of objects: bottles witch different levels of water and plants in different growing levels. The matching criterion could be changed along the task (from water to plant or from plant to water).
- CF2: “The orchard: Your orchard is running empty. You need to make grow a series of plants” - Participants were shown four plant groups; they could see one plant representing each group. Each one of these plants had different characteristics: number of branches, fruit, and growing state. As in the Wisconsin Card Sorting Task, only one of these features was the rule by which a new plant should be matched with one of the four groups. The classification criterion changed during the task (Figure 19).

Figure 20. Screenshot of the SG-CF2. The card on the right should be matched with one of the four groups on the left following one of the classification criterions.

- CF3: "Fuel: Your ship has run out of fuel. Inside the ship you will find some elements that can be used to start the turbine" - In this task, participants needed to activate a turbine, and they were given four elements to do it (water, wind, fire and magnetism), that could be mixed by two. In each trial, there was only one combination that could activate the turbine, but this criterion was changeable along the task (Figure 20).

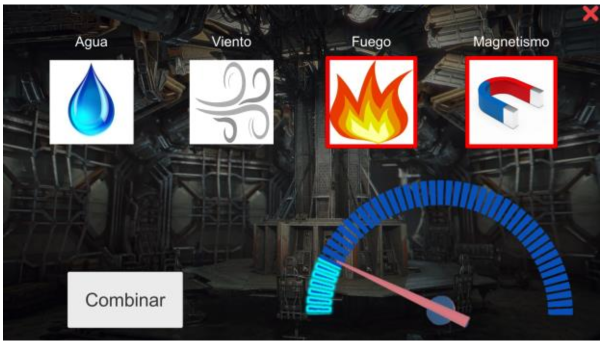


Figure 21. Screenshot of the SG-CF3. There was only one combination that could activate the turbine, but this criterion was changeable along the task.

- PL1: "Escape: You are locked up in the control room. You need to open the door if you want to escape" - The game started in the control room, where participants had stayed locked, such a way that they needed to open the door in order to complete the task. Different objects were randomly distributed along the room, and the subjects must combine and use them in order to accomplish the goal.

Data analysis

The analyses were performed using SPSS version 22.0 (Statistical Package for the Social Sciences for Windows, Chicago, IL) for PC. After identifying and removing outliers using Mahalanobis distance, we verified the assumption of normality applying Kolmogorov Smirnov. Internal consistency of the scales was assessed via Cronbach´s alpha.

Pearson correlations were calculated between standard tasks and self-report, and the SG version dependent variables (DVs). These correlations were consolidated by T-Student test for both standard tasks and self-report with DVs.

Results

The following analysis was carried out with the enduring 301 subjects (see Table 11). The assumption of normality was confirmed (Kolmogorov Smirnov *p* > .05) and the internal consistency of the self-report scales was measured (Cronbach´s alpha $\alpha_{attention}$ = .839, $\alpha_{cognitive\ flexibility}$ = .825, $\alpha_{impulsiveness}$ = .732; bootstrap 95%). Table 12 shows descriptive data for the Standard Task and Serious Games variables.

Table 11

Demographic data of participants (n=301)

Demographic (n=301)	**Mean (SD) [Range]**		
Age	39.78	(8.73)	[25-55]
Gender (M/F)	149/152		
Use of technologies level (High/Low)	143/158		

Table 12

Mean (SD) and [Range] values for Standard Tasks and Serious Games variables

Standard Tasks	**Mean(SD)[Range]**	**Serious Games**	
Dot ProbeTask		**AT1**	**AT2**
Correct answers (%)	0,98 (0.06) [0.46-1.00]	0,98(0.04) [0.53-1.00]	0,98(0.05)[0.53-1.00]
Latency time (s)	0,47 (0.07) [0.13-0.77]	0,47(0.06) [0.34-0.71]	0,46(0.07)[0.07-0.70]
Go/Nogo Task		**AT3**	
Latency time (s) (correct answers-go)	0,21(0.03)[0.11-0.41]	0,32(0.10)[0.004-0.59]	
Stroop Test		**AT4**	
Latency time (s)	1,40(0.26)[0.90-2.45]	2,77(1.12)[0.91-8.10]	
Trail Making Task		**CF1**	
Total time A (s)	39,39(9.25)[21.80-91.15]	60,38(31.87)[19.01-215.98]	

Total time B (s)	55,11(14.61)[31.29-119.97]	57,87(29.76)[22.76-332.36]	
Tower of London		**PL1**	
Total time (s)	338,36(119.91)[158.45-1081.21]	336,94(141.02)[29.42-789.12]	
Initial time (correct answers) (s)	10,81(5.17)[3.72-37.91]	15,43(18.55)[1.36-143.19]	
Total score	26.22(2.72)[18.00-30.00]	11.16(1.68)[2.00-17.00]	
Wisconsin Card Sorting Test		**CF2**	**CF3**
Correct answers (%)	0,63(0.24) [0.01-0.92]	0,54(0.25)[0.01-0.92]	0,59(0.15)[0.08-0.87]
Latency time (s)	1,08(0.44)[0.01-2.08]	1,04(0.47)[0.01-2.10]	0.86(0.74)[0.08-9.29]
Perseverative responses	21,90(23.69)[0.00-118.00]	29,73(25.50)[0.00-118.00]	4.66(0.89)[0-5]

Standard Tasks – Serious Games correlations

Pearson correlations calculated for each Standard Task and its related SG showed statistically significant main relationships between variables (Figure 21).

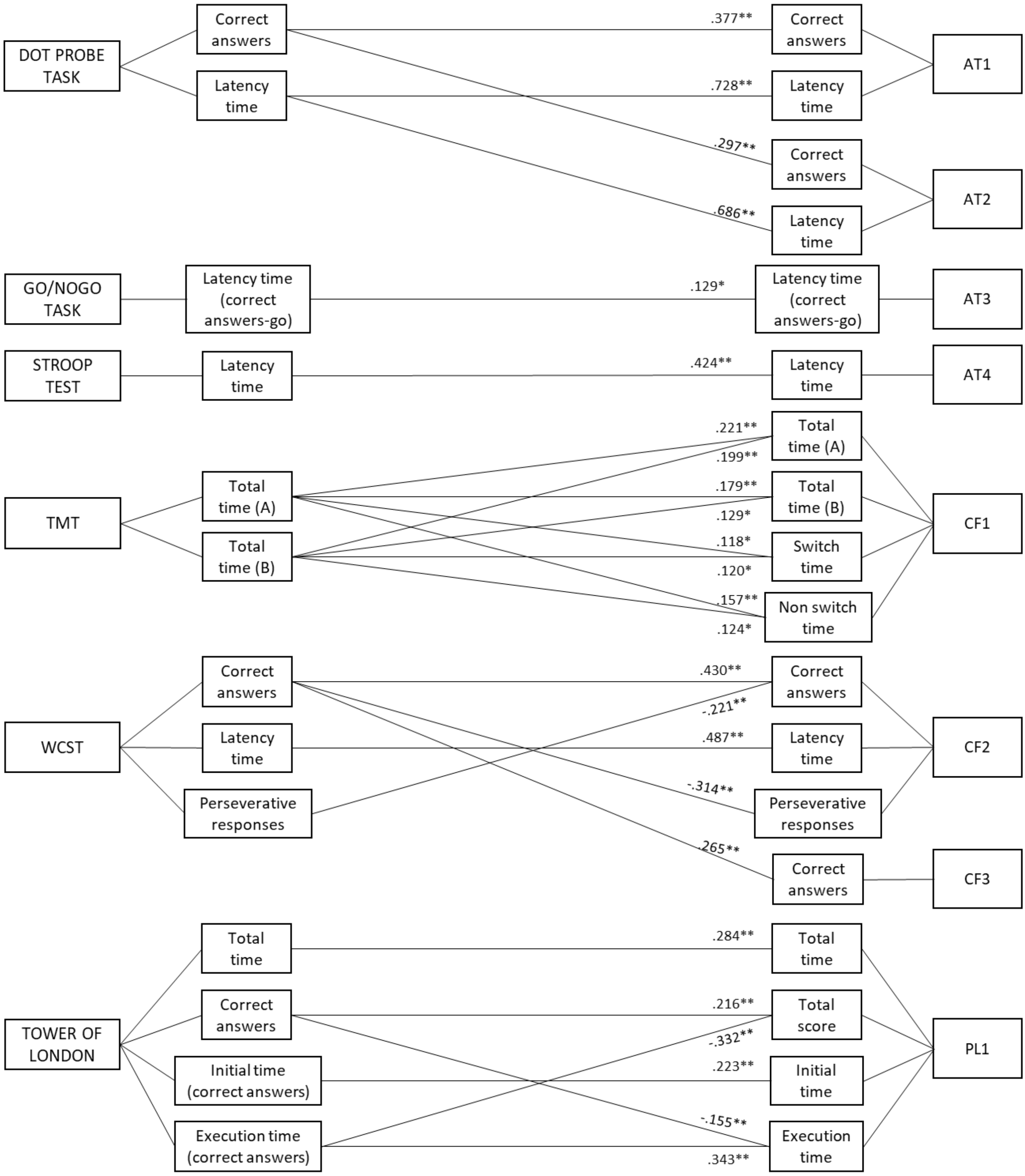


Figure 22. Results of Pearson correlations between associated Standard Tasks and Serious Games Variable. *p<.05 **p<.01.

T-test

Furthermore, for verifying whether the average values of the SG performance differ significantly from the reference neuropsychological standard tasks values we conducted the t-test analyses (Table 13).

Table 13

Results of t-test between Standard Tasks and Serious Games performance. 1 = Dot Probe, 2 = Go/Nogo, 3 = Stroop, 4 = TMT, 5 = WCST,6 = Tower of London; CA = Correct answers, LT = Latency time, TT = Total time, PR = Perseverative responses, IT = Initial time, ET = Execution time, ST = Switch time, NST= Non switch time. LT in Go/Nogo Task and in AT3 = Latency time (go-correct answers). *p<.05 **p<.01.

		CA		**LT**		**TT (A)**		**PR**		**IT**		**ET**	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
							1						
AT2	**LT**			0.01**	0.056								
							2						
AT3	**LT**			-0.107**	0.101								
							3						
AT4	**LT**			-1.318**	0.971								
							4						
CF1	**TT (A)**					-20.916**	31.369						
							5						
CF2	**CA**	0.088**	0.274798626										
	PR							-7.708**	32.577				
CF3	**LT**			0.198**	0.854								
	PR							16.880**	23.958				
							6						
PL1	**IT**									-4.624**	18.112		
	ET											-299.939**	136.450
	TS	-1.342**	1.647										

Questionnaires – Serious Games

Only AT3 latency time (correct answers-go) had weak significant negative correlation (r = -.182**, p-value = .002) with Cognitive Flexibility Scale results that has been confirmed by t-test results. Low cognitive flexibility group registered higher latency times (correct answers-go) (mean = 0.34) in contrast with high cognitive flexibility group (mean = 0.30) (p-value= .000).

Discussion and conclusions

Cognitive functions, such as EFs, are important components of human behavior in daily living activities that can be related to important functional disabilities (Barker et al., 2010; Morton & Barker, 2010; Riva, 2009; Royall et al., 2002; Shallice & Burgess, 1991). Due to the importance of cognitive functions, it is essential to investigate the distinctive structures of the EFs in real life behaviors. As previously described, one of the most crucial limitations is the poor ecological validity of the EF standardized tests (Chan et al., 2008; Chaytor & Schmitter-Edgecombe, 2006). Indeed, studies on EFs have revealed that even if a subject can achieve good results in the traditional tests, he/she may encounter difficulties in real life activities (Barker et al., 2004; Chevignard et al., 2000; Manchester et al., 2004).

SG can provide a more ecological simulated environment in which the participant has an active role allowing evaluating real cognitive behaviors (Riva, 2009). In SG tasks, such as EXPANSE, it is possible to simulate life-like activities, which can involve several and different goals to achieve and the cognitive flexibility and planning to elaborate different strategies to accomplish them and to inhibit inappropriate actions.

Starting from these premises, the main aim of this study was to investigate the plausibility and feasibility of a narrative SG, to integrate the traditional evaluation of EFs with a more ecologically valid assessment. The high Cronbach's alphas for attention, cognitive flexibility, and impulsiveness suggest good internal consistency across trials and the positive correlations with the standard tests provide initial evidences for convergent validity.

Our results yielded four main findings. First, EXPANSE correlated with those standard tasks closely associated. More specifically as for the latency times, the SG tasks attentional-related strongly correlated with the standard attentional tasks, such the Dot Probe Task, as well as moderate correlations were found between SG tasks cognitive flexibility and planning-based and the standard tasks. Second, according to the accuracy of the responses, EXPANSE games moderately correlated with standard tasks. Third, comparing the SG with standard tasks performance, EXPANSE provided longer times and lower correct answers than standard tasks. This last result could depend on the different interaction required by the two technological systems. Fourth, we did not find strong correlations between SG and self-report and we found one low negative correlation between one SG behavioural measure and the cognitive flexibility scales showing that low cognitive flexibility scores registered higher latency times in contrast with high cognitive flexibility group scores. This could suggest discordance between EFs' behavioural and self-report measures on revealing and discriminating different aspects of EFs.

Even though the present findings are relevant, they present some limitations. First, healthy subjects that limit the sensitivity of the results composed the sample. Second, considering the use of a SG, it would be important to also assess the individual's perception of EXPANSE usability, as well as individual differences in cognitive functioning to the field of ICT. Indeed, several studies showed that individual variables, such as game skills or knowledge, personality, age and gender can affect user interactions and performance (Aykin & Aykin, 1991). Further studies are required to examine plausibility, feasibility of EXPANSE game, mainly regarding its sensitivity, including clinical populations, such as patients that show altered performance in the executive domains and/or that present other cognitive symptoms (aphasia, apraxia, neglect, memory deficits) and/or motor problems (hemiparesis to upper and/or lower limbs)as well as its reliability and predictive validity according to the different criterions and the distinctive components of EFs. Nevertheless, the present study has yielded initial evidence on the potential use of a more ecological clinical rehabilitative tool to identify the (dys-) functional cognitive status in real-simulated contexts along with traditional evaluation.

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

Chicchi Giglioli, I. A., de Juan Ripoll, C., Parra, E., & Alcañiz Raya, M. (2019). Are 3D virtual environments better than 2D interfaces in serious games performance? An explorative study for the assessment of executive functions. *Applied Neuropsychology: Adult*, 1-10.

Abstract

Executive functions refer to higher-order cognitive processes that supervise and guide goal-directed and adaptive behaviours 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 behaviours 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.

Introducton

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; Stuss & Alexander, 2000). 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; Nyhus & Barceló, 2009) 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; Chelune & Moehle, 1986; 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; Chaytor et al., 2006; 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; Li et al., 2016). 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 previously mentioned, 2D and 3D-SG are promising tools for neuropsychological assessment of EFs and various environments have been mainly developed and tested for 3D systems (Parsey & Schmitter-Edgecombe, 2013), such as virtual beach (Elkind et al., 2001), classroom (Climent & Banterla, 2012; Díaz-Orueta et al., 2014; 2019; Díaz-Orueta, 2017; 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.

Materials and Method

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 14 shows descriptive data of participants according the two conditions.

Table 14

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%

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.

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, 2005); 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.

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) (Figure 22). These tasks were used to assess attention, inhibition control, impulsivity, planning, and cognitive flexibility. Table 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.

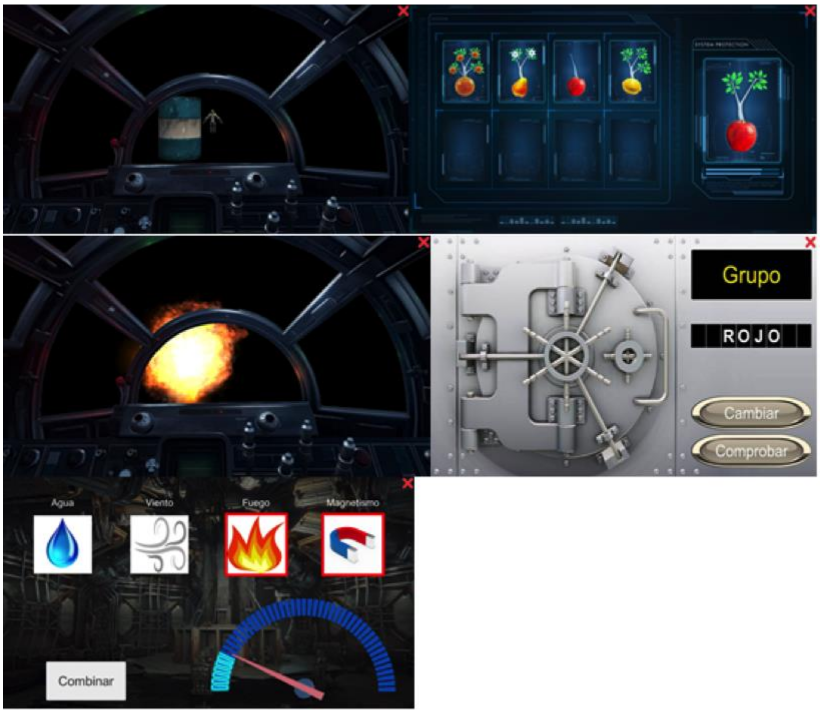


Figure 23. Screen shot of the SG tasks.

Table 15

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

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).

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 $\alpha = 0.05$.

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 16 and 17 show descriptive data for the ST and 2D/3D-Serious Games.

Table 16

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]

Table 17

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]

Questionnaires – Serious Game correlations

Table 18 reports Pearson correlations calculated for each questionnaire and the serious games performance in the two conditions.

Table 18

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.

Standard Tasks – Serious Game correlations

Pearson correlations calculated for each standard task and serious games in the two conditions have been reported in Table 19.

Table 19

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.

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 20).

Table 20

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.

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.

Individuals' variables in cognitive abilities using a narrative serious game

Parra, E., de Juan Ripoll, C., Raya, M. A., & **Chicchi Giglioli, I. A.** (2018, November). Individuals' Variables in Cognitive Abilities Using a Narrative Serious Game. In *Joint International Conference on Serious Games* (pp. 109-119). Springer, Cham.

Abstract

Age, gender, and education represent crucial variables in the assessment and interpretation of traditional neuropsychological measures as regards the executive functions (EF). Currently, traditional measures are showing limitations in capturing real life behaviors and new technologies, such as serious games, are allowing creating more real situations with higher ecological validity. In the present study, we applied a serious game approach to investigate individual variables-related differences in the EF assessment. 268 healthy subjects participated in the study, completing 14 tasks (6 standard tasks; 8 serious games) randomly presented. The results showed that younger participants completed tasks in less time than older and with higher correct answers. Furthermore, males registered shorter reaction times, while females showed higher percentages of correct answers. The university studies group obtained higher total score and correct answers than high school studies group. Finally, since the study involved technology, we divided the group in high and low use technology level, obtaining that participants with a lower level of use technologies reported higher latency times and lower correct answers in high order EF tasks than the group with higher level of use of technology. As the traditional measure, these findings suggest that individuals' differences are critical variables to consider in the development of more ecological measures for the assessment of EFs.

Introduction

Executive functions (EFs) are a set of basic and higher-order cognitive processes involved in the monitoring and control of everyday life behaviors for the achievement of established goals (Diamond, 2013; Stuss & Alexander, 2000). Basic EFs include attention, control inhibition, working memory, and cognitive flexibility, and higher-order EFs involve multi-basic EFs, including planning and problem-solving abilities (Chan et al., 2008; Lezak, 1982; Lezak et al., 2004). In addition, EFs do not only play an important role in cognitive behaviors related to the achievement of established goals, but also in emotional and social situations (Anderson, 2002). Currently, the neuropsychological assessment and an accurate interpretation of EFs performance measures are crucial variables that distinguish normal cognitive functioning from cognitive impairments (Jurado & Rosselli, 2007; Stuss & Alexander, 2000). For accurate neuropsychological interpretation, the tests are adjusted according to socio-demographic variables such as age, gender, and education that influence cognitive performance (Beeri et al., 2006; Gladsjo et al., 1999; Stricks et al., 1998). Numerous age studies on normal functioning showed that the highest performance of EFs abilities is between 20 and 29 years, weakening progressively in later adulthood (Birch & Bloom, 2004; Craik & Salthouse, 2011; De Luca et al., 2003). Otherwise, educational level has related to higher EFs performance (Ardila et al., 2010; Das et al., 2006; Liu et al., 2011; Mathuranath et al., 2007; Ostrosky-Solis et al., 1998). Individuals with higher education performed better than individuals with lower education. Regarding to results on gender, the scientific literature showed more complexity. Some studies, on one hand, showed that women performed better on verbal tasks and men in visual-spatial tasks (Messinis et al., 2007; Proust-Lima et al., 2008; Van Hooren et al., 2007). On other hand, other studies revealed weaker or no differences gender-related on EFs performance (Bagherpoor et al., 2014; Unger, 1979; Zarghi & Zarindast, 2011). Furthermore, the traditional EF tests require simple responses to single stimuli and tend to be decontextualized, and abstract, not reflecting the complex multi-tasks in daily life that demand more composite series of behaviors, limiting in this way their ecological validity (Burgess et al., 2009; Chaytor et al., 2006; Chevignard et al., 2000; Spooner & Pachana, 2006). Indeed, clinical studies have showed that even if patients are able to perform well as healthy subjects on traditional tests, they experience difficulties in everyday life activities (Barker et al., 2004; Chevignard et al., 2000; Manchester et al., 2004). Ecological validity refers to the generalization of the results and individual performance of a research study to real settings of everyday life (Chaytor et al., 2006). Serious games (SG) are games with an established aim that can represent a novel approach to simulate more real-EFs situations and able to capture dynamic performance in real time. Currently, SGs are especially showing efficacy in rehabilitation interventions (Fleming et al., 2017), and less in psychological assessment. The main advantages provided by SGs include appealing, since nowadays technologies are commonly used in daily life, being able to achieve more people (Andrade et al., 2014); engagement because games are fun and able to increase motivation, decreasing patients' drop-out (Fleming et al., 2014); and effectiveness since people can experience new behavior in a safe environment (Fleming et al., 2014).

Starting from these premises, the aim of the present study was to assess, through a narrative SG, three distinct domains of EFs involving attentional and inhibition control, planning and cognitive flexibility among healthy individuals with a wide age range of 25-55 years, considering comparative gender and education differences, as well as differences in using technology.

Material and Methods

Subjects

A total of 268 healthy subjects (Mean age=39.19; SD=8.65) participated in this study. Before participating in the study, each participant received written information about the study and was required to give written consent for inclusion in the study. The study obtained ethical approval by the Ethical Committee of the Polytechnic University of Valencia.

Questionnaire

Participants completed a demographic questionnaire, about their age, gender, education and level of use of technologies.

Tasks

The tasks were developed using Unity 5.5.1f1 software and completed on a personal computer. The Visual Studio tool was used applying c# programming language. Participants completed a total of 14 tasks (6 standard tasks; 8 serious games) randomly presented. Applying a programming code performed the randomization of the tasks.

Each of the SGs was designed according to one of the standard tasks (ST). Table 21 shows the ST administered and its correspondent SG.

Standard tasks

- Dot probe task (Miller & Fillmore, 2010)

A neutral version of the arrangement published by Miller and Fillmore (2010) was administered. During the task, a black cross-appeared in the middle of the screen, followed by a couple of neutral pictures that were presented together, 3 cm apart. We selected these neutral pictures (20 in total) from the International Affective Picture System (IAPS) (Lang et al., 2008). After the images disappear, an “X” emerged on one side of the screen. Participants were instructed to press the “E” key on the keyboard if the target appeared on the left, and the “I” key if the target appeared on the right.

- Go/Nogo Task (Fillmore et al., 2006)

A white rectangle emerged in the middle of the screen. This rectangle could be vertical or horizontal disposed (cue). After the rectangle appeared, it became into blue or green (target). Participants should press the spacebar when the rectangle became into green (go) and they didn´t have to press any key if the rectangle became into blue (nogo).

- Stroop Test (Stroop, 1935)

A colored word appeared in the middle of the screen. Participants were asked to indicate the color of the word, between four options, ignoring its meaning.

- Trail making task ((Fillmore et al., 2006)

This exercise was divided in two parts. In the first one (A), 25 numbers were randomly distributed along the screen; participants were asked to connect them consecutively (1-2-3,… 25), as quickly as they could and using the mouse. In the second part of the task (B), participants should match numbers and letters alternatively, and in consecutive order (1-a-2-b,… 13).

- Wisconsin Card Sorting Task (Grant & Berg, 1948)

Participants were shown four card piles, with its first card faced up. Each one of these cards had different characteristics: number of elements (one, two, three or four), color (red, yellow, blue or green) and shape (cross, triangle, circle or star). One of these features was the criterion by which a new card should be matched with one of the four piles. Participants should put each new card on the related pile, and the system gave feedback about the correct and wrong answers. The classification criterion changed along the task, in such a way that the participants should deduce the correct criterion in each trial based on the system feedback.

- Tower of London (Culberston & Zillmer, 1999)

Participants were shown a structure made by three sticks of different length connected with base and three colored balls (red, green and blue). Three balls filled in the longer stick, two balls filled in the medium one and only one ball filled in the smaller stick. A combination of balls distributed along the sticks appeared, and participants had to reproduce it using the minimum number of moves that they could.

Serious games

The narrative SG has been settled in a spaceship and the aim of the game player was to discover a new "earth". The eight games have been created and developed based on the EFs constructs and contextualized in the storytelling:

- 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”;
- 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”;
- 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”;
- 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”;
- CF2: “The orchard is empty” You have to grow up new plants. You have 4 kinds of plants based on fruit types, number of branches, number of fruits, and color fuits, and you have to decide in which group of plants you joint the new plant”;
- CF3: “Without fuel: your fuel supply is finished. To obtain fuel you have to activate the turbine. For activating you have to combine different elements two by two”;
- 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”;
- 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”.

Table 21

Standard Tasks, Serious Games administered, Outcome measures, and Cognitive functions assessed

ST	SG	Outcome measures	Cognitive functions assessed
Dot Probe Task	AT1	Latency times Correct Answers	Attention Inhibition control
	AT2	Latency times Correct Answers	Attention Inhibition control
Go/NoGo Task	AT3	Latency times Correct Answers	Attention Inhibition control
Stroop Test	AT4	Latency times Correct Answers	Attention Inhibition control
Trail Making Task	CF1	Total times Latency times Correct Answers	Attention Cognitive flexibility

Wisconsin Card Sorting Test	CF2	Latency time Correct answers Perseverative responses	Cognitive flexibility
	CF3	Latency time Correct answers Perseverative responses	Cognitive flexibility
Tower of London	PL1	Total score Initial time Execution time Total time	Planning

We developed each task to measure a specific executive function: four of the SGs were designed to measure attention (AT1, AT2, AT3 and AT4), three of them were aimed to assess cognitive flexibility (CF1, CF2 and CF3) and one of the SGs was thought to evaluate planning abilities (PL1). All these tasks started with a brief contextualization, to create a situational context in which the games will fit in.

Data analysis

The analyses were performed using SPSS version 22.0 (Statistical Package for the Social Sciences for Windows, Chicago, IL) for PC. After verifying the assumption of normality applying Kolmogorov Smirnov, individual differences between groups were obtained. We calculated the individual differences between groups per age, gender and level of use of technologies using t-test, and we applied ANOVA test for analyzing differences among groups of participants per education.

Results

Table 22 shows the descriptive data of participants:

Table 22

Demographic data of participants (n=268)

Demographic data of participants	Mean (SD) [Range]
Age	39,19 (8.65) [25-55]
Age (18-40/41-64)	147/121
Gender (M/F)	133/135
Education (1= High School/ 2= University/ 3= Post-graduate)	107/110/51
Use of technologies level (H/L)	127/141

Individual differences per gender

Regarding the relation between ST tasks and SG-related, we found significant differences between genders in Stroop Test and the associated SG-AT4. In the Stroop-related SG, males registered shorter reaction times, while females showed higher percentages of correct answers. Furthermore, as concerns the exclusive narrative SG the results showed several significant differences. In the SG-PL1, females expend higher execution time than males while males obtained higher total score. Females registered also higher total time in SG-CF1 (A) (Figure 23).

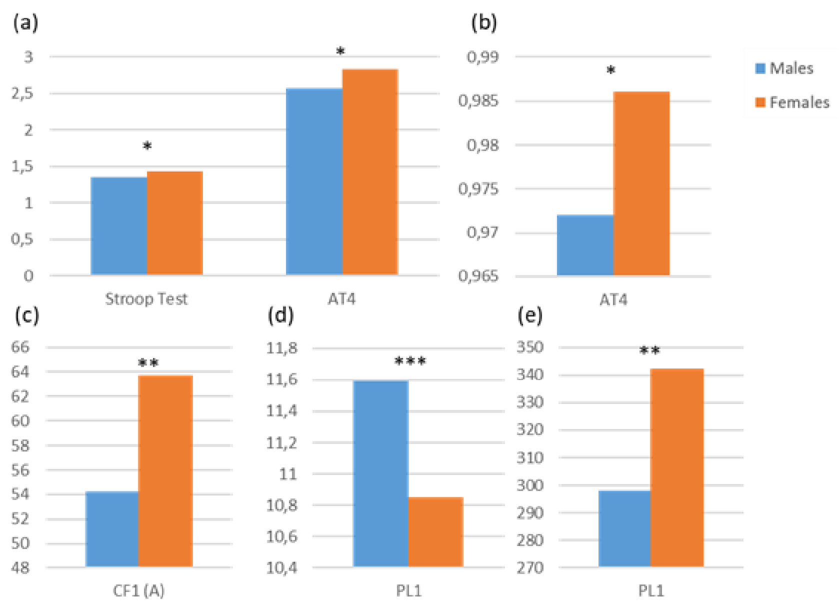


Figure 24. T-test results on significant differences between groups per gender. (a) LT=Latency time, (b) CA=Correct answers, (c) TT=Total time, (d) TS=Total score, (e) ET=Execution time, (f) PR=Perseverative Responses. * p<.05, ** p<.01, *** p<.001.

Individual differences per age

Figure 24 shows the significant differences found among groups per age, ST and SG-related. More in detail, we found analogous significant differences in age between the whole ST tasks and the related SGs: younger participants (18-40) registered lower reaction times and execution times than older subjects (41-64), and higher percentage of correct answers and total score both in the ST tasks and in the SG-related.

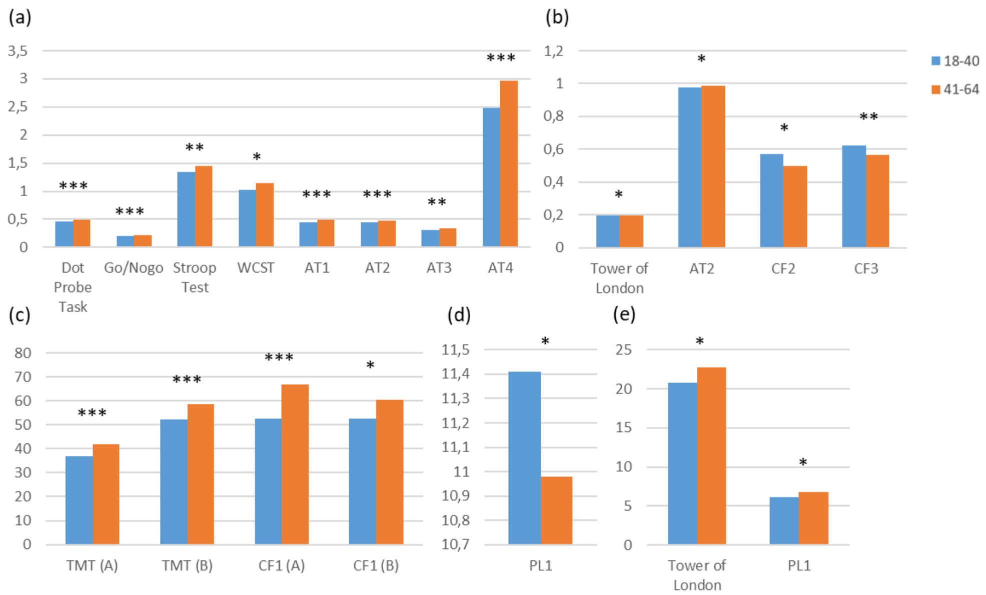


Figure 25. T-test results on significant differences between groups per age. (a) LT=Latency time, (b) CA=Correct answers, (c) TT=Total time, (d) TS=Total score, (e) ET=Execution time, (f) PR=Perseverative Responses. * p<.05, ** p<.01, *** p<.001

Individual differences per education

Table 25 shows the significant differences found among groups per education, ST and SG-related. The university studies group obtained higher total score and correct answers than high school studies group both in the WCST ST task and the SG-related, as well as in the Tower of London.

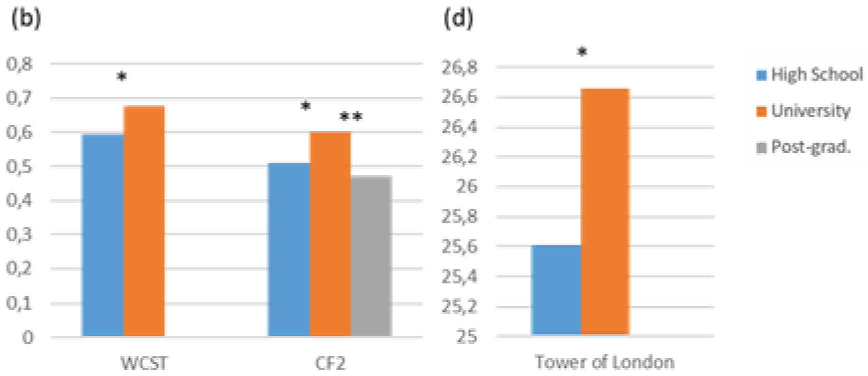
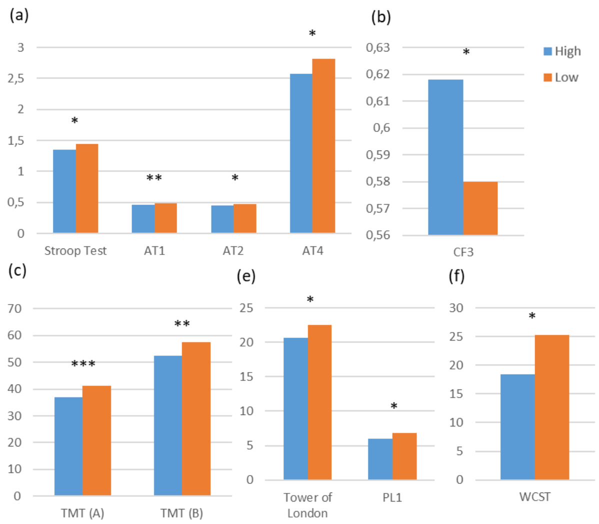


Figure 26. ANOVA results of significant differences between groups per education. (a) LT=Latency time, (b) CA=Correct answers, (c) TT=Total time, (d) TS=Total score, (e) ET=Execution time, (f) PR=Perseverative Responses. * p<.05, ** p<.01, *** p<.001.

Individual differences per level of use of technologies

Figure 26 shows the significant differences found among groups per level of use of technologies, ST and SG related. Participants with a low level of use of technologies registered higher latency times between Stroop test and the attentional SGs, as well as between the Tower of London test and the related SG as concerns the execution times. Furthermore, participants with a low level of use of technology showed lower percentage of correct answers in the SG-CF3, as well as more perseverative responses in WCST.

Figure 27. T-test results on significant differences between groups per level of use of technologies. (a) LT=Latency time, (b) CA=Correct answers, (c) TT=Total time, (d) TS=Total score, (e) ET=Execution time, (f) PR=Perseverative Responses. * p<.05, ** p<.01, *** p<.001.



Discussion and conclusions

The main of this study was to investigate the influence of gender, age, and education on three EFs' domains, including attentional and inhibition control, planning, and cognitive flexibility comparing traditional and standardized EFs' measures with a narrative SG.

The results on gender revealed differences on latency times of attentional (but not in inhibition control) ST and the SG between men and women, showing shorter reaction times in men. Conversely, women have showed higher latency time and better performance than men. Regarding the exclusive narrative SG, we found on cognitive flexibility higher total time in women than men, as well in planning time, and men showed a better performance than women. These results reflect the complexity in better definition the role of gender in EFs activities, showing consistency with the ambiguity of previous studies (Bagherpoor et al., 2014; Unger, 1979; Zarghi & Zarindast, 2011).

Regarding age, our results are consistent with the scientific literature, showing that younger participants registered lower reaction times and execution times than older and higher performance between the whole ST tasks and the related SGs (Birch & Bloom, 2004; Craik & Salthouse, 2011; De Luca et al., 2003). Even results on education reflected coherence with previous works, showing that individuals with higher education performed better than individuals with lower education in ST and the narrative SG (Ardila et al., 2010; Das et al., 2006; Liu et al., 2011; Mathuranath et al., 2007; Ostrosky-Solis et al., 1998). More in detail, the main finding about education refers to higher-order EFs performance, as cognitive flexibility and planning abilities, resulting on better performance in participants with higher educational levels. Finally, the level of use of technologies influenced the performance and our findings showed that people with a low level of technologies' use registered higher performance time both the ST and the SG-related. As well, our results showed that lower correct answers and more preservative responses are two possible effects that seem depend on the level of technologies' use, mainly in higher-order EFs.

In conclusion, our study presented two relevant main findings: firstly, we found similar significant differences in gender, age, and education between ST tasks and the related SG, and, secondly, the only narrative SG has been able to detect more significant differences as compared to the ST, especially in higher-order EFs. As a result, if traditionally neuropsychological tests scores are adjusted based on age and education, the new technological approaches, including more ecological measures for the assessment of EFs, should take into consideration also the level of use of technologies as individual variable.

Although, the present study presents some limitations and further studies, including clinical populations, will be need for enhancing the sensitivity, reliability, and validity in using technological approaches for more ecological neuropsychological tests.

A virtual versus an augmented reality cooking task based-tools: a behavioral and physiological study on the assessment of executive functions.

Chicchi Giglioli, I. A., Bermejo Vidal, C., & Alcañiz Raya, M. (2019). A virtual versus an augmented reality cooking task based-tools: a behavioral and physiological study on the assessment of executive functions. *Frontiers in Psychology*, *10*, 2529.

Abstract

Virtual reality (VR) and augmented reality (AR) are two novel graphics immersive techniques (GIT) that, in the last decade, have been attracting the attention of many researchers, especially in psychological research. VR can provide 3D real-life synthetic environments in which controllers allow human interaction. AR overlays synthetic elements to the real world and the human gaze to target allow hand gesture to act with synthetic elements. Both techniques are providing more ecologically environments than traditional methods, and most of the previous researches, on one side, have more focused on the use of VR for treatment and assessment showing positive effectiveness results. On the other, AR has been proving for the treatment of specific disorders but there are no studies that investigated the feasibility and effectiveness of augmented reality in the neuropsychological assessment. Starting from these premises, the present study aimed to compare the performance and sense of presence using both techniques during an ecological task, such as cooking.

The study included 50 cognitively healthy subjects. The cooking task consisted of 4 levels that increased in difficulty. As the level increased, additional activities appeared. The order of presentation of each exposure condition (AR and VR) was counterbalanced for each participant. The virtual reality-cooking task has been performed through "HTC/VIVE" and augmented reality

through "Microsoft HoloLens". Furthermore, the study recorded and compared the psychophysiological changes (heart rate and skin conductance response) during the cooking task in both conditions. To measure the sense of presence occurring during the two exposure conditions, subjects completed the SUSQ and the ITC-SOPI immediately after each condition.

The behavioral results showed that times are always lower in VR than in AR, increasing constantly in accordance with the difficulty of the tasks. Regarding physiological responses, the findings showed that AR condition produced more individual excitement and activation than VR. Finally, VR was able to produce higher levels of sense of presence than AR condition.

The overall results support that VR currently represents the GIT with greater usability and feasibility compared to AR, probably due to the differences in the human-computer interaction between the two techniques.

Introduction

Virtual reality (VR) and augmented reality (AR) are two novel graphic immersive techniques (GIT) that, in the last decade, have been attracting the attention of many researchers, especially in the fields of psychology and education (Fleming et al., 2017; Jensen and Konradsen, 2018; Chicchi Giglioli et al., 2015; Negut et al., 2016; Cipresso et al., 2018; Ventura et al., 2018; Germine et al., 2019). On one side, VR is an interactive and advanced computer technology that it can create real-simulated three-dimensional (3D) environment. Technologically, VR provides a wide field of view (FOV – the area angular size allowed to a user to see a scene) around 100 degree and the human-computer interaction can be ensured by various devices, such as head-mounted display (HMD) for the visual stimuli, headphone for the acoustic stimuli, controllers for hand interaction. These allow users to navigate and interact with the virtual environment, being felt them totally immersed in the virtual world. The accurate real-simulated 3D environment and the technological presence can help users to generate a sense of presence, defined as the feeling to "being in" the virtual environment (Gregg and Tarrier, 2007; Slater, 2009; Freeman et al., 2017; Parsons, 2015; Valmaggia et al., 2016). On the other, AR is a recent technology in which synthetic elements are incorporated in the physical world adding information to the users (Chicchi Giglioli et al., 2015; Ventura et al., 2018). The FOV is narrower than VR, included between 35 and 45 degrees and the interaction is ensured by various sensors integrated into the headband, like cameras that, through the human gaze to target, allow the real hands' interaction with the synthetic elements. AR, like VR, aims to provide high visual realism, fidelity of the experience, and presence, highly similar to the real one and adding real objects/information to real world. Visual realism and fidelity can depend on the FOV, accuracy, complexity of the systems, as well as on the user's interaction fidelity. Regarding the visual realism and fidelity, a wider FOV allows the user to see more of the scene at once and to use peripheral vision, while a narrower FOV, as in AR systems, may reduce distraction in the periphery and allow the user to focus on the area of interest in the scene (McMahan et al., 2012; Ragan et al., 2010, 2012). Furthermore, high accuracy and complexity on graphics can enhance the level of fidelity of the experience, allowing transferring the VR/AR learned behaviors in real-world or allowing to perform in the AR/VR world, as if the user were in the real-life (Dunkin et al., 2007; Seymour, 2008; Saposnik et al., 2010). Finally, interaction fidelity supposed that more is natural the interaction, higher is the fidelity (McMahan et al., 2012). However, comparison studies on different hand controllers showed that the more familiar, and less natural type of controller provided a best performance, although the participants appreciated the more natural interaction (McMahan et al., 2010). All these features are able to generate immersed and the psychological state to be present in the virtual and augmented environments (Slater, 2009). A valid and reliable measure for the sense of presence is the ITC- Sense of Presence Inventory (ITC-SOPI; Lessiter et al., 2001) that assess four dimensions: sense of physical space, engagement, ecological validity, and negative effects. Tang et al. (2004), compared the sense of presence between a VR and an AR environment, showing significantly higher score for sense of physical space for AR, and no significant differences in the other three dimensions, although all means were higher in the AR than VR condition.

According to this, at present, both techniques are providing advantages along with traditional scientific research procedures, providing accurate real-simulated stimuli control and behavior measurement of reactions times and scores and allowing researcher to address issues that would simply be difficult to pose in natural environments (Bohil et al., 2011; Germine et al., 2012; De Leeuw, 2015; Reimers and Stewart, 2015). In psychology both technologies have been extensively explored in the treatment of certain disorders, such as phobias, allowing patients learning and repeating new behaviors to cope with fearful stimuli in safe and reactive environments generating effectiveness in behavioral changes in real contexts (Chicchi et al, 2015; Ventura et al., 2018; Suso-Ribera et al., 2018). In psychological assessment, conversely, several VR applications have been developed for neuropsychological evaluation in order to improve the ecological validity of them (Pugnetti et al., 1998; Ku et al., 2003, 2004; Rizzo et al., 2004; Rand et al., 2007, 2009; Henry et al., 2012; Parsons et al., 2013; Cipresso et al., 2014; Diaz-Orueta et al., 2012, 2014). Traditional neuropsychological assessment consists of performance-based approach, involving paper-and-pencil and/or computerized tests, to assess a variety of cognitive processes, such as attention, memory, inhibition control, planning, cognitive flexibility, and the higher-order system of executive functions, that govern the cognitive processes to goal-directed and adaptive behaviors. These tests consist of a set of predefined and abstracts' stimuli delivered in a controlled setting that have proved moderate level of ecological validity in predicting real-functional performance (Elkind et al, 2001; Chaytor and Schmitter-Edgecombre, 2003, 2006). For example, the Tower of London is a neuropsychological measure for the assessment of executive functioning, specifically related to planning abilities, in which a target configuration of colored beads are presented to the participant and he/she is asked to compute the minimal number of steps (ranging from 1 to 5) to reach a target configuration. This test is a reliable and valid measure but it is abstract and decontextualized from the real-life activities.

In order to improve similarity between tests and real-life activities, several VR environments have been developed such as virtual mall/supermarket (Rand et al., 2007; 2009; Cipresso et al., 2014), and classroom (Rizzo et al., 2000, 2009; Diaz-Orueta et al., 2014). For example, Cipresso et al.

(2014) tested a virtual supermarket in which participants (patients with normal cognition, patients with mild cognitive impairments and cognitively healthy subjects) had to complete four shopping tasks. Findings revealed that the virtual shopping task was able to discriminate the performance among the three groups and that the virtual supermarket was more sensitive than traditional assessment in detecting cognitive impairments. Furthermore, a recent meta-analytic review (Negut et al., 2016) on VR applications in neuropsychological assessment showed moderate sensitivity and effect size in detecting cognitive impairments by comparing performance between health subjects and patients using both VR applications and traditional measures.

Despite the opportunities, that VR has been providing in psychological assessment, to our knowledge no previous studies have investigated the differences in behavioral responses to ecological tasks presented through AR compared to other methods - particularly VR.

Finally, both systems are also compatible with other neuroscientific tools such as wrist devices able to measure changes in electrodermal activity (EDA) and heart rate variability (HRV) (Poh et al., 2010; Garbarino et al., 2014). EDA and HRV showed consistent results with cognitive and information processing (Dawson et al., 2007; Sequeira et al., 2009) and can provide, together with behavioral data, implicit and objective responses to changing during activities.

Starting from these premises, the first aim of this study was to analyze and compare behavioral and physiological data collected before, during and after performing a cooking task in virtual reality (VR) and augmented reality (AR) environments. Second, the study aimed to determine the degree of presence, or the feeling of "being there", that produced virtual reality through the "HTC Vive" and augmented reality through "Microsoft HoloLens".

Materials and Method

Participants

The experimental sample included 50 healthy individuals (16 males and 34 women). Participants were recruited through local advertisement among college students and workers of the Polytechnic University of Valencia. The mean age was 25.96 ± 6.51. To be included in the study, participants were required to have a score higher than 24 in the “Mini-Mental State Examination” (MMSE) (Folstein et al., 1975). Before participating in the study, each participant was provided with written information about the study and required to give written consent for inclusion in the study. The study received ethical approval by the Ethical Committee of the Polytechnic University of Valencia. Table 23 includes the main sociodemographic data, such as age, gender, and education.

Table 23

Sociodemographic data of the participants (n= 50)

Sociodemographic data	**Mean (Standard Deviation)**	**[Range]**
Age	25.96 (6.51)	[18-48]
Gender (Man/Woman)	16/34	
Education (High school/Bachelor Degree/Postgraduate Degree)	6/28/16	

Psychological assessment

Before the experimental session, the following questionnaires were administered to each participant:

- Attentional Control Scale (ACS) (Derryberry and Reed, 2002): This scale is used to evaluate the attentional control and higher scores show a great ability to maintain voluntarily attention in a task, while low values are related to greater attention stiffness.
- Barratt Impulsiveness Scale (BIS-11) (Barratt, 1959; Oquendo et al., 2001): is a measure of impulsiveness and a score of 72 or more means that the individual is highly impulsive. Between 52 and 71 should be considered within the normal limits of impulsivity. Below 52 represents a subject excessively controlled.
- Cognitive Flexibility Scale (CFS) (Martin and Rubin, 1995): consists of 12 questions that are scored on 6 points Likert-scale; a score of 60 or more indicates that the individual has a high cognitive flexibility.

Furthermore, participants completed a total of 5 standard tasks (ST): Dot Probe Task (DOT) version published by Miller and Fillmore (2010); Go/NoGo Task (Fillmore et al, 2006); Stroop

Test, (Stroop 1992); Trail Making Task (TMTA-B), paper-and-pencil-based version published by Reitan (1958); and Tower of London - Drexler (TOLDX, Culbertson and Zilmer, 1999). The standard tasks were randomly presented and performed on a personal computer. Neuropsychological data performance of the participants are reported in Table 24.

After each exposure condition, the following presence questionnaires were administered to each participant:

• Slater-Usoh-Steed Questionnaire (SUSQ) (Slater and Steed, 2000): This post-hoc test consists of 3 questions that are evaluated on a scale of 7 points. The items evaluate the sensation of being in the environment, the extent to which the medium becomes the dominant reality and the magnitude in which it is remembered as a "place".

• ITC Sense Presence Inventory (ITC-SOPI) (Lessiter et al., 2001): This test consists of 42 items, evaluated on a scale of 5 points, and evaluates 4 dimensions of presence: the sense of physical space or spatial presence (SP), engagement (E), ecological validity (EV), and negative effects (NE).

Descriptive data on presence are reported in Table 25.

Table 24

Mean (M), standard deviation (SD), and range (Min., Max.) of values for questionnaires and standardized tasks. TT=Total Time; CA= Correct Answers; LT= Latency Time; TS=Total Score; ET= Execution Time.

Variables	M	SD	Min.	Max.
CFS	47.36	6.59	34	64
BIS_Cognitive	18.94	2.68	14	25
BIS_Motor	22.14	4.99	14	38
BIS_No Planning	24.16	4.87	15	39
BIS	65.24	9.46	50	91
ACS	55.44	8.07	41	69
DOT_TT	159,01	5,31	151,28	175,00

DOT_CA	0,99	0,01	0,96	1,00
DOT_LT	0,46	0,06	0,36	0,61
GONOGO_CA	0,99	0,02	0,93	1,00
GONOGO_LT	0,41	0,04	0,31	0,53
TMT_TTA	35,54	7,18	22,81	56,08
TMT_TTB	54,32	16,57	28,92	134,08
TMT_CAA	25,00	0,00	25,00	25,00
TMT_CAB	25,00	0,00	25,00	25,00
TORRE_TT	436,64	604,16	124,58	4492,58
TORRE_CA	9,56	0,99	5,00	10,00
TORRE_TS	25,44	3,46	14,00	29,00
TORRE_ET	20,24	7,37	7,18	42,44
STROOP_TT	3208,61	19516,38	75,09	137260,17
STROOP_CA	0,99	0,03	0,82	1,00
STROOP_LT	1,27	0,22		

Table 25

Mean (M) and standard deviation (SD) of values for presence questionnaires.

Variables	M	SD	Min.	Max.
SUSQ_AR	4.11	1.65	1.33	7
SUSQ_VR	5.85	1	3	7
SOPI_SP_AR	3.29	0.64	1.83	4.61
SOPI_E_AR	3.6	0.69	2.08	4.69
SOPI_EV_AR	3.21	0.86	1.6	5

SOPI_NE_AR	1.7	0.64	1	3.4
SOPI_SP_VR	3.81	0.57	2.39	4.67
SOPI_E_VR	4.21	0.49	3.08	5
SOPI_EV_VR	3.93	0.73	1.8	5
SOPI_NE_VR	1.52	0.51	1	3.25

Physiological assessment

At the beginning and during the experimental session, skin conductance response (SCR) and heart rate variability (HRV) were recorded to obtain subjects' physiological responses to VR and AR cooking task. SCR and HRV are considered indexes of arousal responses (Boucsein, 1992; Electrophysiology, 1996). The physiological signals were acquired using Empatica E4 device, including E4 Manager software to record and export raw signals. The sampling frequency in the SCR signal was acquired at 4Hz, and 64 Hz for HRV, inside a window time from 1 to 2.5s with an amplitude > 0.01 μS (microvolts).

The cooking task

The virtual and augmented system was developed using Unity 5.5.1f1 software, applying c# programming language using the Visual Studio tool. Participants performed the virtual cooking task wearing an HMD device (HTC VIVE, https://www.vive.com/eu/product/) and through two hand controllers, and the augmented cooking task using Microsoft HoloLens (https://www.microsoft.com/it-it/hololens). The AR experience was performed in a real kitchen in which the augmented synthetic objects appeared in front of the subject according to the subjective human gaze. The interaction in AR was ensured by various sensors integrated into the headband, like cameras that, through the human gaze to target, allow the real hands' interaction with the synthetic elements.

Before the VR and AR virtual cooking task, participants performed two introductory tasks (tutorial), one for each technology, in order to learn the main body movements and hands' interactions useful to perform the virtual cooking task. The tutorial consisted of a simulated task, similar to the virtual cooking task. In both conditions, body movements were real in the physical space and hands' interaction in the VR was performed through the use of two controllers and in AR, participant interacted with objects with their own hands. Participants could train for as long as necessary, according to the needs of each one. When they felt confident about body and hand movements and interactions, a button pulsed to start the virtual cooking task.

The virtual cooking task consisted of 4 levels of difficulty, involving the abilities to pay attention, planning, and shifting. All were based on cooking a series of food in a set time, avoiding burning (in which the ingredient was in the fire more than the set time) or cooling them (switch off the glass-ceramic switch or remove the food from the pan during cooking). As the level increased, additional activities appeared (Figure 26). In the first level, participants had to cook three foods in one cooker on 2 minutes; in the second level, they had to cook 5 foods on 2 cookers in 3 minutes; in the third level, a dual-task should be performed: a) 5 foods should be cooked on 2 cookers in 4 minutes; b) during the cooking, participants should add the right dressing to the foods; in the last level, another dual-task has been proposed: a) participants should cook 5 foods in 2 cookers in 5 minutes; and b) they should set a table. Each food should be cooked in a scheduled time, as well as the level had a limit time that appeared all the time in the virtual and augmented environment. When the food was cooked, it had to be removed from the pan, turning off the cooker and placed in the dish. The main aim of each level was to cook the foods in the scheduled time without burning and letting them cool. Burning a food means by not taking it out of the pan, or turning the burner off, after the predefined cooking time. Cooling a food means left the food in the pan to cool down after it was cooked. The virtual system gathered various time/performance data for each subtask, including total times, burning times, and cooling times.

Participants exceeded the following level when they have cooked all the foods, completing the level. Before each level, instructions, explaining what activities participants had to be carried out, what time they had to do it, times for each food and remembering to cook foods without burning and letting them cool, have been showed (Figure 27).

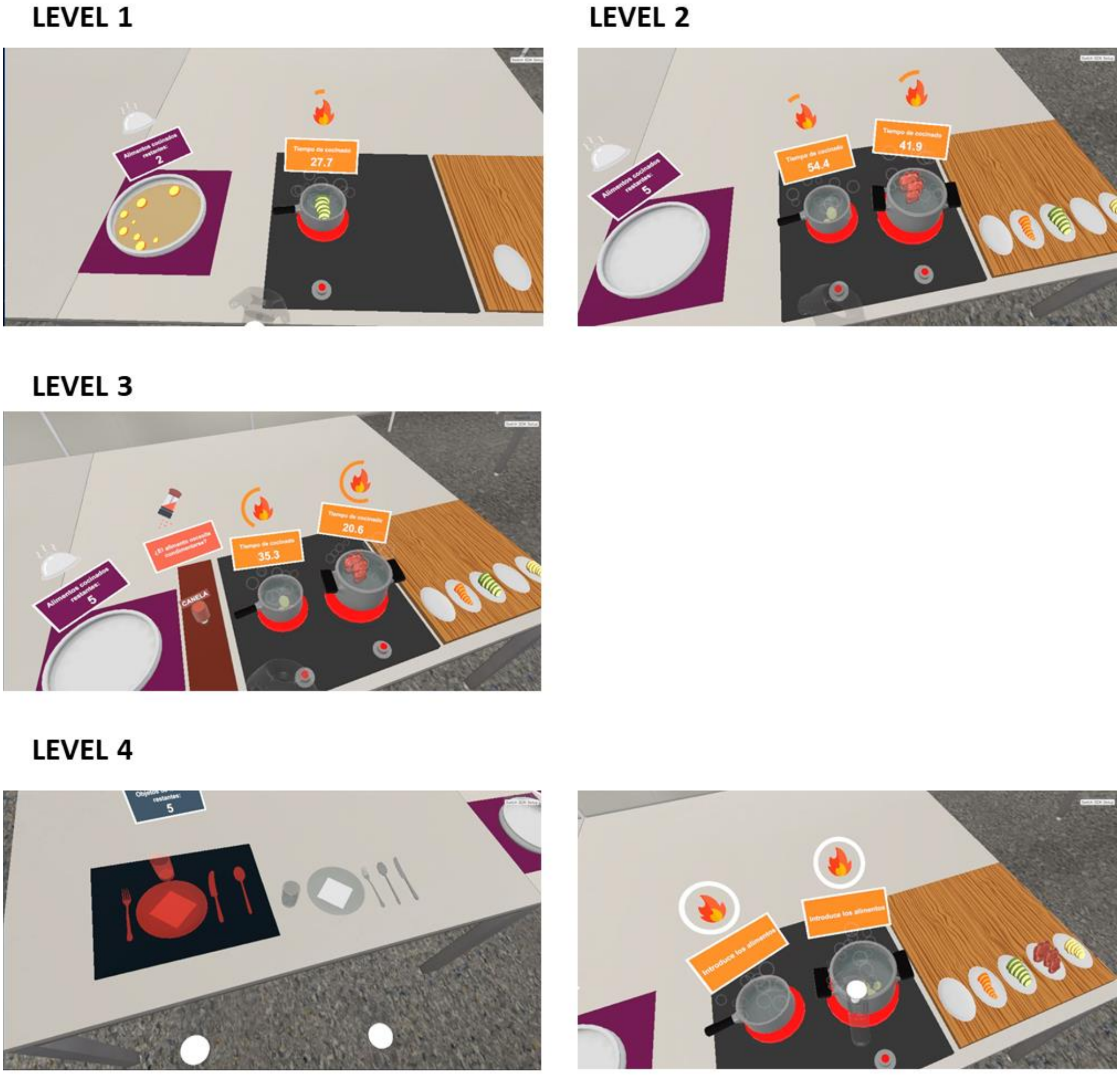


Figure 28. The cooking task levels

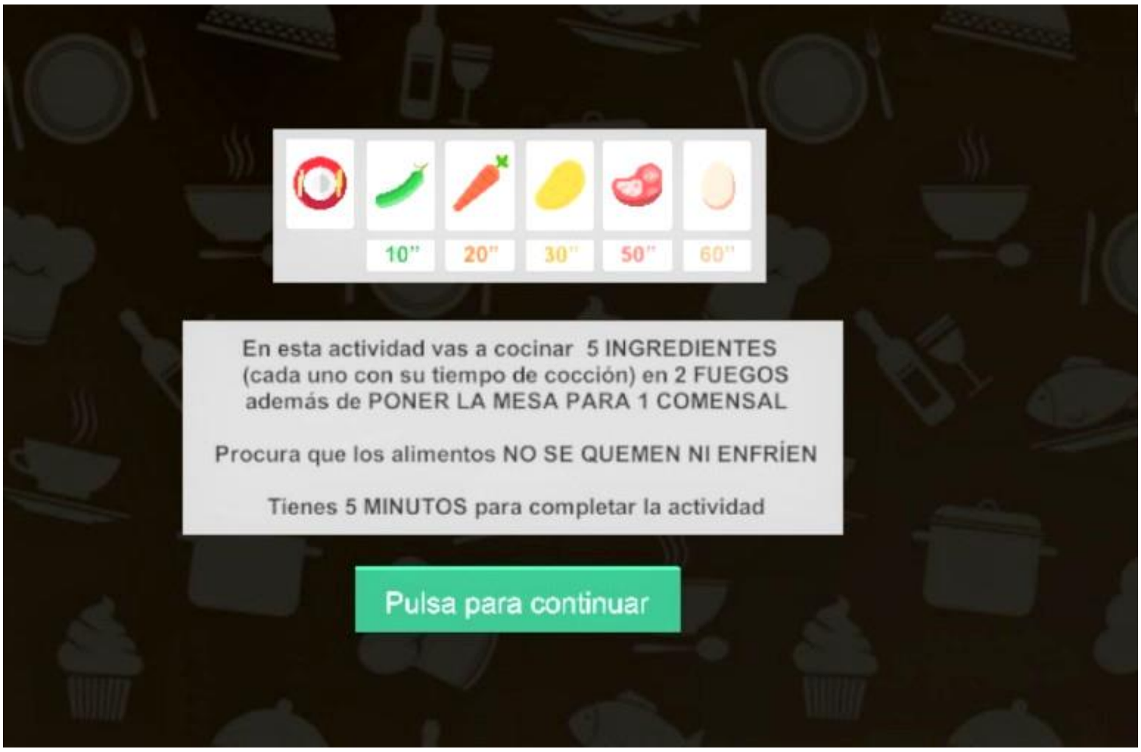


Figure 29. Example of level's instruction.

Experimental procedure

The order of presentation of each exposure condition (AR and VR) was counterbalanced for each participant. Before the beginning of the experiment, participants were administered the MMSE and standard questionnaires (ACS, BIS, CFS) and tasks (DOT, GoNoGo, Stroop, TMTA-B, TOL). Once this first phase was completed, we recorded 3-minutes of EDA and HRV baseline, asking to participants to stay completely relaxed during the recording. Once the physiological baseline was recorded, the experimental session started, and EDA and HRV were continuously recorded until the end of the experiment. To measure the sense of presence occurring during the two exposure conditions, subjects completed the SUSQ and the ITC-SOPI immediately after each condition.

Statistical analyses

The analyses were performed using SPSS version 22.0 (Statistical Package for the Social Sciences for Windows, Chicago, IL) for PC. The biosignals' processing and computation were analyzed using Matlab and Ledalab programs. First, we verified the assumptions of normality applying

Kolmogorov Smirnov and the internal consistency of the scales was assessed via Cronbach's alpha.

Second, it has been verified the normal cognitive functioning and the physiological health (SDNN and rMSSD of HRV values) of the subjects.

Next, four paired t-tests were conducted to compare behavioral, physiological data (SCR and HR), and sense of presence responses in AR and VR conditions. The level of significance was set at $\alpha = 0.05$.

Results

The assumption of normality was confirmed (Kolmogorov Smirnov $p > .05$) and the internal consistency of the self-report scales has been measured (Cronbach's alpha αattention = .819, αcognitive flexibility = .765, αimpulsiveness = .785; αSUSQ_AR =.907; αSUSQ_VR =.702; αITC-SOPI_AR = .946; αITC-SOPI_VR= .937; bootstrap 95%).

Regarding the cognitive functioning (Table 25), the mean total score on cognitive flexibility showed that the subjects had a high cognitive flexibility (CFS TOTAL = 47.36; normal range: 10-60); the mean total value on impulsivity (BIS TOTAL = 65.21) is within the normal limits of impulsivity (normal range: 52-71); and for attentional control, a very high mean score was obtained (ACS TOTAL = 55.44), indicating that subjects were able to voluntarily control their attention. Table 25 also reports the descriptive data on standardized tasks.

Focusing on health at physiological level, the values of beats per minute (BPM) at baseline and during the tasks are in the normal range of 60-100 beats/minute. Also, SDNN values indicate that participants are not in danger of suffering from any cardiac episode since the data is greater than 100 ms, while the rMSSD are also in the normal range (greater than 25 ms) (Macías, 2016; Gámiz et al., 2018) (Table 26).

Table 26

Mean and standard deviation of Heart Rate (HR) SDNN and rMSDD values for condition.

AR			VR	
	M	SD	M	SD
HR_SDNN_Baseline (ms)	118,09	50,48	126,16	44,18
HR_SDNN_Postline (ms)	110,11	48,11	123,44	41,89
HR_rMSSD_ Baseline (ms)	146,71	75,69	164,81	60,75
HR_rMSSD_ Postline (ms)	137,90	76,84	162,02	64,57
HR_SDNN_Level 1 (ms)	180,06	76,17	216,94	57,48
HR_rMSSD_Level 1(ms)	239,58	110,38	295,25	84,80
HR_SDNN_ Level 2(ms)	185,30	85,79	233,99	62,46
HR_rMSSD_Level 2(ms)	243,30	114,33	317,66	94,45
HR_SDNN_ Level 3 (ms)	179,48	81,48	234,20	53,50
HR_rMSSD_ Level 3 (ms)	242,90	107,34	315,49	75,58
HR_SDNN_ Level 4 (ms)	113,27	45,87	253,63	52,69
HR_rMSSD_ Level 4 (ms)	141,99	74,77	341,28	76,68
HR_SDNN_Level 4 (ms)	113,27	45,87	253,63	52,69
HR_rMSSD_Level 4 (ms)	141,99	74,77	341,28	76,68

Behavioral responses to cooking task

Regarding performance, Table 27 shows the mean and standard deviations of behavioral values of the cooking task for both conditions.

A paired t-test was conducted to compare behavioral responses in AR and VR conditions. There were significant differences in the scores for the total 4 levels' time between AR (M= 776.07, SD= 176.89) and VR (M= 574.13, SD= 76.22) conditions; t(49)= 7.75, p=0.00, as well in the total time of level 1 in AR (M= 177.58, SD= 60.47) and VR (M=129.78, SD= 16.45) conditions; t(49)=

3.08, p=0.00 and in the level 2 in AR (M=160.60, SD=60.33) and VR (M=132.09, SD=22.30) conditions; t(49)= -3.08, p=0.00 (Figure 28). Regarding cooling times significant differences between conditions have been found at level 1 [AR (M= 7,57, SD=16.36), VR (M= 0.33, SD=1.97); t(49)= 3.08 p=0.00], level 2 [AR (M=2.23, SD=5.46), VR (M=0.06, SD=0.21); t(49)= -2.81, p=0.01] and level 3 [AR (M=0.66, SD=1.37), VR (M=0.01, SD=0.05); t(49)= 3.37, p=0.00]. Finally, significant differences on burning times have been found between conditions at level 2 [AR (M=1.74, SD=2.02), VR (M=1.02, SD=0.47); t(49)= 2.68, p=0.01], level 3 [AR (M=2.00, SD= 1.50), VR (M=1.22, SD=0.74); t(49)= -3.55, p=0.00], and level 4 [AR (M=1.48, SD=1.36), VR (M=1.06, SD=0.91); t(49)= 2.12, p=0.04] (Figure 29).

Table 27

Mean and standard deviation of values for behavioral responses in AR and VR conditions.

	AR		**VR**	
	M	SD	M	SD
Total time 4 Levels (s)	776,07	176,89	574,13	76,22
Total time Level 1 (s)	177,58	60,47	129,78	16,45
Level 1 Burning Time (s)	1,28	0,98	1,13	0,81
Level 1 Cooling Time (s)	7,57	16,36	0,33	1,97
Total Time Level 2 (s)	160,60	60,33	132,09	22,30
Level 2 Burning Time (s)	1,74	2,02	1,02	0,47
Level 2 Cooling Time (s)	2,23	5,46	0,06	0,21
Total Time Level 3 (s)	165,47	50,62	159,22	31,75
Level 3 Burning Time (s)	2,00	1,50	1,22	0,74
Level 3 Cooling Time (s)	0,66	1,37	0,01	0,05
Total Time Level 4 (s)	158,26	48,85	154,68	55,62
Level 4 Burning Time (s)	1,48	1,36	1,06	0,91
Level 4 Cooling Time (s)	1,60	7,14	0,13	0,70
Total Time Levels' Mean(s)	165,48	39,98	143,94	21,92

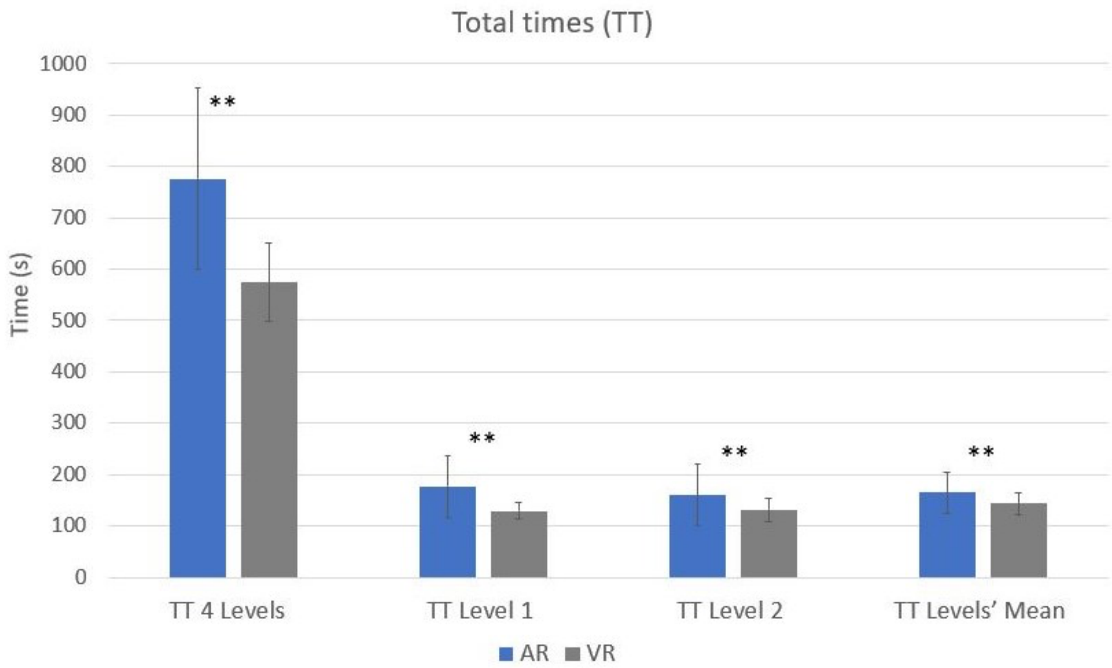


Figure 30. Paired t-test significant differences between conditions for total times (* p ≤ 0.05, **p ≤ 0.01).

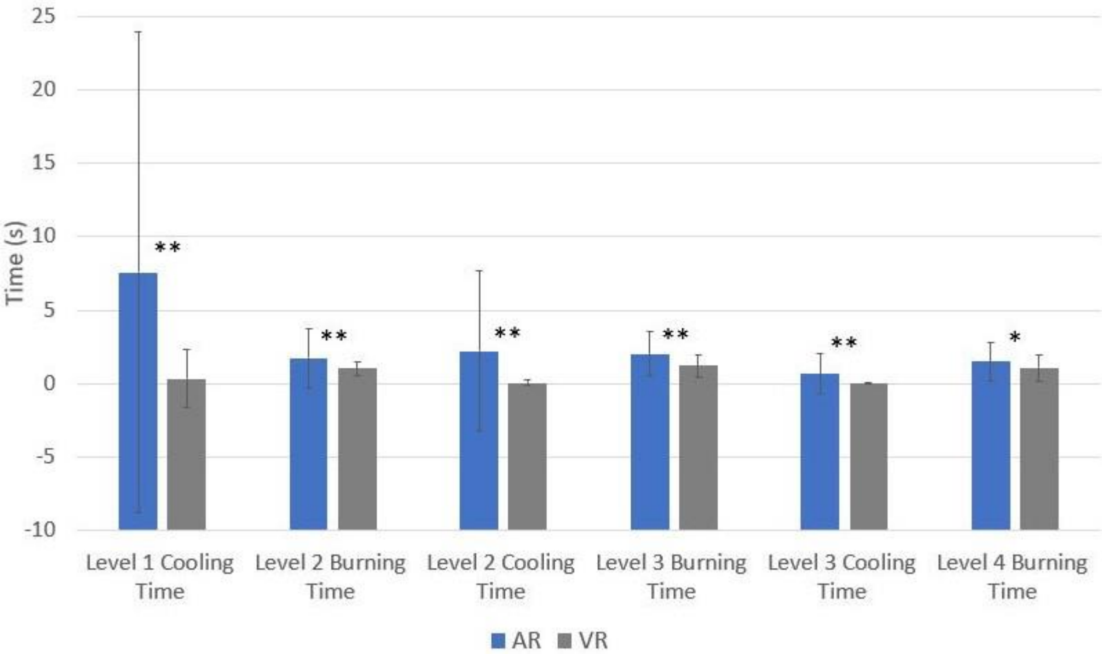


Figure 31. Paired t-test significant differences between conditions for cooling and burning time (* p ≤ 0.05, ** p ≤ 0.01).

Physiological responses to the cooking task

Electrodermal activity (EDA)

First, Table 28 shows the mean and standard deviation of EDA values of the cooking task for both conditions. Second, a paired t-test was computed to compare physiological responses in AR and VR conditions. There was a significant difference in the scores for EDA for AR –pre (M=3.51, SD=5.42) and –post-task (M=6.87, SD=7.81) conditions; t(49)=-5.16, p=0.00, as well for VR –pre (M=2.60, SD=4.25) and –post tasks (M=5.20, SD=5.95) conditions; t(49)=-4.22. Another significant difference in the scores for EDA for AR –post (M=6.87, SD=7.81) and VR –post-task (M=5.20, SD= 5.95) conditions; t(49)=-2.95, p=0.00 has been found. Finally, there was a significant difference in the scores for number of peaks in the first level task between AR (M=59.08, SD=51.30) and VR (M=48.52, SD=30.07) conditions; t(49)=2.01, p=0.05 (Figure 30) No other significant differences in physiological activation during the 4 levels of the cooking task have been found.

Table 28

Mean and standard deviation of EDA values in AR and VR conditions.

AR			VR	
	M	SD	M	SD
EDA_Baseline (µS)	3,51	5,42	2,60	4,25
EDA_Postline (µS)	6,87	7,81	5,20	5,95
EDA_TOT (µS)	1,94	3,17	2,12	3,58
EDA_SCR_TOT (µS)	0,11	0,10	0,12	0,15
EDA_SCL_TOT (µS)	1,85	3,09	2,00	3,46
EDA_N_PEAK_TOT	221,02	178,73	208,96	133,77
Task1_EDA (µS)	1,39	2,34	1,47	2,54
Level 1_SCR (µS)	0,09	0,09	0,10	0,12
Level 1_SCL (µS)	1,30	2,28	1,37	2,46
N_PEAK_Level 1	59,08	51,30	48,52	30,07

Level 2_EDA (μS)	1,70	2,71	1,73	3,03
Level 2_SCR (μS)	0,09	0,09	0,12	0,16
Level 2_SCL (μS)	1,61	2,65	1,61	2,90
N_PEAK_ Level 2	51,82	43,54	50,10	31,99
Level 3_EDA (μS)	2,07	3,53	2,08	3,51
Level 3_SCR (μS)	0,09	0,11	0,13	0,18
Level 3_SCL (μS)	1,97	3,46	1,95	3,36
N_PEAK_ Level 3	55,46	45,27	57,90	42,43
Level 4_EDA (μS)	2,52	3,89	3,00	6,55
Level 4_SCR (μS)	0,11	0,10	0,16	0,20
Level 4_SCL (μS)	2,42	3,81	2,83	6,40
N_PEAK_Level 4	54,66	46,13	52,44	33,26

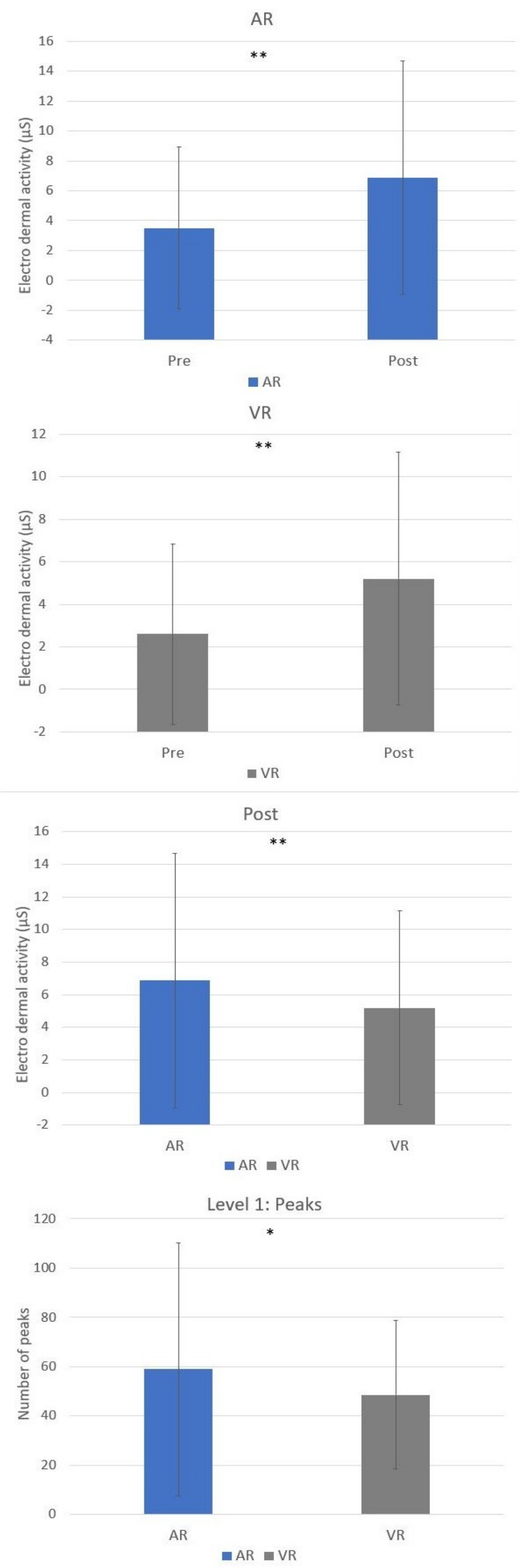


Figure 32. Paired t-test significant differences for EDA between AR and VR conditions (* $p \leq 0.05$, ** $p \leq 0.01$).

Heart Rate Variability (HRV)

Table 29 shows the mean and standard deviation of HRV values of the cooking task for both conditions (AR vs. VR).

A paired t-test was computed to compare HRV in AR and VR conditions. There was a significant difference in the scores for HRV for AR –pre (M=79.57, SD=13.43) and –post (M=81.20, SD=14.07) task; t(49)=-1.97, p=0.05 (Figure 31). No other significant differences in HRV during the 4 levels of the cooking task have been found.

Table 29

Mean and standard deviation for HRV values for AR and VR conditions.

AR			VR	
	M	SD	M	SD
HR_Baseline (bpm)	79,57	13,43	81,93	8,09
HR_Postline (bpm)	81,20	14,07	81,41	6,22
HR_BeatPerMinute_Levels(bpm)	82,26	14,21	81,97	6,15
HR_BeatPerMinute_Level1(bpm)	81,57	13,77	82,90	6,94
HR_BeatPerMinute_Level2(bpm)	81,65	13,81	82,92	6,78
HR_BeatPerMinute_Level3(bpm)	81,83	13,87	82,89	6,66
HR_BeatPerMinute_Level4(bpm)	82,70	7,88	82,75	6,67

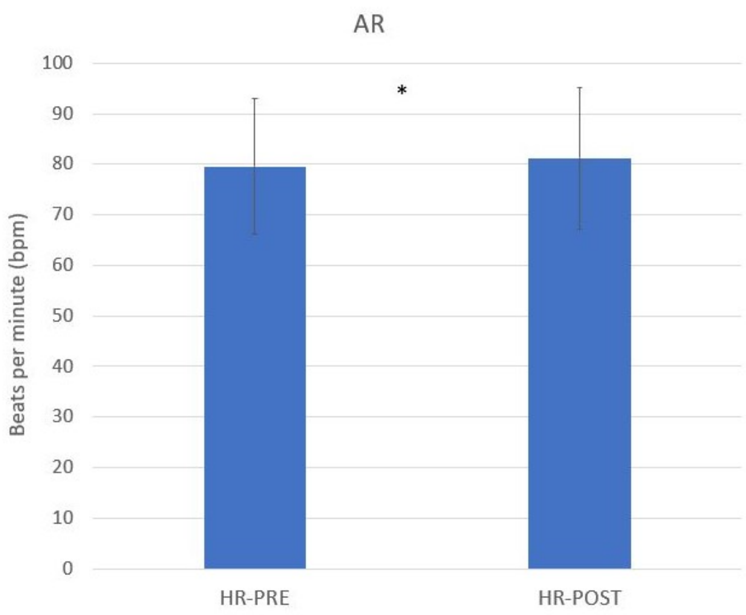


Figure 33. Paired t-test significant difference for HRV for AR -pre and -post task (* $p \leq 0.05$, ** $p \leq 0.01$).

Sense of presence

A paired t-test was computed to compare SUSQ and ITC-SOPI questionnaires in AR and VR conditions. Regarding the SUSQ, there was a significant difference in the scores for AR (M=4.11, SD=1.65) and VR (M=5.85, SD=1.00) conditions; p=0.00. The ITC-SOPI showed significant differences between AR and VR in the four dimension of presence: spatial presence (SP) [AR (M=3.29, SD=0.64); VR (M=3.81, SD=0.57) p=0.00]; engagement (E) [AR (M=3.6, SD=0.69); VR (M=4.21, SD=0.49) p=0.00]; ecological validity (EV) [AR (M=3.21, SD=0.86); VR (M=3.93, SD=0.73)]; and negative effects (NE) [AR (M=1.7, SD=0.64); VR (M=1.52, SD=0.51)] (Figure 32).

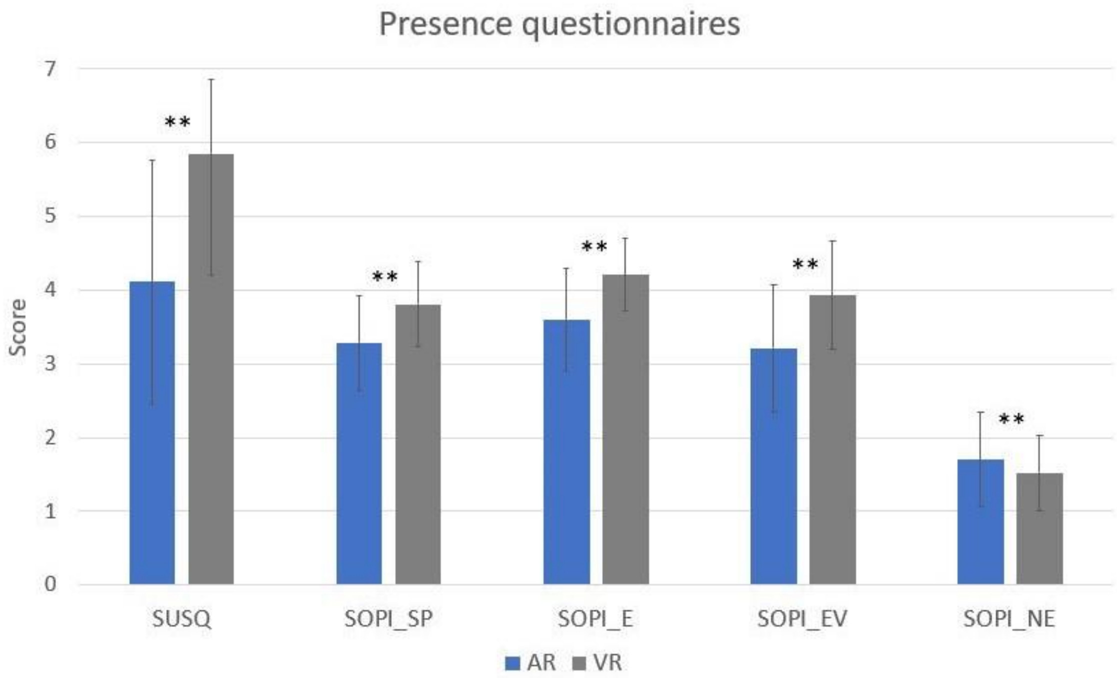


Figure 34. Paired t-test significant difference for presence questionnaires for AR and VR conditions (* $p \leq 0.05$, ** $p \leq 0.01$).

Discussion and conclusions

The first aim of this study was to analyze and compare behavioral and physiological data collected before, during and after performing the cooking task in virtual reality (VR) and augmented reality (AR) environments. The results on the behavioral data comparison showed that times are always lower in VR than AR, both for the times' means of the 4 levels and in the specific levels. This may be because the interaction with VR is usually simpler; however, the VR and AR descriptive data showed that AR levels times decrease, while in VR levels increase (Bermejo Vidal, 2018). The results are partially in opposite with previous works that have compared VR and AR and could depend on the task to perform and on display fidelity, the accuracy and complexity of the technological systems (Arino et al., 2014; Bowman et al., 2012; Botden et al., 2007; Boud et al., 1999; Cidota et al., 2016; Jones et al., 2008; Juan & Pérez, 2010; Khademi et al., 2013; Krichenbauer et al., 2017; Irawati et al., 2008; Lee et al., 2013; Möller et al., 2014). Indeed, the main previous comparison studies implemented non-complex tasks in which one task was proposed at a time, such as object manipulation, and our study included activity of daily life characterized by a succession of actions and/or two tasks at a time in a rich similar real environment. For example, Khademi et al. (2013) compared an AR with VR "pick and place" task for stroke patients, showing that participants performed better in the AR condition than in VR. Möller et al. (2014) performed a study on navigation with a guidance system, showing that participants navigated in VR faster than in AR but committing more errors. Furthermore, each previous study used a different technological system with specific characteristics according to display fidelity, the accuracy, and complexity of the technological systems that could influence the results' variability (Germine et al., 2019).

Coherent with the behavioral data are the physiological results, showing that both conditions produced individual activation with higher values in AR than VR (Bermejo Vidal, 2018). Higher physiological activation in AR could depend on the interaction system differences. More in detail, as mentioned in the description of the cooking task, in VR interaction was ensured by two hand controllers, and in AR depended on the human gaze that allowed real hand interaction with the synthetic elements.

Regarding the second aim on the sense of presence, scores between conditions showed that VR always produces a higher sense of presence than AR (Bermejo Vidal, 2018). Specifically, the higher significant results on VR spatial presence dimension than AR could depend on the fact that VR condition is mostly unmediated. Indeed, VR created a unitary and composite synthetic environment in which the user is totally immersed without interferences from the real world and AR adds synthetic objects to the real world, being able to perceive of discordance between reality and the artificial information in the environment. Regarding the engagement and ecological validity dimensions, we expected that the ecological validity results of AR would be significantly higher than that of VR. Nevertheless, the higher score in VR could depend on the self-report measure used (ITC-SOPI) also for evaluating AR experience. Indeed, the ITC-SOPI items related to ecological validity (5, 11, 15, 20, 27) evaluate if the environment seems natural or if was part of the real world and in AR the environment is the real world. This suggests that in the future studies a change of the scale may be needed for evaluating ecological validity in AR. Finally, participants evaluated AR with a negative connotation with respect to RV as shown in the results and especially in SOPI-EN (Negative effects of the ITC-SOPI) where AR has a higher score than in VR. This result seems to confirm previous results on the comparison between both conditions in situations of acrophobia, in which sense of presence was higher in VR than AR (Juan & Pérez, 2010). This result could also depend on less difficulty and greater familiarity by the subjects in using VR controllers and a feeling of greater naturalness in the interaction in AR, as mentioned in the introduction (McMahan et al., 2010).

Although the results are interesting for their possible applications in neuropsychological assessment, our study has some limitations that could affect the generalizability of the results or that may have influenced the findings. The main issues are related to the small sample size and the specific sample of healthy subjects included in this study. At the technological level, FOV and user interaction differences between the two technological systems can have generated the variability of the test scores. Futures studied are needed to investigate differences in behavioral responses comparing clinical populations and healthy subjects, as well as comparing AR and VR with other condition, such as the real condition. Furthermore, to overpass possible differences between the technological systems results, in the test design it would be important to focus more on the accuracy of the responses rather than in reaction times and also implement an individual baseline on the same or another measure using the different systems before the experimental task (Germine et al., 2019). In this way, it would be possible to consider and control system variability producing a higher generalization of the results. To conclude, VR and AR are two novels GIT with a high ecological validity value applicable to a wide variety of research fields, so it is relevant to understand the effects of various technological systems also on neuropsychological effectiveness. Specifically, we focused on behavioral performance, physiological activation in the virtual cooking task and on the sense of presence, comparing VR and AR. We found higher results on VR than AR condition in all comparison factors.

This research represents a step towards better understanding the differences between AR and VR and opens up several new venues for future research works. In particular, we conclude that future test designs took into consideration some changes in the experimental design – adding an individual technological baseline and considering also the responses accuracy – and in the self-report scale to measure presence in AR.

Part III

Discussion and conclusions

General discussion of the results and conclusions

The general objective of the present thesis project was to design, using Evidence-Centred Design, develop, and compare two ATSNA systems based on cognitive functioning to advance in obtaining an ecologically valid stealth assessment system that could be an alternative method to traditional neuropsychological assessment. Indeed, as reported in the introduction section, traditional measures of cognitive functions are often based on hypothetical constructs that have showed little relevance to predict real-life behaviours (Burgess et al., 1998; 2006; Chaytor et al., 2006; Chevignard et al., 2000; Manchester et al., 2004; Parsons et al., 2017).

Previously to the experimental studies, a network and cluster analysis on the research evolution and development of advanced technological systems, such as virtual and augmented reality, has been conducted from the 1990 to the present. The literature analysis showed a composite background. Looking at the entire period, proceeding papers represented the main research outcomes but in the last five years journal papers overpassed the general production. This result is in accordance with the evolution in subject category in which, initially, computer science and engineering were the two leading categories, and more recently, the number of applications in medical, psychological, and educational areas substantially increased. These results suggest that the continuous development of advanced technologies and the increased commercialization of hardware and software systems allowed the increment of more low-cost developments in various research areas. Together with this, continuous researches have improved graphic resolution and interaction improving the perception of reality. These developments allowed creating environments and scenarios always more like the real ones and the hand contactless devices provided a more natural interaction with virtual objects. All these features, for example in psychological research, guaranteed and showed efficacy in behavioural assessment and in treatment.

Furthermore, about the countries, USA, China, England, and Germany have the most of published papers and in the last five years have emerged also Japan, Canada, Italy and Spain. Furthermore, initially the 48% of the total production concerned proceeding papers, while in the last five years, the 57% of the contributions includes journal papers and the 33% refers to proceeding papers. Finally, regarding the journal citations, the main journals are related to psychological, medical, rehabilitation research area, such as Cyberpsychology & Beahvior, Presence: Teleoperators & Virtual Environments, Advanced Health Telematics and Telemedicine, and Plos One (Cipresso et al., 2018).

In the light of the state of the art, the first experimental study, named “EXPANSE: A novel narrative serious game for the behavioural assessment of cognitive abilities”, proposed a 2D serious game for the assessment of cognitive functions as an alternative method to the traditional one. The cognitive functions included attention, control inhibition, impulsiveness, cognitive flexibility, and planning and each game has been created starting from the cognitive constructs and using ECD as model of reference. The main aim of this study was to examine plausibility and feasibility of a 2D serious game comparing behavioural performance with traditional assessment. The results on 354 subjects showed good internal consistency as regards attention, cognitive flexibility, and impulsiveness, and initial evidences for convergent validity between behavioural performance and traditional tests. Specifically, EXPANSE showed strong and moderate correlations between traditional tests and games related to attention, cognitive flexibility, and planning skills. Furthermore, the comparison results showed that participants spent more time solving tasks in the serious game and gave fewer correct answers compared to performance in traditional tests. Traditional tests, as mentioned in the introduction, are abstracts and out of reality contexts and real life activities are situated within a complex context, which includes the appearance of different stimuli at the same time, distracting stimuli, social interactions, etc.. This result is coherent with previous studies in which several authors showed that traditional tests are not able to well discriminate and identify real cognitive (dys-) functional status (Burgess et al., 1998; 2006; Chaytor et al., 2006; Chevignard et al., 2000; Manchester et al., 2004) and ATSNA, such as serious games and/or virtual reality, are showing effectiveness in the assessment of cognitive functions (Fleming et al., 2017; Kane & Parsons, 2019; Negut et al., 2014; Valladares-Rodríguez et al., 2016).

The second experimental study applied the same methodological paradigm of the first study, adding the immersive three-dimensional condition and comparing the performance between a 2D-SG and a 3D-SG on 94 healthy subjects. As in the first study, the results exhibited good psychometrics properties related to internal consistency of the measures and convergent validity between traditional measures and the SG performance. On convergent validity, an interesting result showed that 3D condition identified higher correlation than 2D condition, and the comparison between the two systems revealed that 3D was able to generate lower reaction times, higher correct answers, and lower perseverative responses in attention abilities, inhibition control, and cognitive shifting than 2D condition. According to a recent critical overview of the state-of-the-art of Germine et al. (2018), ATSNA systems, on one hand, can provide more precise stimuli control and quantification of reaction times and accuracy in responses than traditional assessment; on the other, device characteristics, such as visual displays, or screen sizes, or the execution of visual display elements, as well as the interaction devices (e.g. mouse, keyboard, or controllers) and people's familiarity with those devices, can influence performance and the interpretation of the data results, generating systematic measurement bias in validation and collection of normative data. For example, a comparison analysis on 8304 participants between iPad, personal computer, and iPhone version of the trail-making test for the assessment of attention and cognitive flexibility showed that the test took 24-30% less time to perform on iPad than personal computer and 30-31% more time on iPhone, suggesting that the screen size con impact the performance (Germine et al., 2018).

Finally, the normative data on traditional neuropsychological tests are adjusted taking into account age, gender, and educational levels. The congress paper of this thesis project concerned the influence of gender, age, and education on cognitive functions comparing traditional tests and the serious game. The results on 268 healthy subjects were consistent with the scientific literature.

Specifically, regarding age variable, younger (18-40 years) showed lower reaction times and higher correct answers than older (41-64), and according to the literature the highest performance is between 20 and 29 years old, diminishing progressively in later adulthood (Bravo & Hébert, 1997; Craik & Salthouse, 2011; De Luca et al., 2003; Moraes et al., 2010; Rossetti et al., 2011). Regarding gender and attention abilities, women showed higher performance times but a better performance than men. Contrariwise, on cognitive flexibility and planning abilities women performed the SG in more time than men and men also showed a better performance than women. These results are consistent with the complex panorama on literature about gender (Bagherpoor & Akbar, 2014; Morgado et al., 2019; Van Hooren et al., 2007; Zarghi & Zarindast, 2011). Indeed, some studies revealed better performance of women on verbal tasks and men in visual-spatial tasks, and other studies showed weak or no differences gender-related. Similar results on education reflected consistence with previous works, showing that individuals with higher education performed better than individuals with lower education, mostly in higher-order cognitive functions, such as planning and cognitive flexibility (Ardila et al., 2010; Bravo & Hébert, 1997; Moraes et al., 2010; Ostrosky-Solis et al., 1998; Rossetti et al., 2011).

The third experimental study mainly investigated the behavioural differences on 50 healthy subjects during a cooking task, comparing the performance in VR and AR. Secondly, the study compared the individual sense of presence in both systems. Regarding the main aim, the results showed that the total times and the level times are lower in VR than AR. These results are partially on opposite with previous research and could depend on the differences in system used and the interaction tools, as well as on the complexity of the task (Arino et al., 2014; Bowman et al., 2012; Botden et al., 2007; Boud et al., 1999; Cidota et al., 2016; Jones et al., 2008; Juan & Pérez, 2010; Khademi et al., 2013; Krichenbauer et al., 2017; Irawati et al., 2008; Lee et al., 2013; Möller et al., 2014). Indeed, previous studied have compared the two systems using a simple task, such as object manipulation or a pick and place task (Khademi et al., 2013), while our task based on a dailylife activity, characterized by a succession of actions and with more than one action at the same time. Regarding the second aim, VR produced a higher sense of presence than AR (Bermejo

Vidal, 2018). Specifically, on spatial presence dimension, the results could due to fact that VR is an unmediated experience. As mentioned in the introduction section, VR is able to generate a unitary and rich environment in which the user is totally immersed while in AR the synthetic object are added to real-world, generating probably a mediated experience. Furthermore, the higher VR scores on ecological validity dimension than AR could depend on the self-report measure used – the ITC-SOPI – that evaluates if the environment could seem natural or part of the real world. This result suggested using, in future works, and another self-report measure to assess presence in AR. Finally, regarding the negative effects dimesion, AR was evaluated to have more negative effects than VR. This result could depend on greater familiarity by the subjects in using VR controllers and an expactation of greater naturalness in the AR interaction, as mentioned in the introduction (McMahan et al., 2010).

So far as these studies have shown positive results, supporting the use of ASTNA in neuropsychological evaluation, they have limitations. First, healthy subjects composed the sample examined in the studies, limiting the sensitivity of the results, and as a consequence, the external validity to the generalization of the results. Second, the studies examined convergent validity that, as mentioned in the introduction, is a subtype of construct validity. However, construct validity in composed both by convergent and discriminant validity and the theoretical approach stresses the importance of using both validation techniques when assessing new tests (Campbell & Fiske, 1959). The presented studies showed initial evidence on convergence but not on discrimination validity. Third, the use of different devices, such as desktop or HMD and different interaction devices, as mouse or controllers, can influence the test performance, limiting the validity of the results. Indeed, each hardware and software presents latency characteristics to record individual responses that can vary from one device to another. So, the total measured response time are representing by the individual response plus the time to record the response by the specific device. In neuropsychological assessment using ATSNA, this latency time differences across devices could: a) detect individual impairments when he/she functioning is normal and vice-versa; b) impact the results on group differences in cognitive performance; and c) alter the temporal stability of the results if an individual is examined at two time points using different devices.

Future research

In accordance with the limitations, further studies are required to examine sensitivity of the ATSNA, including also clinical populations, as well as its reliability and predictive and content validity according to the different criterions and the distinctive components of cognitive functions. Furthermore, to improve the classification between healthy and impaired population, as well as to predict that ASTNA situations and activities are accurately related to the cognitive constructs, further analyses using machine learning methods can enhance the accuracy of the performance outcomes and also their sensitity and specificity among other measures. Finally, further studies will be conducted, developing more ecological contexts, similar to the real ones in order to make more accurate predictions about a person's ability to work, to attend to school, and other activities.

Conclusions

The most advantage of ATSNA tools for cognitive assessment, and specifically of this thesis, is to increase the accessibility of neuropsychological instruments, providing ecologically valid cognitive measures and to reduce the cost of assessment, reaching larger number of people. Through scientific rigor in methodology, ASTNA tools in this thesis showed that can facilitate: a) more reliable evaluations of cognitive functioning; b) the ability to capture variability of a person's cognitive functioning; c) better ecological validity. Specifically, regardless of the proposed activity or situation that was a real-life simulation or a fantasy game, 3D virtual reality tool and environment showed control over the other two conditions of 2D and AR. The VR control over the 2D condition is particularly interesting since the major commercialized and popular tools are 2D desktop/laptop and tablet in which the human computer interaction is characterized by keyboard, mouse, or trackpad or touchscreen. Although less widespread, VR is showing to be a potential tool also for assessment and not only for psychological treatments. A relevant actual issue in scientific community is the method to develop and implement VR environments to evaluate empirically the claims that ASTNA can be a means to assess constructs and behaviros with great validity than traditional approaches. In other words, which situations/activities and what characteristics of these contribute to enhance validity in the assessment of cognitive functions? Stealth assessment based on an evidence-centre designing has showed to be a valid method to develop VR environments for psychological assessment. Finally, this thesis took into consideration individual's variables, such as age and education, that are relevant in the traditional assessment and that will be important in the development of ASTNA applications for cognitive functioning assessment and to generate normalize data.

To conclude, I believe that the future of neuropsychology will lie in ASTNA tools, although the change will be critical and long over the time. To reach this aim, scientific community should work in the same direction to share open source platforms to obtain larger sample data that will allow to create robust results and more consistent and standardised validations.

Scientific contributions

The publications derived from this thesis are the following:

<u>*Papers in journals indexed in JCR*</u>:

- **Chicchi Giglioli, I. A.**, Pravettoni, G., Sutil Martín, D. L., Parra, E., & Raya, M. A. (2017). A novel integrating virtual reality approach for the assessment of the attachment behavioral system. *Frontiers in psychology*, *8*, 959.
- Cipresso, P., **Chicchi Giglioli, I. A.**, Raya, M. A., & Riva, G. (2018). The past, present, and future of virtual and augmented reality research: a network and cluster analysis of the literature. *Frontiers in psychology*, *9*.
- **Chicchi Giglioli, I. A.**, 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.
- Olmos-Raya, E., Ferreira-Cavalcanti, J., Contero, M., Castellanos-Baena, M. C., **Chicchi-Giglioli, I. A.**, & Alcañiz, M. (2018). Mobile virtual reality as an educational platform: A pilot study on the impact of immersion and positive emotion induction in the learning process. *Eurasia Journal of Mathematics Science and Technology Education*, *14*(6), 2045-2057.
- **Chicchi Giglioli, I. A.**, de Juan Ripoll, C., Parra, E., & Alcañiz Raya, M. (2019). Are 3D virtual environments better than 2D interfaces in serious games performance? An explorative study for the assessment of executive functions. Applied Neuropsychology: Adult, 1-10.
- Raya, A., Luis, M., **Chicchi Giglioli, I. A.,** & Parra Vargas, E. (2018). Virtual reality as an emerging methodology for leadership assessment and training. *Frontiers in psychology*, *9*, 1658.

- **Chicchi Giglioli, I. A.,** Bermejo Vidal, C., & Alcañiz Raya, M. (2019). A virtual versus an augmented reality cooking task based-tools: a behavioral and physiological study on the assessment of executive functions. *Frontiers in Psychology*, *10*, 2529.

Papers in conferences:

- **Chicchi Giglioli, I. A.,** Parra, E., Cardenas-Lopez, G., Riva, G., & Raya, M. A. (2017, November). Virtual Stealth Assessment: A New Methodological Approach for Assessing Psychological Needs. In *Joint International Conference on Serious Games* (pp. 1-11). Springer, Cham.
- Parra, E., de Juan Ripoll, C., Raya, M. A., & **Chicchi Giglioli, I. A.** (2018, November). Individuals' Variables in Cognitive Abilities Using a Narrative Serious Game. In *Joint International Conference on Serious Games* (pp. 109-119). Springer, Cham.
- **Chicchi Giglioli, I.A.**, de Juan Ripoll, C., Llorens, R., Alcañiz Raya, M.L., (2018). Feasibility of serious games for assessing attentional abilities in people with different educational level. *10th International Conference on Education and New Learning Technologies*. DOI: 10.21125/edulearn.2018.2194.
- Llorens, R., Navarro, M.D., **Chicchi Giglioli, I.A.,** Alcañiz, M. (2018). Identification of information needs and use of ICT tools by caregivers of personas with brain injury. Preliminary efficacy of an educational app. *10th International Conference on Education and New Learning Technologies.* 10.21125/edulearn.2018.2288.

Reference

Akçayır, M., & Akçayır, G. (2017). Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educational Research Review, 20*, 1-11.

Alderman, N., Burgess, P. W., Knight, C., & Henman, C. (2003). Ecological validity of a simplified version of the multiple errands shopping test. *Journal of the International Neuropsychological Society*, *9*(1), 31-44.

Alexander, T., Westhoven, M., & Conradi, J. (2017). Virtual Environments for Competency-Oriented Education and Training. In *Advances in Human Factors, Business Management, Training and Education* (pp. 23-29). Springer International Publishing.

Andersen, S. M., & Thorpe, J. S. (2009). An IF–THEN theory of personality: Significant others and the relational self. *Journal of Research in Personality, 43*(2), 163-170.

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child neuropsychology*, *8*(2), 71-82.

Andrade, L. H., Alonso, J., Mneimneh, Z., Wells, J. E., Al-Hamzawi, A., Borges, G., ... & Florescu, S. (2014). Barriers to mental health treatment: results from the WHO World Mental Health surveys. *Psychological medicine*, *44*(6), 1303-1317.

Ardila, A., Bertolucci, P. H., Braga, L. W., Castro-Caldas, A., Judd, T., Kosmidis, M. H., ... & Rosselli, M. (2010). Illiteracy: the neuropsychology of cognition without reading. *Archives of clinical neuropsychology, 25*(8), 689-712.

Aykin, N. M., & Aykin, T. (1991). Individual differences in human-computer interaction. *Computers & industrial engineering*, *20*(3), 373-379.

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE computer graphics and applications*, *21*(6), 34-47.

Bailenson, J. N., Yee, N., Merget, D., & Schroeder, R. (2006). The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *Presence: Teleoperators and Virtual Environments*, *15*(4), 359-372.

Bacca, J., Baldiris, S., Fabregat, R., & Graf, S. (2014). Augmented reality trends in education: a systematic review of research and applications. Journal of Educational Technology & Society, 17(4), 133.Baños, R. M., Botella, C., Garcia-Palacios, A., Villa, H., Perpiñá, C., & Alcaniz, M. (2000). Presence and reality judgment in virtual environments: a unitary construct?. *CyberPsychology & Behavior*, *3*(3), 327-335.

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

Bagherpoor, L. S. K., & Akbar, A. (2014). The influence of educational type (virtual and non-virtual) and gender on the executive functions (mental flexibility and selective attention) of students. *Bull. Environ. Pharmacol. Life Sci*, 360-367.

Baler, R. D., & Volkow, N. D. (2006). Drug addiction: the neurobiology of disrupted self-control. *Trends in molecular medicine*, *12*(12), 559-566.

Baños, R. M., Quero, S., Salvador, S., & Botella, C. (2005, August). Role of presence and reality judgment in virtual environments in clinical psychology. In *CYBERPSYCHOLOGY & BEHAVIOR* (Vol. 8, No. 4, pp. 303-304). 140 HUGUENOT STREET, 3RD FL, NEW ROCHELLE, NY 10801 USA: MARY ANN LIEBERT INC.

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.

Barker, L. A., Andrade, J., Morton, N., Romanowski, C. A. J., & Bowles, D. P. (2010). Investigating the 'latent'deficit hypothesis: Age at time of head injury, implicit and executive functions and behavioral insight. *Neuropsychologia*, *48*(9), 2550-2563.

Barker, L. A., Morton, N., Morrison, T. G., & McGuire, B. E. (2011). Inter-rater reliability of the Dysexecutive Questionnaire (DEX): comparative data from non-clinician respondents—all raters are not equal. *Brain injury*, *25*(10), 997-1004.

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

Batterham, P. J., Neil, A. L., Bennett, K., Griffiths, K. M., & Christensen, H. (2008). Predictors of adherence among community users of a cognitive behavior therapy website. *Patient preference and adherence*, *2*, 97.

Baus, O., & Bouchard, S. (2014). Moving from virtual reality exposure-based therapy to augmented reality exposure-based therapy: a review. *Frontiers in human neuroscience*, *8*.

Beeri, M. S., Schmeidler, J., Sano, M., Wang, J., Lally, R., Grossman, H., & Silverman, J. M. (2006). Age, gender, and education norms on the CERAD neuropsychological battery in the oldest old. *Neurology*, *67*(6), 1006-1010.

Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *The Journal of general psychology*, *39*(1), 15-22.

Bermejo Vidal, C. (2018). Estudio de las funciones ejecutivas y su validez ecológica en la vida real: Realidad Aumentada (RA) vs. Realidad Virtual (RV) utilizando medidas explícitas (cuestionarios) e implícita (comportamentales y fisiológicas). http://hdl.handle.net/10251/110020.

Besnard, J., Richard, P., Banville, F., Nolin, P., Aubin, G., Le Gall, D., ... & Allain, P. (2016). Virtual reality and neuropsychological assessment: The reliability of a virtual kitchen to assess daily-life activities in victims of traumatic brain injury. *Applied Neuropsychology: Adult*, *23*(3), 223-235.

Biocca, F. (1997). The cyborg's dilemma: progressive embodiment in virtual environments. *J. Comput. Mediat. Commun. 3*. doi: 10.1111/j.1083-6101.1997. tb00070.x

Biocca, F., Harms, C., & Gregg, J. (2001, May). The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In *4th annual international workshop on presence, Philadelphia, PA* (pp. 1-9).

Birch, S. A., & Bloom, P. (2004). Understanding children's and adults' limitations in mental state reasoning. *Trends in cognitive sciences*, *8*(6), 255-260.

Bohil, C. J., Alicea, B., & Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. Nature reviews neuroscience, 12(12), 752. Borrego, A., Latorre, J., Llorens, R., Alcañiz, M., and Noé, E. (2016). Feasibility of a walking virtual reality system for rehabilitation: objective and subjective parameters. *J. Neuroeng. Rehabil. 13*, 68. doi: 10.1186/s12984-016-0174-1.

Botella, C. M., Juan, M. C., Baños, R. M., Alcañiz, M., Guillén, V., & Rey, B. (2005). Mixing realities? An application of augmented reality for the treatment of cockroach phobia. *Cyberpsychology & Behavior*, *8*(2), 162-171.

Botella, C., Bretón-López, J., Quero, S., Baños, R. M., and García-Palacios, A. (2010). Treating cockroach phobia with augmented reality. *Behav. Ther. 41*, 401–413. doi:10.1016/j.beth.2009.07.002

Botella, C., Fernández-Álvarez, J., Guillén, V., García-Palacios, A., & Baños, R. (2017). Recent Progress in Virtual Reality Exposure Therapy for Phobias: A Systematic Review. *Current Psychiatry Reports, 19*(7), 42.

Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of mathematical sociology, 25*(2), 163-177.

Bravo, G., & Hébert, R. (1997). Age-and education-specific reference values for the Mini-Mental and Modified Mini-Mental State Examinations derived from a non-demented elderly population. *International journal of geriatric psychiatry, 12*(10), 1008-1018.

Bretón-López, J., Quero, S., Botella, C., García-Palacios, A., Baños, R. M., & Alcañiz, M. (2010). An augmented reality system validation for the treatment of cockroach phobia. *Cyberpsychology, Behavior, and Social Networking, 13*(6), 705-710.

Brown, A., & Green, T. (2016). Virtual reality: Low-cost tools and resources for the classroom. *TechTrends, 60*(5), 517-519.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological review*, *62*(3), 193.

Bu, Y., Liu, T. Y., & Huang, W. B. (2016). MACA: a modified author co-citation analysis method combined with general descriptive metadata of citations. *Scientometrics, 108*(1), 143-166.

Burke, J. W., McNeill, M. D. J., Charles, D. K., Morrow, P. J., Crosbie, J. H., & McDonough, S. M. (2009). Optimising engagement for stroke rehabilitation using serious games. *The Visual Computer*, *25*(12), 1085.

Burdea, G., Richard, P., & Coiffet, P. (1996). Multimodal virtual reality: Input-output devices, system integration, and human factors. International *Journal of Human-Computer Interaction, 8*(1), 5-24.

Burdea, G. C., & Coiffet, P. (2003). *Virtual reality technology* (Vol. 1). John Wiley & Sons.

Burgess, P. W., Alderman, N., Evans, J., Emslie, H., & Wilson, B. A. (1998). The ecological validity of tests of executive function. *Journal of the international neuropsychological society, 4*(6), 547-558.

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.

Burgess, P. W., Alderman, N., Volle, E., Benoit, R. G., & Gilbert, S. J. (2009). Mesulam's frontal lobe mystery re-examined. *Restorative neurology and neuroscience*, *27*(5), 493-506.

Burke, J. W., McNeill, M. D. J., Charles, D. K., Morrow, P. J., Crosbie, J. H., & McDonough, S. M. (2009). Optimising engagement for stroke rehabilitation using serious games. *The Visual Computer*, *25*(12), 1085.

Calvert, E. J., & Waterfall, R. C. (1982). A comparison of conventional and automated administration of Raven's Standard Progressive Matrices. *International Journal of Man-Machine Studies*, 17(3), 305-310.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin, 56*(2), 81.

Carmigniani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., & Ivkovic, M. (2011). Augmented reality technologies, systems and applications. *Multimedia Tools and Applications, 51*(1), 341-377.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment* (Vol. 17). Sage publications.

Castelvecchi, D. (2016). Low-cost headsets boost virtual reality's lab appeal. *Nature, 533*(7602).

Cathy. (2011).The History of Augmented Reality. The Optical Vision Site. Available at: http://www.theopticalvisionsite.com/history-of-eyewear/the-history-of-augmented-reality/#.UelAUmeAOyA.

Chan, R. C., Shum, D., Toulopoulou, T., & Chen, E. Y. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of clinical neuropsychology*, *23*(2), 201-216.

Chatham, R. E. (2007). Games for training. *Communications of the ACM, 50*(7), 36-43.

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 ofexecutive functioning assessment. *Archives of Clinical Neuropsychology*,*21*, 217 – 227.

Cheek, C., Fleming, T., Lucassen, M. F., Bridgman, H., Stasiak, K., Shepherd, M., & Orpin, P. (2015). Integrating health behaviour theory and design elements in serious games. *JMIR mental health*, *2*(2), e11.

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

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the Association for Information Science and Technology, 57*(3), 359-377.

Chen, C., Ibekwe-SanJuan, F., & Hou, J. (2010). The structure and dynamics of cocitation clusters: A multipleperspective cocitation analysis. *Journal of the Association for Information Science and Technology, 61*(7), 1386-1409.

Chen, Y. C., Chi, H. L., Hung, W. H., & Kang, S. C. (2011). Use of tangible and augmented reality models in engineering graphics courses. *Journal of Professional Issues in Engineering Education & Practice, 137*(4), 267-276.

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.

Chicchi Giglioli, I. A., Pallavicini, F., Pedroli, E., Serino, S., & Riva, G. (2015). Augmented reality: a brand new challenge for the assessment and treatment of psychological disorders. *Computational and mathematical methods in medicine*, 2015.

Chicchi Giglioli, I. A., 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.

Chien, C. H., Chen, C. H., & Jeng, T. S. (2010, March). An interactive augmented reality system for learning anatomy structure. In *proceedings of the international multiconference of engineers and computer scientists* (Vol. 1, pp. 17-19). Hong Kong, China: International Association of Engineers.

Cipresso, P., Albani, G., Serino, S., Pedroli, E., Pallavicini, F., Mauro, A., & Riva, G. (2014). Virtual multiple errands test (VMET): a virtual reality-based tool to detect early executive functions deficit in Parkinson's disease. *Frontiers in behavioral neuroscience, 8*, 405.

Cipresso, P., & Serino, S. (2014). *Virtual reality: Technologies, medical applications and challenges*. Nova Science Publishers, Inc..

Cipresso, P. (2015). Modeling behavior dynamics using computational psychometrics within virtual worlds. *Frontiers in psychology*, *6*.

Cipresso, P., Serino, S., & Riva, G. (2016). Psychometric assessment and behavioral experiments using a free virtual reality platform and computational science. *BMC medical informatics and decision making*, *16*(1), 37.

Cipresso, P., Chicchi Giglioli, I. A., Raya, M. A., & Riva, G. (2018). The Past, Present, and Future of Virtual and Augmented Reality Research: A network and cluster analysis of the literature. *Frontiers in psychology, 9*.

Choi, S., Jung, K., & Noh, S. D. (2015). Virtual reality applications in manufacturing industries: Past research, present findings, and future directions. *Concurrent Engineering*, *23*(1), 40-63.

Csikszentmihalyi, M. (1997). Flow and the psychology of discovery and invention. *HarperPerennial, New York*, *39*.

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.

Craik, F. I., & Salthouse, T. A. (Eds.). (2011). *The handbook of aging and cognition*. Psychology press.

Cruz-Neira, C. (1993, July). Virtual reality overview. In *SIGGRAPH* (Vol. 93, No. 23, pp. 1-1).

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

Cummings, J., Gould, H., & Zhong, K. (2012). Advances in designs for Alzheimer's disease clinical trials. *American journal of neurodegenerative disease*, *1*(3), 205.

Das, S. K., Banerjee, T. K., Mukherjee, C. S., Bose, P., Biswas, A., Hazra, A., ... & Roy, T. (2006). An urban community-based study of cognitive function among non-demented elderly population in India. *Neurol Asia*, *11*, 37-48.

Dawson, M. E., Schell, A. M., & Filion, D. L. (2007). The electrodermal system. *Handbook of psychophysiology*, *2*, 200-223.

De Buck, S., Maes, F., Ector, J., Bogaert, J., Dymarkowski, S., Heidbuchel, H., & Suetens, P. (2005). An augmented reality system for patient-specific guidance of cardiac catheter ablation procedures. *IEEE transactions on medical imaging*, *24*(11), 1512-1524.

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior research methods*, *47*(1), 1-12.

De Luca, C. R., Wood, S. J., Anderson, V., Buchanan, J. A., Proffitt, T. M., Mahony, K., & Pantelis, C. (2003). Normative data from the CANTAB. I: development of executive function over the lifespan. *Journal of clinical and experimental neuropsychology*, *25*(2), 242-254.

Denckla, M. B. (1994). Measurement of executive function.

Denckla, M. B. (1996). A theory and model of executive function: *A neuropsychological perspective*.

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

Diamond, A. (2005). Attention-deficit disorder (attention-deficit/hyperactivity disorder without hyperactivity): a neurobiologically and behaviorally distinct disorder from attention-deficit/hyperactivity disorder (with hyperactivity). *Development and psychopathology*, *17*(3), 807-825.

Diamond, A. (2013). Executive functions. *Annual review of psychology*, *64*, 135-168.

Diaz-Orueta, U., Iriarte, Y., Climent, G., & Banterla, F. (2012). AULA: An ecological virtual reality test with distractors for evaluating attention in children and adolescents. *Virtual Reality and Scientific Visualization Journal*, *5*(2), 1-20.

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.

Dickey, M. D. (2011). Murder on Grimm Isle: The impact of game narrative design in an educational game-based learning environment. *British Journal of Educational Technology*, *42*(3), 456-469.

Di Serio, Á., Ibáñez, M. B., & Kloos, C. D. (2013). Impact of an augmented reality system on students' motivation for a visual art course. *Computers & Education*, *68*, 586-596.

Dunkin, B., Adrales, G. L., Apelgren, K., & Mellinger, J. D. (2007). Surgical simulation: a current review. *Surgical endoscopy*, *21*(3), 357-366.

Durand, V. M., & Barlow, D. H. (2012). Essentials of abnormal psychology. *Cengage Learning*.

Ebert C. (2015). Looking into the Future. *A+ A, 13*: 55.

Eckerman, D. A. (1985). An approach to brief field testing for neurotoxicity. *Neurobehavioral Toxicology & Teratology*.

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.

Elwood, R. W. (2001). MicroCog: assessment of cognitive functioning. *Neuropsychology review*, *11*(2), 89-100.

Englund, C., Olofsson, A. D., & Price, L. (2017). Teaching with technology in higher education: understanding conceptual change and development in practice. *Higher Education Research & Development*, *36*(1), 73-87.

Eisinga, R., Te Grotenhuis, M., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown?. *International journal of public health*, *58*(4), 637-642.

Fajkowska, M., & Derryberry, D. (2010). Psychometric properties of Attentional Control Scale: The preliminary study on a Polish sample. *Polish Psychological Bulletin, 41*(1), 1-7.

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. *The FASEB journal*, *22*(2), 338-342.

Feiner, S., MacIntyre, B., Hollerer, T., & Webster, A. (1997, October). A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. In *Digest of Papers. First International Symposium on Wearable Computers* (pp. 74-81). IEEE.

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., De Beurs, D., Khazaal, Y., Gaggioli, A., Riva, G., Botella, C., ... & Merry, S. (2016). Maximizing the impact of e-therapy and serious gaming: time for a paradigm shift. *Frontiers in psychiatry*, *7*, 65.

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.

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.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*, *12*(3), 189-198.

Fortuny, L. I. A., & Heaton, R. K. (1996). Standard versus computerized administration of the Wisconsin Card Sorting Test. *The Clinical Neuropsychologist*, 10(4), 419-424.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 35-41.

Freeman, D., Reeve, S., Robinson, A., Ehlers, A., Clark, D., Spanlang, B., & Slater, M. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychological Medicine*, 1-8.

Franzen, M. D., & Wilhelm, K. L. (1996). Conceptual foundations of ecological validity in neuropsychological assessment.

French, C. C., & Beaumont, J. G. (1990). A clinical study of the automated assessment on intelligence by the Mill Hill vocabulary test and the standard progressive matrices test. *Journal of Clinical Psychology*, 46, 129–140.

Fuchs, H., & Bishop, G. (1992). Research directions in virtual environments.

Gladsjo, J. A., Schuman, C. C., Evans, J. D., Peavy, G. M., Miller, S. W., & Heaton, R. K. (1999). Norms for letter and category fluency: demographic corrections for age, education, and ethnicity. *Assessment*, *6*(2), 147-178.

Gallagher, A. G., Ritter, E. M., Champion, H., Higgins, G., Fried, M. P., Moses, G., ... & Satava, R. M. (2005). Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Annals of surgery*, *241*(2), 364.

Garbarino, M., Lai, M., Bender, D., Picard, R. W., & Tognetti, S. (2014, November). Empatica E3—A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. In *2014 4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)* (pp. 39-42). IEEE.

Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic bulletin & review, 19*(5), 847-857.

Germine, L., Reinecke, K., & Chaytor, N. S. (2019). Digital neuropsychology: Challenges and opportunities at the intersection of science and software. *The Clinical Neuropsychologist*, *33*(2), 271-286.

Gigante, M. A. (1993). Virtual reality: definitions, history and applications. *Virtual Reality Systems*, 3-14.

Gilberstadt, H., Lushene, R., & Buegel, B. (1976). Automated assessment of intelligence: The TAPAC test battery and computerized report writing. *Perceptual and motor skills*, *43*(2), 627-635.

Goldstein, K. (1944). The mental changes due to frontal lobe damage. *The Journal of Psychology, 17*(2), 187-208.

González-Teruel, A., González-Alcaide, G., Barrios, M., & Abad-García, M. F. (2015). Mapping recent information behavior research: an analysis of co-authorship and co-citation networks. *Scientometrics*, *103*(2), 687-705.

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.

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of personality and social psychology*, *79*(5), 701.

Green, M. C., Brock, T. C., & Kaufman, G. F. (2004). Understanding media enjoyment: The role of transportation into narrative worlds. *Communication Theory*, *14*(4), 311-327.

Gualtieri, C. T., & Johnson, L. G. (2005). Neurocognitive testing supports a broader concept of mild cognitive impairment. *American Journal of Alzheimer's Disease & Other Dementias®*, *20*(6), 359-366.

Heaton, R. K. (1981). Wisconsin card sorting test manual; revised and expanded. *Psychological Assessment Resources*, 5-57.

Heaton, R. K., & Staff, P. (1999). Wisconsin card sorting test: computer version 3 for Windows. *Odessa, FL: Research Edition Psychological Assessment Resource*.

Heeter, C. (1992). Being there: The subjective experience of presence. *Presence: Teleoperators & Virtual Environments*, *1*(2), 262-271.

Heeter, C. (2000). Interactivity in the context of designed experiences. *Journal of Interactive Advertising, 1*(1), 3-14.

Heilig, M. (1962). Sensorama Simulator.(1962). United States Patent and Trade Office, Virginia, USA, US-3,050,870.

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* (pp. 429-445). Academic Press.

Ibáñez, M. B., Di Serio, Á., Villarán, D., & Kloos, C. D. (2014). Experimenting with electromagnetism using augmented reality: Impact on flow student experience and educational effectiveness. *Computers & Education*, *71*, 1-13.

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.

Jenkins, H. (2004). Game design as narrative. *Computer*, *44*(53), 118-130.

Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, *23*(4), 1515-1529.

Jeonghun Ku, W. G., Kim, J. H., & Kim, K. U. (2004). The development of a VR system for the cognitive & behavioral assessment of schizophrenia. *Stud Health Technol Inform*, *98*, 180.

Johnson, D. M. (2005). *Introduction to and review of simulator sickness research*. ARMY RESEARCH INST FIELD UNIT FORT RUCKER AL.

Jovanovski, D., Zakzanis, K., Campbell, Z., Erb, S., & Nussbaum, D. (2012). Development of a novel, ecologically oriented virtual reality measure of executive function: The Multitasking in the City Test. *Applied Neuropsychology: Adult*, *19*(3), 171-182.

Jovanovski, D., Zakzanis, K., Ruttan, L., Campbell, Z., Erb, S., & Nussbaum, D. (2012). Ecologically valid assessment of executive dysfunction using a novel virtual reality task in patients with acquired brain injury. *Applied Neuropsychology: Adult*, *19*(3), 207-220.

Juan, M. C., Alcaniz, M., Monserrat, C., Botella, C., Baños, R. M., & Guerrero, B. (2005). Using augmented reality to treat phobias. *IEEE computer graphics and applications*, *25*(6), 31-37.

Juan, M. C., Alcañiz, M., Calatrava, J., Zaragozá, I., Baños, R., and Botella, C. (2007). "An optical see-through augmented reality system for the treatment of phobia to small animals," in Virtual

Reality, *HCII 2007 Lecture Notes in Computer Science*, Vol. 4563, ed. R. Schumaker (Springer: Berlin Heidelberg), 651–659.

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: a review of our current understanding. *Neuropsychology review*, *17*(3), 213-233.

Khademi, M., Hondori, H. M., Dodakian, L., Cramer, S., & Lopes, C. V. (2013, July). Comparing "pick and place" task in spatial augmented reality versus non-immersive virtual reality for rehabilitation setting. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 4613-4616). IEEE.

Kim, G. J. (2005). A SWOT analysis of the field of virtual reality rehabilitation and therapy. *Presence: Teleoperators and Virtual Environments*, *14*(2), 119-146.

Klavans, R., & Boyack, K. W. (2015). Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge? *Journal of the Association for Information Science and Technology, 68*(4), 984-998.

Kleinberg J. Bursty and hierarchical structure in streams. (2002) Paper presented at the *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; 2002; Edmonton, Alberta, Canada.

Kleinberg, J. (2003). Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery, 7*(4), 373-397.

Klinger, E., Chemin, I., Lebreton, S., & Marié, R. M. (2004). A virtual supermarket to assess cognitive planning. *Cyberpsychol Behav*, *7*(3), 292-293.

Knights, R. M., Richardson, D. H., & McNarry, L. R. (1973). Automated vs clinical administration of the Peabody Picture Vocabulary Test and the Coloured Progressive Matrices. *American Journal of Mental Deficiency*.

Knight, R. G., & Titov, N. (2009). Use of virtual reality tasks to assess prospective memory: applicability and evidence. *Brain impairment*, *10*(1), 3-13.

Korolov, M. (2014). The real risks of virtual reality. *Risk Management, 61*(8), 20-24.

Krichenbauer, M., Yamamoto, G., Taketom, T., Sandor, C., & Kato, H. (2017). Augmented reality versus virtual reality for 3d object manipulation. *IEEE transactions on visualization and computer graphics*, *24*(2), 1038-1048.

Krueger, M. W., Gionfriddo, T., & Hinrichsen, K. (1985, April). VIDEOPLACE—an artificial reality. In *ACM SIGCHI Bulletin* (Vol. 16, No. 4, pp. 35-40). ACM.

Ku, J., Cho, W., Kim, J. J., Peled, A., Wiederhold, B. K., Wiederhold, M. D., ... & Kim, S. I. (2003). A virtual environment for investigating schizophrenic patients' characteristics: assessment of cognitive and navigation ability. *CyberPsychology & Behavior, 6*(4), 397-404.

Lamberts, K. F., Evans, J. J., & Spikman, J. M. (2010). A real-life, ecologically valid test of executive functioning: The executive secretarial task. *Journal of Clinical and Experimental Neuropsychology*, *32*(1), 56-65.

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

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators & Virtual Environments, 10*(3), 282-297.

Levin, H. S., Eisenberg, H. M., & Benton, A. L. (Eds.). (1991). Frontal lobe function and dysfunction. *Oxford University Press*, USA.

Levy, L. M., Solomon, R., Johnson, J., Wilson, J., Lambeth, A. J., Gandy, M., ... & Liu, R. (2016). Grouches, Extraverts, and Jellyfish: Assessment Validity and Game Mechanics in a Gamified Assessment. In *DiGRA/FDG*.

Lezak, M. D. (1982). The problem of assessing executive functions. *International journal of Psychology*, *17*(1-4), 281-297.

Lezak, M. D., Howieson, D. B., Loring, D. W., & Fischer, J. S. (2004). Neuropsychological assessment. *Oxford University Press*, USA.

Li, K., Robertson, J., Ramos, J., & Gella, S. (2013). Computer-based cognitive retraining for adults with chronic acquired brain injury: a pilot study. *Occupational therapy in health care*, *27*(4), 333-344.

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.

Lin, C. H., & Hsu, P. H. (2017). Integrating Procedural Modelling Process and Immersive VR Environment for Architectural Design Education. In *MATEC Web of Conferences* (Vol. 104, p. 03007). EDP Sciences.

Liu, K. P., Kuo, M. C., Tang, K. C., Chau, A. W., Ho, I. H., Kwok, M. P., ... & Chu, L. W. (2011). Effects of age, education and gender in the Consortium to Establish a Registry for the Alzheimer's Disease (CERAD)-Neuropsychological Assessment Battery for Cantonese-speaking Chinese elders. *International Psychogeriatrics*, *23*(10), 1575-1581.

Llorens, R., Noé, E., Ferri, J., and Alcañiz, M. (2014). Virtual reality-based telerehabilitation program for balance recovery. A pilot study in hemiparetic individuals with acquired brain injury. *Brain Inj.* *28*, 169.

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.

Logie, R. H., Trawley, S., & Law, A. (2011). Multitasking: Multiple, domain-specific cognitive functions in a virtual environment. *Memory & cognition*, *39*(8), 1561-1574.

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication, 3*(2), 0-0.

Loomis, J. M., Blascovich, J. J., & Beall, A. C. (1999). Immersive virtual environment technology as a basic research tool in psychology. *Behavior Research Methods, Instruments, & Computers, 31*(4), 557-564.

Loomis, J. M., Golledge, R. G., & Klatzky, R. L. (1998). Navigation system for the blind: Auditory display modes and guidance. *Presence: Teleoperators and Virtual Environments*, *7*(2), 193-203.

Luckerson, V. (2014). Facebook Buying Oculus Virtual-Reality Company for $2 Billion. *Retrieved November, 13*, 2014.

Lui, M., & Tannock, R. (2007). Working memory and inattentive behaviour in a community sample of children. *Behavioral and Brain Functions, 3*(1), 12.

Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., & Munafò, M. R. (2016). Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. *JMIR serious games*, *4*(2), e11.

Luria, A. R.(1966). Human brain and psychological processes. *Trans. B. Haigh*. New York: Harper and Row.

Luria, A. R. (1969). Frontal lobe syndromes in man.

Luria, A. R. (1973). The working Brain. An Introduction to Neuropsychology (London, Allen).

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

Mazuryk, T., & Gervautz, M. (1996). Virtual reality-history, applications, technology and future.

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

Marchand, A., & Hennig-Thurau, T. (2013). Value creation in the video game industry: Industry economics, consumer benefits, and research opportunities. *Journal of Interactive Marketing*, *27*(3), 141-157.

Marié, R. M., Klinger, E., Chemin, I., & Josset, M. (2003). Cognitive planning assessed by virtual reality. In *VRIC 2003, Laval Virtual Conference* (pp. 119-125).

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

Martínez-Pernía, D., González-Castán, Ó., & Huepe, D. (2017). From ancient Greece to the cognitive revolution: a comprehensive view of physical rehabilitation sciences. *Physiotherapy theory and practice*, *33*(2), 89-102.

Martoni, R. M., Salgari, G., Galimberti, E., Cavallini, M. C., & O'Neill, J. (2015). Effects of gender and executive function on visuospatial working memory in adult obsessive–compulsive disorder. *European archives of psychiatry and clinical neuroscience*, *265*(8), 707-718.

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.

Mathuranath, P. S., Cherian, J. P., Mathew, R., George, A., Alexander, A., & Sarma, S. P. (2007). Mini Mental State Examination and the Addenbrooke's Cognitive Examination: Effect of education and norms for a multicultural population. *Neurology India*, *55*(2), 106.

Maurugeon, G. (2011).New D'Fusion supports iPhone4S and 3DSMax 2012. Available at: http://www.t-immersion.com/blog/2011-12-07/augmented-reality-dfusion-iphone-3dsmax.

McDaniel, R., Fiore, S. M., & Nicholson, D. (2010). Serious storytelling: Narrative considerations for serious games researchers and developers. In *Serious game design and development: Technologies for training and learning* (pp. 13-30). IGI Global.

McGeorge, P., Phillips, L. H., Crawford, J. R., Garden, S. E., Sala, S. D., Milne, A. B., ... & Callender, J. S. (2001). Using virtual environments in the assessment of executive dysfunction. *Presence: teleoperators & virtual environments*, *10*(4), 375-383.

McMahan, R. P., Alon, A. J. D., Lazem, S., Beaton, R. J., Machaj, D., Schaefer, M., ... & Bowman, D. A. (2010, March). Evaluating natural interaction techniques in video games. In *2010 IEEE Symposium on 3D User Interfaces (3DUI)* (pp. 11-14). IEEE.

McMahan, R. P., Bowman, D. A., Zielinski, D. J., & Brady, R. B. (2012). Evaluating display fidelity and interaction fidelity in a virtual reality game. *IEEE transactions on visualization and computer graphics*, *18*(4), 626-633.

McQuiggan, S. W., Rowe, J. P., Lee, S., & Lester, J. C. (2008a). Story-based learning: The impact of narrative on learning experiences and outcomes. In International Conference on Intelligent Tutoring Systems (pp. 530-539). Springer, Berlin, Heidelberg.

McQuiggan, S. W., Rowe, J. P., & Lester, J. C. (2008, April). The effects of empathetic virtual characters on presence in narrative-centered learning environments. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1511-1520). ACM.

Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological bulletin*, *114*(3), 449.

Meldrum, D., Glennon, A., Herdman, S., Murray, D., & McConn-Walsh, R. (2012). Virtual reality rehabilitation of balance: assessment of the usability of the Nintendo Wii® Fit Plus. *Disability and rehabilitation: assistive technology, 7*(3), 205-210.

Marchand, A., & Hennig-Thurau, T. (2013). Value creation in the video game industry: Industry economics, consumer benefits, and research opportunities. *Journal of Interactive Marketing*, *27*(3), 141-157.

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.

Messinis, L., Tsakona, I., Malefaki, S., & Papathanasopoulos, P. (2007). Normative data and discriminant validity of Rey's Verbal Learning Test for the Greek adult population. *Archives of Clinical Neuropsychology*, *22*(6), 739-752.

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, *77*(12), 1321-1329.

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.

Minderer, M., Harvey, C. D., Donato, F., & Moser, E. I. (2016). Neuroscience: virtual reality explored. *Nature*, *533*(7603), 324-325.

Moraes, C., Pinto, J. A., Lopes, M. A., Litvoc, J., & Bottino, C. M. (2010). Impact of sociodemographic and health variables on mini-mental state examination in a community-based sample of older people. *European archives of psychiatry and clinical neuroscience*, *260*(7), 535-542.

Morgado, J., Rocha, C. S., Maruta, C., Guerreiro, M., & Martins, I. P. (2010). Cut-off scores in MMSE: a moving target?. *European Journal of Neurology*, *17*(5), 692-695.

Morton, N., & Barker, L. (2010). The contribution of injury severity, executive and implicit functions to awareness of deficits after traumatic brain injury (TBI). *Journal of the International Neuropsychological Society*, *16*(6), 1089-1098.

Neguț, A., Matu, S. A., Sava, F. A., & David, D. (2016). Virtual reality measures in neuropsychological assessment: a meta-analytic review. *The Clinical Neuropsychologist*, *30*(2), 165-184.

Neri, S. G., Cardoso, J. R., Cruz, L., Lima, R. M., de Oliveira, R. J., Iversen, M. D., & Carregaro, R. L. (2017). Do virtual reality games improve mobility skills and balance measurements in community-dwelling older adults? Systematic review and meta-analysis. *Clinical Rehabilitation*, 0269215517694677.

Nigg, J. T. (2005). Neuropsychologic theory and findings in attention-deficit/hyperactivity disorder: the state of the field and salient challenges for the coming decade. *Biological psychiatry*, *57*(11), 1424-1435.

Nincarean, D., Alia, M. B., Halim, N. D. A., & Rahman, M. H. A. (2013). Mobile augmented reality: The potential for education. *Procedia-Social and Behavioral Sciences*, *103*, 657-664.

Nyhus, Erika, and Francisco Barceló. "The Wisconsin Card Sorting Test and the cognitive assessment of prefrontal executive functions: a critical update." *Brain and cognition* 71.3 (2009): 437-451.

Oquendo, M. A., Baca-García, E., Graver, R., Morales, M., Montalvan, V., & Mann, J. (2001). Spanish adaptation of the Barratt impulsiveness scale (BIS-11). *Graver, Journal of Psychiatry, 15*(3), 147-155.

Orosz, K., Farkas, I. J., & Pollner, P. (2016). Quantifying the changing role of past publications. *Scientometrics*, *108*(2), 829-853.

Ostrosky-Solis, F., Ardila, A., Rosselli, M., Lopez-Arango, G., & Uriel-Mendoza, V. (1998). Neuropsychological test performance in illiterate subjects. *Archives of Clinical Neuropsychology*, *13*(7), 645-660.

Ozbek, C. S., Giesler, B., & Dillmann, R. (2004, January). Jedi training: playful evaluation of head-mounted augmented reality display systems. In *Proceedings of SPIE* (Vol. 5291, pp. 454-463).

Park, N., Lee, K. M., Jin, S. A. A., & Kang, S. (2010). Effects of pre-game stories on feelings of presence and evaluation of computer games. *International Journal of Human-Computer Studies*, *68*(11), 822-833.

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.

Parsons, T. D., & Rizzo, A. A. (2008). Initial validation of a virtual environment for assessment of memory functioning: virtual reality cognitive performance assessment test. *CyberPsychology & Behavior*, *11*(1), 17-25.

Parsons, T. D. (2011). Neuropsychological assessment using virtual environments: enhanced assessment technology for improved ecological validity. In *Advanced computational intelligence paradigms in healthcare 6. Virtual reality in psychotherapy, rehabilitation, and assessment* (pp. 271-289). Springer, Berlin, Heidelberg.

Parsons, T. D., Courtney, C. G., & Dawson, M. E. (2013). Virtual reality Stroop task for assessment of supervisory attentional processing. *Journal of Clinical and Experimental Neuropsychology,35*, 812–826. doi:10.1080/13803395.2013.82455

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

Parsons, T. D., & Parsons. (2016). *Clinical neuropsychology and technology* (Vol. 2016). Springer.

Parsons, T., Gaggioli, A., & Riva, G. (2017). Virtual reality for research in social neuroscience. *Brain sciences*, *7*(4), 42.

Parsons, T. D., Carlew, A. R., Magtoto, J., & Stonecipher, K. (2017). The potential of function-led virtual environments for ecologically valid measures of executive function in experimental and clinical neuropsychology. *Neuropsychological rehabilitation*, *27*(5), 777-807.

Perfetti, B., Varanese, S., Mercuri, P., Mancino, E., Saggino, A., & Onofrj, M. (2010). Behavioural assessment of dysexecutive syndrome in Parkinson's disease without dementia: a comparison with other clinical executive tasks. *Parkinsonism & related disorders*, *16*(1), 46-50.

Perry, S. (2008). Wikitude: Android App with Augmented Reality: Mind Blowing. *Digital Lifestyles*.

Poh, M. Z., Swenson, N. C., & Picard, R. W. (2010). A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE transactions on Biomedical engineering*, *57*(5), 1243-1252.

Poulin, V., Korner-Bitensky, N., & Dawson, D. R. (2013). Stroke-specific executive function assessment: A literature review of performance-based tools. *Australian Occupational Therapy Journal*, *60*(1), 3-19.

Proust-Lima, C., Amieva, H., Letenneur, L., Orgogozo, J. M., Jacqmin-Gadda, H., & Dartigues, J. F. (2008). Gender and education impact on brain aging: A general cognitive factor approach. *Psychology and aging*, *23*(3), 608.

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.

Radu, I. (2012, November). Why should my students use AR? A comparative review of the educational impacts of augmented-reality. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 313-314). IEEE.

Radu, I. (2014). Augmented reality in education: a meta-review and cross-media analysis. *Personal and Ubiquitous Computing*, *18*(6), 1533-1543.

Rand, D., Katz, N., Shahar, M., Kizony, R., & Weiss, P. L. (2005). The virtual mall: A functional virtual environment for stroke rehabilitation. *Annual Review of Cybertherapy and Telemedicine: A decade of VR*, *3*, 193-198.

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.

Ragan, E. D., Sowndararajan, A., Kopper, R., & Bowman, D. A. (2010). The effects of higher levels of presence on procedure memorization performance and implications for educational virtual environments. *Presence: Teleoperators and Virtual Environments*, *19*(6), 527-543.

Ragan, E. D., Kopper, R., Schuchardt, P., & Bowman, D. A. (2012). Studying the effects of stereo, head tracking, and field of regard on a small-scale spatial judgment task. *IEEE transactions on visualization and computer graphics*, *19*(5), 886-896.

Reimers, S., & Stewart, N. (2015). Presentation and response timing accuracy in Adobe Flash and HTML5/JavaScript Web experiments. Behavior research methods, 47(2), 309-327.

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.

Riva, G. (2009). Virtual reality: an experiential tool for clinical psychology. *British Journal of Guidance & Counselling*, *37*(3), 337-345.

Riva, G., Gaggioli, A., Grassi, A., Raspelli, S., Cipresso, P., Pallavicini, F., ... & Donvito, G. (2011). NeuroVR 2-A free virtual reality platform for the assessment and treatment in behavioral healthcare. *In MMVR* (pp. 493-495).

Rizzo, A. A., Buckwalter, J. G., Bowerly, T., Van Der Zaag, C., Humphrey, L., Neumann, U., ... & Sisemore, D. (2000). The virtual classroom: a virtual reality environment for the assessment and rehabilitation of attention deficits. *CyberPsychology & Behavior, 3*(3), 483-499.

Rizzo, A. A., Schultheis, M., Kerns, K. A., & Mateer, C. (2004). Analysis of assets for virtual reality applications in neuropsychology. *Neuropsychological rehabilitation*, *14*(1-2), 207-239.

Rizzo, A. A., Bowerly, T., Buckwalter, J. G., Klimchuk, D., Mitura, R., & Parsons, T. D. (2009). A virtual reality scenario for all seasons: the virtual classroom. *Cns Spectrums, 11*(1), 35-44.

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.

Roberts, A. C., Robbins, T. W., & Weiskrantz, L. E. (1998). The prefrontal cortex: executive and cognitive functions. *Oxford University Press*.

Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1994). *TEA: The Test of Everyday Attention*. Pearson.

Rock, D. L., & Nolen, P. A. (1982). Comparison of the standard and computerized versions of the Raven Coloured Progressive Matrices Test. *Perceptual and motor skills*, *54*(1), 40-42.

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.

Roy, A., Allain, P., Roulin, J. L., Fournet, N., & Le Gall, D. (2015). Ecological approach of executive functions using the Behavioural Assessment of the Dysexecutive Syndrome for Children (BADS-C): Developmental and validity study. *Journal of clinical and experimental neuropsychology*, *37*(9), 956-971.

Rosenberg, L. (1993). The use of virtual fixtures to enhance telemanipulation with time delay. In *Proceedings of the ASME Winter Anual Meeting on Advances in Robotics, Mechatronics, and Haptic Interfaces* (Vol. 49, pp. 29-36).

Rossetti, H. C., Lacritz, L. H., Cullum, C. M., & Weiner, M. F. (2011). Normative data for the Montreal Cognitive Assessment (MoCA) in a population-based sample. *Neurology*, *77*(13), 1272-1275.

Royall, D. R., Lauterbach, E. C., Cummings, J. L., Reeve, A., Rummans, T. A., Kaufer, D. I., ... & Coffey, C. E. (2002). Executive control function: a review of its promise and challenges for clinical research. A report from the Committee on Research of the American Neuropsychiatric Association. *The Journal of neuropsychiatry and clinical neurosciences*, *14*(4), 377-405.

Saposnik, G., Mamdani, M., Bayley, M., Thorpe, K. E., Hall, J., Cohen, L. G., & Teasell, R. (2010). Effectiveness of Virtual Reality Exercises in STroke Rehabilitation (EVREST): rationale, design, and protocol of a pilot randomized clinical trial assessing the Wii gaming system. *International Journal of Stroke*, *5*(1), 47-51.

Saxton, J., Morrow, L., Eschman, A., Archer, G., Luther, J., & Zuccolotto, A. (2009). Computer assessment of mild cognitive impairment. *Postgraduate medicine*, *121*(2), 177-185.

Schultheis, M. T., Himelstein, J., & Rizzo, A. A. (2002). Virtual reality and neuropsychology: upgrading the current tools. *The Journal of head trauma rehabilitation*, *17*(5), 378-394.

Shallice, T. I. M., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, *114*(2), 727-741.

Schmidt, M., Beck, D., Glaser, N., & Schmidt, C. (2017, June). A Prototype Immersive, Multi-user 3D Virtual Learning Environment for Individuals with Autism to Learn Social and Life Skills: A Virtuoso DBR Update. In *International Conference on Immersive Learning* (pp. 185-188). Springer, Cham.

Schwald, B., & De Laval, B. (2003). An augmented reality system for training and assistance to maintenance in the industrial context.

Sequeira, H., Hot, P., Silvert, L., & Delplanque, S. (2009). Electrical autonomic correlates of emotion. *International journal of psychophysiology*, *71*(1), 50-56.

Serino, S., Cipresso, P., Morganti, F., & Riva, G. (2014). The role of egocentric and allocentric abilities in Alzheimer's disease: a systematic review. *Ageing research reviews*, *16*, 32-44.

Seymour, N. E. (2008). VR to OR: a review of the evidence that virtual reality simulation improves operating room performance. *World journal of surgery*, *32*(2), 182-188.

Shallice, T. I. M., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, *114*(2), 727-741.

Shumaker, R. (Ed.). (2011). *Virtual and Mixed Reality-Systems and Applications: International Conference, Virtual and Mixed Reality 2011, Held as Part of HCI International 2011, Orlando, FL, USA, July 9-14, 2011, Proceedings* (Vol. 6774). Springer.

Skalski, P., & Tamborini, R. (2007). The role of social presence in interactive agent-based persuasion. *Media psychology*, *10*(3), 385-413.

Skarbez, R., Brooks Jr, F. P., & Whitton, M. C. (2018). A survey of presence and related concepts. *ACM Computing Surveys (CSUR)*, *50*(6), 96.

Skinner, B. F. (1953). *Science and human behavior* (No. 92904). Simon and Schuster.

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 of London B: *Biological Sciences, 364*(1535), 3549-3557.

Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI*, *3*, 74.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the Association for Information Science and Technology*, *24*(4), 265-269.

Song, H., Chen, F., Peng, Q., Zhang, J., & Gu, P. (2017). Improvement of user experience using virtual reality in open-architecture product design. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 0954405417711736.

Sonuga-Barke, E. J., Sergeant, J. A., Nigg, J., & Willcutt, E. (2008). Executive dysfunction and delay aversion in attention deficit hyperactivity disorder: nosologic and diagnostic implications. *Child and adolescent psychiatric clinics of North America*, *17*(2), 367-384.

Suso-Ribera, C., Fernández-Álvarez, J., García-Palacios, A., Hoffman, H. G., Bretón-López, J., Banos, R. M., ... & Botella, C. (2018). Virtual Reality, Augmented Reality, and In Vivo Exposure Therapy: A Preliminary Comparison of Treatment Efficacy in Small Animal Phobia. *Cyberpsychology, Behavior, and Social Networking*, *22*(1), 31-38.

Spitoni, G. F., Aragonaa, M., Bevacqua, S., Cotugno, A., & Antonucci, G. (2018). An ecological approach to the behavioral assessment of executive functions in anorexia nervosa. *Psychiatry research*, *259*, 283-288.

Spooner, D. M., & Pachana, N. A. (2006). Ecological validity in neuropsychological assessment: A case for greater consideration in research with neurologically intact populations. *Archives of clinical neuropsychology*, *21*(4), 327-337.

Stricks, L., Pittman, J., Jacobs, D. M., Sano, M., & Stern, Y. (1998). Normative data for a brief neuropsychological battery administered to English-and Spanish-speaking community-dwelling elders. *Journal of the International Neuropsychological Society*, *4*(4), 311-318.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, *18*(6), 643.

Stroop, J. R. (1992). Studies of interference in serial verbal reactions. Journal of Experimental *Psychology: General, 121*(1), 15.

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.

Stuss, D. T., & Knight, R. T. (Eds.). (2013). Principles of frontal lobe function. *Oxford University Press*.

Sundar, S. S., Xu, Q., & Bellur, S. (2010, April). Designing interactivity in media interfaces: A communications perspective. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2247-2256). ACM.

Sutherland, I. E. (1965). The ultimate display. Multimedia: From Wagner to virtual reality.

Sutherland, I. E. (1968, December). A head-mounted three dimensional display. In *Proceedings of the fall joint computer conference, part I* (pp. 757-764). ACM.

Tavares, J. V. T., Clark, L., Cannon, D. M., Erickson, K., Drevets, W. C., & Sahakian, B. J. (2007). Distinct profiles of neurocognitive function in unmedicated unipolar depression and bipolar II depression. *Biological psychiatry*, *62*(8), 917-924.

Thomas, B., Close, B., Donoghue, J., Squires, J., De Bondi, P., Morris, M., & Piekarski, W. (2000, October). ARQuake: An outdoor/indoor augmented reality first person application. In *the fourth international symposium on Wearable computers* (pp. 139-146). IEEE.

Tirapu-Ustárroz, J., Muñoz-Céspedes, J. M., & Pelegrín-Valero, C. (2002). Funciones ejecutivas: necesidad de una integración conceptual. *Revista de neurología*, *34*(7), 673-685.

Tornatore, J. B., Hill, E., Laboff, J. A., & McGann, M. E. (2005). Self-administered screening for mild cognitive impairment: initial validation of a computerized test battery. *The Journal of neuropsychiatry and clinical neurosciences*, 17(1), 98-105.

Trochim, W. M., & Donnelly, J. P. (2005). Research methods: The concise knowledge base. Cincinnati, OH: *Atomic Dog Publishing*.

Trochim, W. M. (2006). Theory of reliability.

Unger, R. K. (1979). Toward a redefinition of sex and gender. *American Psychologist, 34*(11), 1085.

Valmaggia, L. R., Latif, L., Kempton, M. J., & Rus-Calafell, M. (2016). Virtual reality in the psychological treatment for mental health problems: An systematic review of recent evidence. *Psychiatry Research, 236*, 189-195.

Van Hooren, S. A. H., Valentijn, A. M., Bosma, H., Ponds, R. W. H. M., Van Boxtel, M. P. J., & Jolles, J. (2007). Cognitive functioning in healthy older adults aged 64–81: a cohort study into the effects of age, sex, and education. *Aging, Neuropsychology, and Cognition*, *14*(1), 40-54.

Ventura, S., Baños, R. M., & Botella, C. (2018). Virtual and Augmented Reality: New Frontiers for Clinical Psychology. In *State of the Art Virtual Reality and Augmented Reality Knowhow. IntechOpen.*

Ware, C., Arthur, K., & Booth, K. S. (1993, May). Fish tank virtual reality. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems* (pp. 37-42). ACM.

Waterfall, R. C. (1979). Automating standard intelligence tests. *Journal of Audiovisual media in Medicine*, *2*(1), 21-24.

Wattanasoontorn, V., Boada, I., García, R., & Sbert, M. (2013). Serious games for health. *Entertainment Computing*, *4*(4), 231-247.

Wechsler, D. (1945). Wechsler memory scale.

Wechsler, D. (2008). Wechsler adult intelligence scale–Fourth Edition (WAIS–IV). San Antonio, *TX: NCS Pearson, 22*, 498.

Werner, P., Rabinowitz, S., Klinger, E., Korczyn, A. D., & Josman, N. (2009). Use of the virtual action planning supermarket for the diagnosis of mild cognitive impairment. *Dementia and geriatric cognitive disorders*, *27*(4), 301-309.

Wexelblat, A. (Ed.). (2014). Virtual reality: applications and explorations. *Academic Press*.

White, H. D., & Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the Association for Information Science and Technology, 32*(3), 163-171.

Wiloth, S., Lemke, N., Werner, C., & Hauer, K. (2016). Validation of a computerized, game-based assessment strategy to measure training effects on motor-cognitive functions in people with dementia. *JMIR serious games*, *4*(2), e12.

Wilson, B. A., Alderman, N., Paul W.. Burgess, Emslie, H., & Jonathan J.. Evans. (1996). *BADS: Behavioural assessment of the dysexecutive syndrome*.

Wilson, B. A., Evans, J. J., Alderman, N., Burgess, P. W., & Emslie, H. (1997). Behavioural assessment of the dysexecutive syndrome. *Methodology of frontal and executive function*, *239*, 250.

Wilson, B. A., Evans, J. J., Emslie, H., Alderman, N., & Burgess, P. (1998). The development of an ecologically valid test for assessing patients with a dysexecutive syndrome. *Neuropsychological rehabilitation*, *8*(3), 213-228.

Wrzesien, M., Burkhardt, J. M., Alcañiz, M., & Botella, C. (2011). How technology influences the therapeutic process: a comparative field evaluation of augmented reality and in vivo exposure therapy for phobia of small animals. *Human-Computer Interaction–INTERACT 2011*, 523-540.

Wrzesien, M., Burkhardt, J. M., Alcañiz Raya, M., & Botella, C. (2011, May). Mixing psychology and HCI in evaluation of augmented reality mental health technology. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems* (pp. 2119-2124). ACM.

Wrzesien, M., Alcañiz, M., Botella, C., Burkhardt, J. M., Bretón-López, J., Ortega, M., & Brotons, D. B. (2013). The therapeutic lamp: treating small-animal phobias. *IEEE computer graphics and applications, 33*(1), 80-86.

Zarghi, A., & Zarindast, M. R. (2011). Demographic variables and selective, sustained attention and planning through cognitive tasks among healthy adults. *Basic and clinical Neuroscience*, *2*(3), 58-67.

Zyda, M. (2005). From visual simulation to virtual reality to games. *Computer, 38*(9), 25-32.