Διάλογοι! Θεωρία και πράξη στις επιστήμες αγωγής και εκπαίδευσης

Τόμ. 5 (2019)

Αξιολόγηση 2D και 3D ψηφιακών εκπαιδευτικών παιχνιδιών από φοιτητές-παίκτες: η σημασία των χαρακτηριστικών τους

Polyxeni Kaimara, Emmanuel Fokides, Andreas Oikonomou, Penelope Atsikpasi, Ioannis Deliyannis

doi: [10.12681/dial.20704](https://doi.org/10.12681/dial.20704)

Dialogoi! Theory & Praxis in Education
2019, Volume 5, pp. 36-56
e-ISSN: 2459-3737

http://dx.doi.org/10.12681/dial.20704

Evaluating 2D and 3D serious games: The significance of student-player characteristics

Polyxeni Kaimara[1], Emmanuel Fokides[2], Andreas Oikonomou[3], Penelope Atsikpasi[2], Ioannis Deliyannis[1]

[1]Ionian University, [2]University of the Aegean, [3]School of Pedagogical and Technological Education

Abstract

Serious games are gaining an ever-increasing interest of many scholars of the learning process of all educational levels. Important aspects of serious games implementation in education relate to their features, their relevance to pedagogy and learners' views. This quantitative research focused on characteristics of pedagogical departments' students (gender, scientific background, prior gaming and ICT skills) and factors that affect and shape their learning experience when playing 2D or 3D serious games, i.e. immersion, enjoyment, perceived usefulness-knowledge improvement, perceived narratives' adequacy, perceived realism, perceived feedback's adequacy, perceived audiovisual adequacy, perceived relevance to personal interests, perceived goal's clarity, perceived ease of use, adequacy of the learning material and motivation. A total of 542 university students participated in the study. A series of non-parametric tests (Mann-Whitney U tests and Kruskal-Wallis H tests) revealed that students preferred the 2D game. Their gender, scientific background and ICT skills did not have an impact on their views. On the other hand, students with high game-playing competence seemed to have a better learning experience, as their scores in most of the factors were higher compared to the ones with lower game-playing competence. Those results led us to consider other aspects/skills of learners beyond ordinary demographics such as self-regulation, spatial cognition and mental rotation and to examine the potential of serious games to improve these skills.

Keywords: 2D game, 3D game, mental rotation, playing experience, serious games, spatial cognition, user experience, quantitative approach.

Περίληψη

Τα σοβαρά παιχνίδια (ψηφιακά εκπαιδευτικά παιχνίδια μαθησιακού σκοπού) κερδίζουν ολοένα και περισσότερο το ενδιαφέρον πολλών μελετητών της μαθησιακής διαδικασίας όλων των εκπαιδευτικών επιπέδων. Σημαντικές πτυχές της εφαρμογής τους στην εκπαίδευση σχετίζονται με τα χαρακτηριστικά τους, τη συνάφειά τους με την παιδαγωγική και τις απόψεις των μαθητών. Αυτή η ποσοτική έρευνα επικεντρώθηκε στα χαρακτηριστικά των φοιτητών παιδαγωγικών τμημάτων (φύλο, επιστημονικό υπόβαθρο, προηγούμενη εμπειρία με παιχνίδια και δεξιότητες ΤΠΕ) και στους παράγοντες που επηρεάζουν και διαμορφώνουν την μαθησιακή εμπειρία τους όταν παίζουν σοβαρά παιχνίδια δύο (2D) και τριών (3D) διαστάσεων, όπως είναι η εμβύθιση, η απόλαυση, η υποκειμενική αντίληψη των μαθητών σχετικά με τη χρησιμότητά τους για τη βελτίωση της γνώσης, την επάρκεια της αφήγησης, τον ρεαλισμό, την ανατροφοδότηση, την οπτικοακουστική επάρκεια, τη συνάφειά τους με τα

Correspondent author: *Polyxeni Kaimara,* a16kaim@ionio.gr, PhD Candidate, Department of Audiovisual Arts, Ionian University

e-publisher: *National Documentation Centre, National Hellenic Research Foundation*
URL: http://ejournals.epublishing.ekt.gr/index.php/dialogoi

προσωπικά ενδιαφέροντά τους, τη σαφήνεια του στόχου τους, την ευκολία χρήσης τους, την καταλληλότητα του μαθησιακού υλικού και τα κίνητρα. Συνολικά 542 φοιτητές συμμετείχαν στη μελέτη. Σειρά μη παραμετρικών αναλύσεων (Mann-Whitney U tests και Kruskal-Wallis H tests) αποκάλυψε ότι οι φοιτητές προτίμησαν το 2D παιχνίδι. Το φύλο, το επιστημονικό υπόβαθρο και οι δεξιότητες στις ΤΠΕ δεν είχαν αντίκτυπο στις απόψεις τους. Από την άλλη πλευρά, οι φοιτητές με υψηλή ικανότητα στη χρήση των παιχνιδιών φαίνεται να έχουν καλύτερη μαθησιακή εμπειρία, καθώς οι βαθμολογίες τους στους περισσότερους παράγοντες ήταν υψηλότερες σε σύγκριση με τους φοιτητές με χαμηλότερη ικανότητα. Αυτά τα αποτελέσματα μας οδήγησαν να προσανατολιστούμε στην εξέταση άλλων πτυχών/δεξιοτήτων των φοιτητών πέρα από τα συνήθη δημογραφικά στοιχεία, όπως είναι η αυτορρύθμιση, η χωρική ικανότητα και η νοητική περιστροφή και να ερευνήσουμε τη δυνατότητα των σοβαρών παιχνιδιών να βελτιώσουν αυτές τις δεξιότητες.

Λέξεις-κλειδιά: 2D παιχνίδι, 3D παιχνίδι, εμπειρία του παιχνιδιού, εμπειρία του χρήστη, νοητικός προσανατολισμός, ποσοτική προσέγγιση σοβαρά παιχνίδια, χωρική ικανότητα.

Introduction

Digital technology is present in almost every aspect of daily life and is encapsulated within most human activities. Its use in education has actively been researched for at least five decades (Law & Sun, 2012). Most 21st century educational systems adopt new pedagogical models that enable technology-driven learning. Information and Communication Technologies (ICT) lead to the emergence of new forms of literacy, such as audio-visual literacy and multimedia learning. Computer-based learning and its capacity to enhance the learning process is based on four pillars: learner, content, pedagogy and context (Sims, 2000). Examples of digital learning resources include online instructional presentations, interactive lessons, e-courses, computer-supported in-class presentations, virtual reality, 3D multi-user virtual environments (MUVEs), simulations and games (Cai, Goei, & Trooster, 2016; Gee, 2003; Fokides, 2017; Mayer, 2016; Mayer et al., 2014; Zhonggen, 2019).

The term "serious game" (SG) is often encountered in the literature when searching for games designed for educational purposes. Although there is a perception that SGs are not entertaining as their purpose is other than fun, Abt (1970) argued that SGs can (and should be) enjoyable. Since then, the range of SGs continues to grow but concerns regarding their effectiveness in the learning process still remain. Perhaps one of the main reasons for these concerns is related to the significant differences between game design and instructional design (Kirkley, Tomblin & Kirkley, 2005; Van Eck, Shute, & Rieber, 2017). Indeed, SGs have to balance entertainment, engagement and learning (Franzwa, Tang, Johnson, & Bielefeldt, 2014; Kaimara & Deliyannis, 2019; Westera, 2019). This task requires design teams able to formulate effective teaching/learning models embedded in innovative games (de Freitas, 2018; Kirkley, Tomblin, & Kirkley, 2005).

During the above process, various issues surfaced, with the user experience (UX) to be one of the major concerns for both researchers and practitioners (Lallemand, 2015). UX is a dynamic, context-dependent and subjective concept that emerges from interacting with a product, system, service or an object (Law, Roto, Hassenzahl, Vermeeren, & Kort, 2009). UX is important in serious gaming since the effect of an SG on players' behavioural change can be witnessed (Nacke, Drachen, & Göbel, 2010). Although UX design can play a pivotal role in ones learning experience, it has not received sufficient recognition for its role in determining the success of an educational product and future levels of adoption by educators and learners (O'Brien, 2016). The technology-enhanced learning systems with the appropriate tools and methods, such as SGs, enable the development of media-rich, highly-responsive and customized user-experiences designed to offer the end-user a rewarding, interesting and captivating learning process (Deliyannis & Kaimara, 2019). This exactly why in this study we focused on the user-based evaluation of UX.

User Experience

In the 1990s, Donald Norman (as cited in Lallemand, 2015) was among the first to use the term "user experience". He introduced this term because he believed that the term "usability" fails to holistically represent human-computer interactions. Usability refers to attributes that make a product easy to use. "Usability: the capability of the software product to be understood, learned, used and attractive to the user, when used under specified conditions" (Bevan, 2001, p. 537). Usability also refers to the extent to which a product can be used by specific users for achieving specific goals in specific contexts of use. However, a clear definition of UX is still lacking, as there are definitional problems at both conceptualization and measurement levels (Bernhaupt, 2010; Buck, Khan, Fagan & Coman, 2018; Hassenzahl, 2008; Koeffel, Hochleitner, Leitner, Haller, Geven & Tscheligi, 2010). This is due to the fact that practitioners and academics understand the concept of UX differently (Hassenzahl, 2008). For the industry, UX is perceived as a synonym of usability and user-centered-design, while academics notice differences between usability and UX. What is certain is that UX goes beyond usability by bringing experiential aspects into the process. Accordingly, experience design goes beyond user-centered design, as it puts more emphasis on the quality of the users' experience (Lallemand, Gronier, & Koenig, 2015).

UX assessment includes user-based and expert-based evaluation using qualitative as well as quantitative methods such as physiological measurements, self-reported measures, usability tests, expert evaluation, heuristics, cognitive walkthroughs and guidelines reviews (Almeida et al., 2018; Bernhaupt & Mueller, 2016). Moreover, during the design of educational material, it is necessary to take into account the specific characteristics of the people to whom it is addressed and the context in which it will be applied. Users' subjectivity is another critical factor, as it includes temporal, spatial, social and personal factors, as well as their literacy level. Coming to SGs, UX evaluation follows three methodological approaches: (a) the quality of the product (game system experience), (b) the quality of human-product interaction (individual player experience) and (c) the quality of this interaction in a given social, temporal, spatial or other context (Nacke, Drachen, & Göbel, 2010). The UX in SGs has been evaluated using a variety of factors including but not limited to immersion, fun, presence, involvement, engagement, flow, play and playability (Bernhaupt, 2010).

Playability is a set of attributes such as satisfaction, learning, efficiency, immersion, motivation, emotion and socialization (Sánchez, Zea, & Gutiérrez, 2009). Humanities scholars have chosen, from the scientific literature on virtual reality, the term presence, defined as “as the feeling of being there”. The terms immersion and presence are seen as to be interchangeable (McMahan, 2003). Overall “immersion means the player is caught up in the world of the game’s story, but it also refers to the player’s love of the game and the strategy that goes into it” (McMahan, 2003, p. 68). It also defines the level of enjoyment and fun (Koeffel et al., 2010). Enjoyment and fun motivate learners to play a game and they are both related to the concept of flow that is the sense of presence when fully involved in an activity. “The state of flow is felt when opportunities for action are in balance with the actor’s skills” (Csikszentmihalyi, 1975, p. 49). Csikszentmihalyi was the first who introduced the concept of flow and emphasized that people perceive opportunities for action according to their capabilities. Motivation and engagement are components of the player's satisfaction (Kaimara, Deliyannis, Oikonomou, Papadopoulou, & Fokides, 2018; Kaimara & Deliyannis, 2019). Satisfaction includes several dimensions such as flow, immersion, fun, aesthetics, compelling experiences, presence, pleasure and enjoyment. Phan, Keebler and Chaparro (2016) concluded that the satisfaction that users receive when playing games is composed of nine factors: usability/playability, narratives, play engrossment, enjoyment, creative freedom, audio aesthetics, personal gratification, social connectivity and visual aesthetics. Learner's satisfaction is a multifaceted process, which depends on both its internal motives of user/player and the SG itself.

Research also takes into account the demographic characteristics of users/players/learners and how these characteristics are related to different game elements that can be incorporated into factors (e.g., gameplay, usability, engagement and motivation). The impact of age, gender and prior gaming experience play a significant role in gameplay performance. They also play a significant role in the game design and development process, as they give designers some guidelines for users/players profiles and how to incorporate mechanics and dynamics in order to create an effective learning environment that is the SGs' objective (Erfani et al., 2010; Spieler & Slany, 2018, Wang, Rajan, Sankar, & Raju, 2016). Previous studies suggested that the amount of time spent playing video games was significantly higher in males than females. Hu and Liu (2010) noted that users perceive games differently depending on their game experience and gender. In 2018, although women accounted for nearly 45 percent of all gamers in the United States (Statista, 2018), this report does not provide more information about the genre of games they prefer. Other studies concluded that playing games in terms of quantity (i.e., how much and how often one plays games and quality (i.e., what kind of games one plays), is associated with social and gender stereotypes (Wasserman & Rittenour, 2019). Several researchers found gender differences according to game mode (e.g. Massive Multiplayer Online Games, 3D environment) and genre (e.g., puzzles, sports games, strategy and role-playing games). Game content moderates the effect of gender on learning achievement and motivation. Females prefer brain-oriented and exploratory genres such as 2D board games, puzzles, quests and skill games, while males prefer 3D, shooter games, role-playing and strategy games (Chung & Chang, 2017; Dindar, 2018; Gecu & Cagiltay, 2015; Veltri, Krasnova, Baumann, & Kalayamthanam, 2014).

It is obvious that flow, enjoyment, fun and immersion are factors that concern academics and designers not only of digital games but also of games in general. Assessing a product by the users related to its usability, understanding, attractiveness and usefulness is an important process shaping the manufacture of the product. When this product is additionally designed for learning purposes, all these factors need to be related to the target group characteristics. Thus, factors such as age, gender, educational level, special skills are particularly significant as they identify the interests and abilities of the target group. SGs are slowly being introduced into the educational process, often complementary to more traditional teaching methods, such as lectures or more modern ones, such as asynchronous education, mainly in higher education and postgraduate studies. What, however, are students' views on their use in general and their effectiveness in learning?

Purpose of the study

The authors, after an extensive literature review on the evaluation of digital games in education, found that although many studies have shown that a student performs better when engaged in learning, something that games can do, unlike traditional curricula and methodologies, more research is needed that will lead to the standardization of criteria which the designers of serious games need to adopt. On the other hand, fewer studies examined the users' learning experience when playing serious games. Thus, they developed a scale (questionnaire) for measuring the factors that affect and ultimately shape, the learning experience. The questionnaire was tested for its validity and reliability several times (Fokides, Kaimara, Deliyannis, & Atsikpasi, 2019; Fokides, Atsikpasi, Kaimara & Deliyannis, 2019a, 2019b; Kaimara, Fokides, Plerou, Atsikpasi & Deliyannis, 2020). Utilizing the questionnaire, the purpose of this quantitative research was to correlate learners' characteristics such us gender, scientific background, prior gaming and ICT skills with twelve factors that are considered to affect and shape their learning experience when playing 2D and 3D serious games: immersion, enjoyment, perceived usefulness-knowledge improvement, perceived narratives' adequacy, perceived realism, perceived feedback's adequacy, perceived audiovisual adequacy, perceived relevance to personal interests, perceived goal's clarity, perceived ease of use, adequacy of the learning material and motivation.

Method

In our study, the UX was evaluated while university students played two ready-made products, a typical 2D and 3D SG. In order to evaluate UX, to examine the possible statistically significant differences among users and to record if users pay attention to different aspects of games according to their particular characteristics, the SGs' type (2D or 3D) and the quality of human-game interaction were taken into account. In our previous researches (Fokides, Kaimara, Deliyannis, & Atsikpasi, 2019; Fokides, Atsikpasi, Kaimara & Deliyannis, 2019a, 2019b; Kaimara, Fokides, Plerou, Atsikpasi, & Deliyannis, 2020), we concluded that twelve factors can be used for assessing SGs: immersion, enjoyment, perceived usefulness-knowledge improvement, perceived narratives' adequacy, perceived realism, perceived feedback's adequacy, perceived audiovisual adequacy, perceived relevance to personal interests, perceived goal's clarity, perceived ease of use, adequacy of the learning material and motivation. In the present study, learners' experience was evaluated via a questionnaire which examined the above factors (see Appendix). For the purposes of the survey, the data was collected in accordance with the ethical rules of the Universities that participated, the games were played in the Departments' Laboratories and the completed questionnaires were anonymous.

Research Questions

Our literature review revealed that important factors that can affect UX include the product design itself (2D or 3D gaming type), gender and prior gaming experience. It was also examined whether the technological competence of gamers affects their user experience. In this context, five research questions had been formulated:

1. Are there any statistically significant differences between the 2D and 3D games?
2. Can gender diversify the user experience?
3. Are there any statistically significant differences depending on the users' scientific background?
4. Are there any differences depending on their ICT competencies?
5. Are there differences depending on their game-playing competencies?

Participants and duration of the project

Students from the Department of Audiovisual Arts of the Ionian University in Corfu, Greece and the Department of Primary Education of the University of the Aegean in Rhodes, Greece were the study's participants. Both groups of students are potential users of the serious games used in this study (presented in the "Materials" section). An invitation was posted to Facebook groups of these two departments and on the e-class platforms and addressed to those students interested in participating in the project. The participants were informed that they would play one SG (or two SGs if they were interested in doing so) and then fill out a short questionnaire. They were also informed that the survey was conducted on a voluntary basis, that consent to participation was considered to have been given by completing the anonymous questionnaire. The total number of students enrolled in this process was 542. The SGs were available to be played in the Laboratories for a two-month period, from mid-January to mid-March 2018.

Materials

The survey's material was based on two games, "ARTé Mecenas" and "Variant: Limits", developed by Triseum (https://triseum.com/) both for the high school and university level students, classified as SGs.

"ARTé Mecenas" a 2D resource-management game, supports traditional college-level Art History and Art Appreciation courses. It teaches the interconnectedness of local and international economies in Renaissance Italy, how those economies influenced art and art patronage and give players/learners a unique perspective on the Italian Renaissance of the 15th to 16th centuries. It is designed to provide a learning experience where the player/learner can develop knowledge of famous artists, artworks, vocabulary and language of art and art history. Given a variety of scenarios, the player-learner: (a) develops an understanding of the interconnected networks of Renaissance economics, art patronage and production, including art markets, conventional banking, trade and alternative banking practices such as usury, (b) develops strategies to evaluate the impact of art and architecture patronage on generating spiritual and religious status and social and political prestige and (c) distinguishes between major artistic media, forms, techniques and theoretical and critical concepts to develop a more holistic interpretation of the Renaissance era.

"Variant: Limits" is a 3D game designed for teaching advanced mathematics. Students are engaged with an interface that allows them to develop a conceptual understanding of calculus via an experiential learning environment without relying on terminology, formulas and calculations. The player/learner (a) learns the nature of limits, the value of a limit and identifies when a function has continuity, (b) relates the graphical and algebraic representations of a function and applies the rules and principles of limits to determine the limit of a function, (c) uses the concept of continuity and relates it to the nature of limits and learns and applies the Intermediate Value Theorem in various contexts and (d) learns to determine function behaviors as x infinitely increases or decreases and identifies vertical asymptotes and oscillating behaviors of functions.

Instrument

For data collection purposes a questionnaire available online was used in this study. It was developed for evaluating digital educational material (serious games included) by measuring a total of twelve subjective factors. Besides demographic information (such as age, gender, scientific background, ICT and game-playing competence), respondents were asked to indicate their agreement to fifty-four statements in a five-point Likert-type scale (worded strongly agree to strongly disagree). Its reliability and factorial structure were tested and confirmed in previous studies (Fokides, Kaimara, Deliyannis, & Atsikpasi, 2019; Fokides, Atsikpasi, Kaimara, & Deliyannis, 2019a, 2019b; Kaimara, Fokides, Plerou, Atsikpasi, & Deliyannis, 2020). The questionnaire's items of the "Scale for Measuring the Learning Experience in Serious Games" are presented in the Appendix.

Procedure

Participants were asked to choose and play one of the two games (or both of them) and their gaming process was completed when certain conditions were met: they were instructed to play the game for a minimum of two hours and/or complete at least two levels. Both games feature an introductory/tutoring level, enabling players to familiarize with the use of the controls and user-interface that was not counted into the overall gaming time. After playing the SGs, each student-player filled an electronic questionnaire.

Results

All in all, 303 questionnaires were for the 2D game and 239 for the 3D game. Most participants came from the Department of Primary Education (DoPE), University of the Aegean (*N* = 343) and the rest (*N* = 199) came from the Department of Audiovisual Arts (Avarts), Ionian University. The data were imputed into SPSS 25 for statistical analyses. As expected, females were, by far, more than males (66% and 33% respectively). More than half of the participants were between 19 and 23 years of age, while very few were above the age of 28. Since most participants came from the Department of Primary Education, it was quite logical that most had a social sciences background rather than an inclination towards natural sciences (64% and 36% respectively). As for their ICT-related skills, these were slightly above the mean (*M* = 3.58, *SD* = 0.80), while their games-related skills were very close to the mean (*M* = 3.11, *SD* = 1.10). In order to discover which factors played an important role in both games, the average of the questions corresponding to each factor was calculated (Table 1). It has to be noted that a reliability analysis was run for the questionnaire as a total and for each of its constructs. Cronbach's alpha for the questionnaire was .87 and for the factors, it ranged between .88 and .95, well above DeVellis's (2016) recommendations (> .70).

	2D game		3D game	
	M	*SD*	*M*	*SD*
Immersion	2.52	0.97	2.49	0.98
Enjoyment	3.62	0.99	3.30	1.00
Perceived learning effectiveness	3.30	0.97	2.93	1.00
Perceived realism	2.52	0.97	2.24	0.95
Perceived narration's adequacy	3.42	0.97	3.06	1.00
Perceived audiovisual adequacy	3.52	0.95	3.45	0.92
Perceived goals' clarity	3.52	0.96	3.08	0.90
Perceived feedback's adequacy	3.90	0.65	3.65	0.65
Perceived ease of use/playability	3.75	0.91	3.56	0.91
Perceived adequacy of the learning material	3.53	0.85	3.11	0.99
Perceived motivation	3.83	0.98	3.56	0.96
Perceived relevance to personal interests	2.80	0.96	2.46	0.93

Table 1
Means per factor and per game

The sample size was considered more than adequate for inferential statistical testing. That is because, for 2 groups with N_{2D} game = 303 and N_{3D} game = 239, a significance level of .05, and an expected effect size of .10, the power value was .91, which is considered excellent (Cohen, 2013). One-way ANOVA tests were conducted for comparing both games, in order to determine if they had any statistically significant differences. Prior to conducting these tests, we checked whether the assumptions for ANOVA testing were violated. We found that: (a) the number of participants was not the same in both games, (b) the data were not normally distributed in many cases, as assessed by Q-Q plots and the Shapiro-Wilk test, and (c) the homogeneity of variance was also violated in some cases, as assessed by Levene's Test of Homogeneity of Variance. Given that the assumptions for ANOVA testing were violated, it was decided to proceed using non-parametric tests, namely the Mann-Whitney U test and the Kruskal-Wallis H test. Although these tests do not require normally distributed data, they require similarly shaped data distributions (Corder & Foreman, 2009; Siegel & Castellan, 1988), as was the case in the present study. The results are presented in the sections to follow.

Are there any statistically significant differences between the 2D and 3D games?

Differences between the two games were noted in almost all factors. Indeed, the 2D game received higher evaluation scores than the 3D game in all but two cases (immersion and perceived audiovisual adequacy). Furthermore, it seems that the differences between the two games were small (r < 0.30) (Table 2).

	2D game	3D game	Z	*p*

	Mean rank scores		**Mann-Whitney U**			**Effect size (*r*)**
Immersion	273.03	269.55	35743.50	-0.26	.797	-
Enjoyment	295.46	241.12	28947.50	-4.02	< .001	0.17
Perceived learning effectiveness	297.61	238.40	28297.50	-4.38	< .001	0.19
Perceived realism	291.40	246.27	30179.50	-3.34	.001	0.14
Perceived narration's adequacy	295.49	241.09	28940.00	-4.03	< .001	0.17
Perceived audiovisual adequacy	277.78	263.54	34305.50	-1.05	.293	-
Perceived goals' clarity	305.03	228.99	26047.50	-5.65	< .001	0.24
Perceived feedback's adequacy	298.67	237.05	27975.00	-4.56	< .001	0.21
Perceived ease of use/playability	287.17	251.63	31460.00	-2.63	.009	0.11
Perceived adequacy of the learning material	299.76	235.68	27646.50	-4.75	< .001	0.20
Perceived motivation	293.58	243.51	29518.00	-3.72	< .001	0.16
Perceived relevance to personal interests	296.07	240.35	28764.00	-4.14	< .001	0.18

Table 2

Differences between the two games

Given that there were differences between the games, in an attempt to explain them, it was decided to examine whether these differences were due to variances in gender, scientific background, ICT or game-playing competencies.

Can gender diversify the user experience?

In the 3D game, there were no differences between genders in any of the questionnaire's factors. In the 2D game, two differences were observed. Males gave higher ratings than females in perceived goals' clarity (mean $rank_{males}$ = 171.05, mean $rank_{females}$ = 142.05, *U* = 8367.00, *Z* = -2.75, *p* = .006, *r* = 0.27). The same applied for the perceived relevance to personal interests (mean $rank_{males}$ = 174.88, mean $rank_{females}$ = 140.04, *U* = 7968.50, *Z* = -3.31, *p* = .001, *r* = 0.23). Given that in both cases the effect size was small, it can be concluded that the participants' gender did not have any effect on how they viewed both games.

Are there any statistically significant differences depending on the players' scientific background?

The participants' scientific background (social sciences/natural sciences) seems to have played role in both games, but without affecting the same (or too many) factors. Participants having a natural sciences background rated the 2D game higher that the ones having a social sciences background in (a) immersion (mean $rank_{social}$ = 143.83, mean $rank_{natural}$ = 167.87, *U* = 8665.00, *Z* = -2.271, *p* = .023, *r* = 0.13), (b) perceived learning effectiveness (mean $rank_{social}$ = 143.02, mean $rank_{natural}$ = 169.45, *U* = 8503.00, *Z* = -2.49, *p* = .013, r = 0.14), and (c) perceived ease of use (mean $rank_{social}$ = 137.99, mean $rank_{natural}$ = 179.21, *U* = 7497.00, *Z* = -3.89, *p* <001 , *r* = 0.22). Participants with social sciences background rated the 3D game higher that the ones with natural sciences background only in perceived goals' clarity (mean $rank_{social}$ = 127.61, mean $rank_{natural}$ = 107.17, *U* = 5533.50, *Z* = -2.23, *p* = .026, *r* = 0.15). Then again, participants having a natural sciences background rated the 3D game higher that the ones having a social sciences background in (a) perceived ease of use (mean $rank_{social}$ = 112.72, mean $rank_{natural}$ = 132.28, *U* =

5582.50, $Z = -2.12$, $p = .034$, $r = 0.14$) and (b) relevance to personal interests (mean $\text{rank}_{social} = 109.70$, mean $\text{rank}_{natural} = 137.35$, $U = 5130.50$, $Z = -3.01$, $p = .003$, $r = 0.20$). In all cases the effect size was small.

Are there any differences depending on the players' ICT competencies?

Coming to the impact the participants' ICT competence had on both games, rather interesting differences were observed. In the 2D game, participants highly competent in ICT compared to participants not competent in ICT, rated statistically significantly higher (a) enjoyment [$H(4) = 13.69$, $p = .008$], (b) perceived learning effectiveness [$H(4) = 12.69$, $p = .015$], (c) perceived goals' clarity [$H(4) = 20.36$, $p < .001$], (d) perceived ease of use [$H(4) = 22.13$, $p < .001$], and (e) relevance to personal interests [$H(4) = 22.04$, $p < .001$]. On the other hand, in the 3D game, participants with high ICT competences rated higher only perceived feedback's adequacy [$H(4) = 11.29$, $p = .024$], while no other statistically significant differences were observed.

Are there any differences depending on the players' game-playing competencies?

Finally, the impact the participants' game-playing competence had on both games was the most prominent one. In the 2D game, participants highly competent in playing games compared to participants not so competent, gave statistically significantly higher scores to (a) immersion [$H(4) = 12.68$, $p = .013$], (b) enjoyment [$H(4) = 16.56$, $p = .002$], (c) perceived learning effectiveness [$H(4) = 13.11$, $p = .011$], (d) realism [$H(4) = 13.83$, $p = .008$], (e) perceived narration's adequacy [$H(4) = 12.47$, $p = .014$], (f) perceived goals' clarity [$H(4) = 23.78$, $p < .001$], (g) perceived ease of use/playability [$H(4) = 27.49$, $p < .001$], (h) perceived adequacy of the learning material [$H(4) = 11.98$, $p = .017$], and (i) relevance to personal interests [$H(4) = 44.40$, $p < .001$]. In the 3D game, participants highly competent in playing games gave higher scores to (a) immersion [$H(4) = 16.71$, $p = .002$], (b) perceived learning effectiveness [$H(4) = 11.73$, $p = .019$], (c) perceived realism [$H(4) = 21.85$, $p < .001$], (d) perceived goals' clarity [$H(4) = 13.04$, $p = .011$], and (e) perceived feedback's adequacy [$H(4) = 12.88$, $p = .012$].

Discussion

Table 3 summarizes the study's findings. Evidently, the 2D game was considered better in many factors (see Table 1 and Table 2), but the effect size was small, meaning that the difference was statistically significant but not of practical interest. Plass and colleagues (2019) found a small emotional effect related to presence when they compared 2D and 3D characters on screen-based computer games. Maybe the fact that our students played the 3D game (Variant: Limits) on screen could explain why the game failed to put the students-players in full immersion. Participants' gender was not important as it affected just a couple of factors. The relevant literature discussing gender differences in 2D and 3D games as well as in SGs, indicated that educational computer games in schools are considered as effective and motivational learning environments, regardless of students' gender (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012; Hainey, Connolly, Boyle, Wilson, & Razak, 2016; Mayer, 2019; Papastergiou, 2009). It seems that the decisive factor modifying the findings of gender surveys is the time spent in playing games. The more the females play, the more experienced and skillful they are in handling a game (Statista, 2018). These results are in concordance with the findings of other studies noting that the amount of time spent playing video games was significantly higher in males than females (Hu & Liu, 2010) and consequently, males are more familiar and experienced with game-playing. On the other hand, given enough time, males and females devote the same amount of time in playing games, so this gap is narrowed (Gecu & Cagiltay, 2015; Statista, 2018) and, thus, their differences are practically not significant. Dealing with a project systematically creates self-regulatory conditions and feedback resulting in satisfaction and learning. Self-regulating learning is referred to the learning process in which learners use self-regulatory skills such as self-assessment, self-directing, control and adaptation to acquire knowledge (Zimmerman, 1989).

Referring to the results related to the type of game (2D or 3D) and gender, spatial cognition is important to be mentioned. In the literature, gender differences in spatial cognition have been well documented (Moreau, Mansy-Dannay, Clerc, & Guerrien, 2010; Peters, Lehmann, Takahira, Takeuchi, & Jordan, 2006; Spence & Feng, 2010). Although there is evidence of the existence of gender differences in spatial cognition based on a variety of biological and environmental explanatory what is emphasized is that spatial skills are crucial for both females and males because mental rotation (one of the many spatial tasks) is related to STEM (science, technology, engineering and math) achievement (Yurt & Tünkler, 2016). Spatial thinking is malleable and improves with training and playing video games could reduce gender differences in spatial cognition (Levine, Foley, Lourenco, Ehrlich, & Ratliff, 2016; Zemiek, 2006).

ICT knowledge gives advantages/convenience to those who possess it, but only to the 2D game. Passing from 2D to 3D creates mental/cognitive requirements that invalidate the advantage of any ICT knowledge possessed by players. An interpretation could be that the 3D game was essentially a game that did not incorporate all the features of the 3D games as it was played on a 2D screen and could, therefore, it did not cause them much interest.

	2D/3D game	Gender		Background		ICT comp.		Game comp.	
		2D game	3D game	2D game	3D game	2D game	3D game	2D game	3D game
Immersion	-	-	-	N	-	-	-	H	H
Enjoyment	2D	-	-	-	-	H	-	H	-
Perceived learning effectiveness	2D	-	-	N	-	H	-	H	H
Perceived realism	2D	-	-	-	-	-	-	H	H
Perceived narration's adequacy	2D	-	-	-	-	-	-	H	-
Perceived audiovisual adequacy	-	-	-	-	-	-	-	-	-
Perceived goals' clarity	2D	M	-	-	S	H	-	H	H
Perceived feedback's adequacy	2D	-	-	-	-	-	H	-	H
Perceived ease of use/playability	2D	-	-	N	N	H	-	H	-
Perceived adequacy of the learning material	2D	-	-	-	-	-	-	H	-
Perceived motivation	2D	-	-	-	-	-	-	-	-
Perceived relevance to personal interests	2D	M	-	-	N	H	-	H	-

Notes: - = (NS) not statistically significant difference; M = males; F= females; S = social sciences background; N = natural sciences background; H = high ICT or game playing competence

Table 3
Results' summary

We also found that participants' scientific background was not so important. It was expected participants having a natural sciences background to give higher scores to the 3D game. The first interpretation is based on the content of the 3D game that is related to their interests associated with mathematics and calculus. Second, the 3D game players who have achieved increased spatial processing skills might have higher mental rotation abilities than both 2D computer game players and non-players (Gecu & Cagiltay, 2015). The relationship between spatial ability and success in science and mathematics has been reported in several publications (Charlesworth, Drummer, Hungwe, & Sorby, 2005; Dawson, 2019). Excellence in science, technology, engineering and math fields (STEM) is strongly correlated with the spatial ability and spatial skills are associated with performance in mathematics and science courses as well as the choice of mathematics and science courses in college (Spence & Feng, 2010).

It was also expected participants having a social sciences background to give higher scores to the 2D in the same factor, but this expectation was not confirmed. According to Yurt and Tünkler (2016), it is indicated that spatial visualization and mental rotations abilities of social studies teachers' are at a low level. However, going back to the content, the 2D game under research was about art history and in order to proceed several decisions had to be made, as searching the world web for answers. This requires problem-solving skills. These skills are not related to the scientific background but to the general ability to gather information as information processing skill.

Participants with a background related to natural sciences gave higher scores in the 2D game to immersion, learning effectiveness and ease of use. In the 3D game, they gave higher scores to ease of use and relevance to personal interests. The Keller's model, ARCS-V (Keller, 2010) which refers to the five principles of the learning process related to student motivation that is attention, relevance, confidence, satisfaction and self-regulation, can explain these findings. The common factor in both games was ease of use. Participants having a natural sciences background have more self-regulation, because when they play games obviously they do not just interact with the content but also try to understand the mechanics and other game elements. Thus, students were either from natural or social studies, motivated or not to play the game according to their interests, the curiosity, the relevance of the content of the game to the subject of their studies, the satisfaction but also their self-regulation. The same conclusion was reached by Tiede and Grafe (2018, p. 1), who evaluated the concrete games using Keller's model: "the overall results show that both games were successful in stimulating motivation and classroom engagement with the students, even though the effects varied between the two games in certain regards and were discovered to depend on numerous factors in the context of interpersonal differences".

These results can lead us to the conclusion that the scientific background of the respondents is not decisive, at least not as much as other factors not measured in our research, for example, participants' learning style, spatial cognition and mental rotation

(Garmen, et al., 2019; Raptis, Fidas, & Avouris, 2016). In addition, participants highly ICT skilled rated higher, in the 2D game, enjoyment, learning effectiveness, goals' clarity and ease of use. In the 3D game, they rated higher only feedback's adequacy. Only ease of use, which was also expected, was confirmed. Students with high ICT competencies do not face problems in handling computers nor to their access the world web site.

Finally, it seems that participants highly competent in playing games appreciated the 2D game more than the 3D game. Probably this is due to the difficulty of participants to take into account elements from the 3D environment owing to the exhaustion of their mental resources. According to Anderson and colleagues (2019), learning activities employing 3D models require larger working memory resources and this results in less free capacity in total working memory to engage in learning activity itself. Thus, the transition from 2D to 3D environment appears to be equally difficult for all participants.

Implications for research and practice

The quantitative research was conducted using a questionnaire designed to evaluate games with different content and type. It is a sensitive tool tracking differences in a set of variables related to player characteristics and game type. Game designers, developers and educators of all educational levels could benefit using this questionnaire for research and design purposes**.** Investigating serious games can also take into account the psychological aspects of students, such as self-regulation, spatial cognition and mental rotation.

Limitations and future work

As is the case with any empirical research, the present study has limitations that reduce the generalizability of our findings regarding the concrete games, which were employed. First, our sample is not representative but comes from two specific university departments selected based on ease of access. Second, the two games have different learning content. It would be interesting to examine the same content in different game environments so that their learning effectiveness could be attributed to their type (2D or 3D). It would be, also, interesting the questionnaire could be used in a population of students and a sample of representative studies/faculties that they could choose based on study content in order for spatial cognition to be investigated. Taking into account the findings and research on spatial cognition, authors are geared towards exploring applications that not only share the same content, but will also be played on different platforms to cover the three conditions: (a) the same content, (b) comparison of 2D vs 3D application and (c) comparison of 3D application on a full immersion and isolation platform vs screens. In addition, our future work is about the evaluation of the different type (2D/3D) and genre of games (puzzles, simulation, etc.) by teachers and students, utilizing the same questionnaire, exploiting cutting-edge technologies like virtual reality and augmented reality, in deferent educational levels (primary and secondary) and sectors (general, special education and inclusive educational settings).

Conclusion

User experience is one of the most significant factors related to any product. Regarding the SGs, players' characteristics as users, were examined gender, scientific background, prior gaming experience, ICT knowledge) over twelve factors that shape the UX in SGs (i.e., immersion, enjoyment, perceived usefulness-knowledge improvement, perceived narratives' adequacy, perceived realism, perceived feedback's adequacy, perceived audiovisual adequacy, perceived relevance to personal interests, perceived goal's clarity, perceived ease of use, adequacy of the learning material and motivation).

Players had the convenience of counting a number of variables at a time. For this reason and as their number grows, the more difficult the differences between expert and non-expert players initiate or fewer initiates in ICT, are identified. Our results lead us to conclude that 2D games offered more to identify differences in their users in the various variables. On the other hand, it is useful to introduce psychological variables as spatial cognition, mental rotation, motives, short-term/working memory and intelligence into further research that will better explain possible differences that will be identified. When playing serious games, players use working memory resources, which are different for each player and are related to their information processing system and experience in related works. When the information increases, the task becomes more difficult, so that players who have experience in related environments do better. What is important from this work is that it seemed that the specific evaluation tool designed to measure the perceptions of the players about the learning games has the sensitivity to discern even a few differences.

References

Abt C. C. (1970). *Serious games*. New York: Viking Press. https://doi.org/10.1177/000276427001400113

Almeida, P., Abreu, J., Silva, T., Varsori, E., Oliveira, E., Velhinho, A., … Oliveira, D. (2018). Iterative user experience evaluation of a user interface for the unification of tv contents. *Communications in Computer and Information Science*, *813*, 44-57. https://doi.org/10.1007/978-3-319-90170-1_4

Anderson, S. J., Jamniczky, H., Williams, C., Krigolson, O., Coderre, S., & Hecker, K. (2019). Learning from Two-Dimensional (2D) versus Three-Dimensional Anatomical Models: Assessing Working Memory Requirements Using Electroencephalography (EEG). *The FASEB Journal, 33*(1_supplement), 328-5.

Bernhaupt, R. (2010). User Experience Evaluation in Entertainment. In R. Bernhaupt (Ed.), *Evaluating user experience in games: Concepts and methods* (pp. 3-7). London: Springer. https://doi.org/10.1007/978-1-84882-963-3_1

Bernhaupt, R., & Mueller, F. F. (2016, May). Game user experience evaluation. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 940-943. ACM. https://doi.org/10.1145/2851581.2856683

Bevan, N. (2001). International standards for HCI and usability. *International journal of human-computer studies*, *55*(4), 533-552. https://doi.org/10.1006/ijhc.2001.0483

Buck, R., Khan, M., Fagan, M., & Coman, E. (2018). The User Affective Experience Scale: A measure of emotions anticipated in response to pop-up computer warnings. *International Journal of Human-Computer Interaction*, *34*(1), 25-34. https://doi.org/10.1080/10447318.2017.1314612

Cai, Y., Goei, S. L., & Trooster, W. (Eds.) (2016). *Simulation and serious games for education*. Singapore: Springer. https://doi.org/10.1007/978-981-10-0861-0

Charlesworth, P., Drummer, T., Hungwe, K., & Sorby, S. (2005, June). Developing 3D spatial visualization skills for non engineering students. *Proceedings of the 2005 American Society for Engineering Education Annual Conference & Exposition*, 10.428.1-10.428.11. American Society for Engineering Education.

Chung, L. Y., & Chang, R. C. (2017). The effect of gender on motivation and student achievement in digital game-based learning: A case study of a contented-based classroom. *Eurasia Journal of Mathematics, Science and Technology Education*, *13*(6), 2309-2327. https://doi.org/10.12973/eurasia.2017.01227a

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. New York: Routledge

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. https://doi.org/10.1016/j.compedu.2012.03.004

Corder, G. W., & Foreman, D. I. (2009). *Nonparametric statistics for non-statisticians: A step-by-step approach*. John Wiley & Sons.

Csikszentmihalyi, M., (1975). *Beyond boredom and anxiety*. San Francisco: Jossey-Bass. Retrieved from http://psy2.ucsd.edu/~nchristenfeld/Happiness_Readings_files/Class%207%20-%20Csikszentmihalyi%201975.pdf

Dawson, C. (2019). Tackling Limited Spatial Ability: Lowering One Barrier into STEM?. *European Journal of Science and Mathematics Education*, *7*(1), 14-31. Retrieved from https://files.eric.ed.gov/fulltext/EJ1203669.pdf

de Freitas, S. (2018). Are games effective learning tools? A review of educational games. *Journal of Educational Technology & Society*, *21*(2), 74-84. Retrieved from https://psycnet.apa.org/record/2018-18436-006

Deliyannis, I., & Kaimara, P. (2019). Developing Smart Learning Environments Using Gamification Techniques and Video Game Technologies. *Didactics of Smart Pedagogy*, 285-307. https://doi.org/10.1007/978-3-030-01551-0_15

DeVellis, R. F. (2016). *Scale development: Theory and applications* (Vol. 26). Sage publications.

Dindar, M. (2018). An empirical study on gender, video game play, academic success and complex problem solving skills. *Computers & Education*, *125*, 39-52. https://doi.org/10.1016/j.compedu.2018.05.018

Erfani, M., El-Nasr, M. S., Milam, D., Aghabeigi, B., Lameman, B. A., Riecke, B. E., Maygoli, H., & Mah, S. (2010). The effect of age, gender, and previous gaming experience on game play performance. In P. Forbrig, F. Paternó, A. Mark Pejtersen (Eds.), *Human-Computer Interaction. HCIS 2010. IFIP Advances in Information and Communication Technology, vol 332.* Berlin: Springer. https://doi.org/10.1007/978-3-642-15231-3_33

Fokides, E., Kaimara, P., Deliyannis, I., & Atsikpasi, P. (2019). Development of a scale for measuring the learning experience in serious games. Preliminary results. *Proceedings of 1st International Conference on Digital Culture & AudioVisual Challenges*, Department of Audiovisual Arts, Ionian University, Corfu June 1-2, 2018 (in press).

Fokides, E., Atsikpasi, P., Kaimara, P., & Deliyannis, I. (2019a). Factors influencing the subjective learning effectiveness of serious games. *Journal of Information Technology Education: Research, 18*, 437-466. https://doi.org/10.28945/4441

Fokides, E., Atsikpasi, P., Kaimara, P., & Deliyannis, I. (2019b). Let players evaluate serious games. Design and validation of the Serious Games Evaluation Scale. *International Computer Games Association Journal, 31*(3), 1-22. https://doi.org/10.3233/ICG-190111

Fokides, E. (2017). Pre-service teachers' intention to use MUVES as practitioners. A structural equation modeling approach. *Journal of Information Technology Education*, *16*, 47-68. https://doi.org/10.28945/3645

Franzwa, C., Tang, Y., Johnson, A., & Bielefeldt, T. (2014). Balancing fun and learning in a serious game design. *International Journal of Game-Based Learning (IJGBL)*, *4*(4), 37-57. https://doi.org/10.4018/ijgbl.2014100103

Garmen, P., Rodríguez, C., García-Redondo, P., & San-Pedro-Veledo, J. C. (2019). Multiple intelligences and video games: Assessment and intervention with TOI software. *Comunicar: Media Education Research Journal*, *27*(58), 95-104. https://doi.org/10.1111/j.1365-2729.2012.00489.x

Gecu, Z., & Cagiltay, K. (2015). Mental Rotation Ability and Computer Game Experience. *International Journal of Game-Based Learning (IJGBL)*, *5*(4), 15-26. https://doi.org/10.4018/ijgbl.2015100102

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment*, *1*(1), 20-20. https://doi.org/10.1145/950566.950595

Hassenzahl, M. (2008). User experience (UX): towards an experiential perspective on product quality. In *Proceedings of the 20th international Conference of the Association Francophone D'interaction Homme-Machine*, 11-15. ACM. https://doi.org/10.1145/1512714.1512717

Hainey, T., Connolly, T. M., Boyle, E. A., Wilson, A., & Razak, A. (2016). A systematic literature review of games-based learning empirical evidence in primary education. *Computers & Education*, *102*, 202-223. https://doi.org/10.1016/j.compedu.2016.09.001

Hu, F., & Liu, Y. (2010, October). Impact of experience and gender differences on users' perceptions on mobile game. *Proceedings of the Multimedia Technology (ICMT), 2010 International Conference,* 1-2. IEEE. https://doi.org/10.1109/icmult.2010.5629762

Kaimara, P., Deliyannis, I., Papadopoulou, A., Oikonomou, A., & Fokides, E. (2018). Smart Education: Τα ψηφιακά παιχνίδια ως κίνητρο για μάθηση [Smart Education: Digital games as motivation for learning]. *Proceedings of the 2nd International Experiential Conference on Applied Teaching*, 385-391. Drama, Greece: Educational Circle.

Kaimara, P., & Deliyannis, I. (2019). Why should I play this game? The role of motivation in Smart Pedagogy. In L. Daniela (Ed.), *Didactics of Smart Pedagogy* (pp. 113-137). Springer, Cham. https://doi.org/10.1007/978-3-030-01551-0_6

Kaimara, P., Fokides, E., Plerou, A., Atsikpasi, P., & Deliyannis, I. (2020). Serious games effect analysis on player's characteristics. *International Journal of Smart Education and Urban Society*, *11*(1), 75-91. https://doi.org/10.4018/IJSEUS.2020010106

Keller, J. M. (2010). Five fundamental requirements for motivation and volition in technology-assisted distributed learning environments. *Revista Inter Ação*, *35*(2), 305-322. https://doi.org/10.5216/ia.v35i2.12668

Kirkley, S. E., Tomblin, S., & Kirkley, J. (2005). Instructional design authoring support for the development of serious games and mixed reality training. *Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC)*. Retrieved from https://pdfs.semanticscholar.org/abeb/00c31a0cbc8920c37493f7922b95f333bc4c.pdf

Koeffel, C., Hochleitner, W., Leitner, J., Haller, M., Geven, A., & Tscheligi, M. (2010). Using heuristics to evaluate the overall user experience of video games and advanced interaction games. In R. Bernhaupt (Ed.), *Evaluating user experience in games: Concepts and Methods* (pp. 233-256). London: Springer. https://doi.org/10.1007/978-1-84882-963-3_13

Lallemand, C. (2015). *Towards consolidated methods for the design and evaluation of user experience* (Doctoral dissertation). University of Luxembourg, Luxembourg. Retrieved from http://orbilu.uni.lu/handle/10993/21463

Lallemand, C., Gronier, G., & Koenig, V. (2015). User experience: A concept without consensus? Exploring practitioners' perspectives through an international survey. *Computers in Human Behavior*, *43*, 35-48. https://doi.org/10.1016/j.chb.2014.10.048

Law, E. L. C., & Sun, X. (2012). Evaluating user experience of adaptive digital educational games with Activity Theory. *International Journal of Human-Computer Studies*, *70*(7), 478-497. https://doi.org/10.1016/j.ijhcs.2012.01.007

Law, E. L. C., Roto, V., Hassenzahl, M., Vermeeren, A. P., & Kort, J. (2009, April). Understanding, scoping and defining user experience: a survey approach. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 719-728). ACM. https://doi.org/10.1145/1518701.1518813

Levine, S. C., Foley, A., Lourenco, S., Ehrlich, S., & Ratliff, K. (2016). Sex differences in spatial cognition: Advancing the conversation. *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*(2), 127-155. https://doi.org/10.1002/wcs.1380

Mayer, R. E. (2016). What should be the role of computer games in education? *Policy Insights from the Behavioral and Brain Sciences*, *3*(1), 20-26. https://doi.org/10.1177/2372732215621311

Mayer, R. E. (2019). Computer games in education. *Annual review of Psychology*, *70*, 531-549. https://doi.org/10.1146/annurev-psych-010418-102744

Mayer, I., Bekebrede, G., Harteveld, C., Warmelink, H., Zhou, Q., van Ruijven, T., Lo, J., Kortmann, R., & Wenzler, I. (2014). The research and evaluation of serious games: Toward a comprehensive methodology. *British Journal of Educational Technology*, *45*(3), 502-527. https://doi.org/10.1111/bjet.12067

McMahan, A. (2003). Immersion, engagement, and presence: A method for analyzing 3-D video games. In M. J. P. Wolf & P. Bernard (Eds.), *The video game theory reader* (pp. 67–86). New York, NY: Routledge. Retrieved from https://alisonmcmahan.com/sites/default/files/pdfs/papers/ch3-McMahanrev.pdf

Moreau, D., Mansy-Dannay, A., Clerc, J., & Guerrien, A. (2010). Academic program and mental rotation performance: Evidence for a developmental effect on individual differences in early adulthood. *Education Sciences & Psychology*, *17*(2), 21-28.

Nacke, L. E., Drachen, A., & Göbel, S. (2010). Methods for evaluating gameplay experience in a serious gaming context. *International Journal of Computer Science in Sport*, *9*(2), 1-12. Retrieved from http://iacss.org/fileadmin/user_upload/IJCSS_FullPaper/Vol9_Ed2/IJCSS-Volume9_Edition2_4_nacke.pdf

O'Brien, S. (2016, June). Why User Experience Design is Critical to Driving and Maintaining User Engagement and Motivation for Online and Mobile Educational Tools. *Proceedings of the Future of Education Conference* (p. 328). Padova, Italy: Libreriauniversitaria.iit Edizioni Retrieved from https://conference.pixel-online.net/FOE/files/foe/ed0006/FP/2819-ENT1831-FP-FOE6.pdf

Papastergiou, M. (2009). Digital game-based learning in high school computer science education. *Computers & Education, 52*(1), 1-12. https://doi.org/10.1016/j.compedu.2008.06.004

Peters, M., Lehmann, W., Takahira, S., Takeuchi, Y., & Jordan, K. (2006). Mental rotation test performance in four cross-cultural samples (n= 3367): overall sex differences and the role of academic program in performance. *Cortex*, *42*(7), 1005-1014. https://doi.org/10.1016/s0010-9452(08)70206-5

Phan, M. H., Keebler J. R., & Chaparro, B. S. (2016). The Development and Validation of the Game User Experience Satisfaction Scale (GUESS). *Human factors*, *58*(8), 1217-1247. https://doi.org/10.1177/0018720816669646

Plass, J. L., Homer, B. D., MacNamara, A., Ober, T., Rose, M. C., Pawar, S., ... & Olsen, A. (2019). Emotional design for digital games for learning: The effect of expression, color, shape, and dimensionality on the affective quality of game characters. *Learning and Instruction*. https://doi.org/10.1016/j.learninstruc.2019.01.005

Raptis, G. E., Fidas, C. A., & Avouris, N. M. (2016, October). Do field dependence-independence differences of game players affect performance and behaviour in cultural heritage games? *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (pp. 38-43). ACM. https://doi.org/10.1145/2967934.2968107

Sánchez, J. G., Zea, N. P., & Gutiérrez, F. L. (2009, August). Playability: How to identify the player experience in a video game. *Proceedings of the IFIP Conference on Human-Computer Interaction*, 356-359. Berlin: Springer. https://doi.org/10.1007/978-3-642-03655-2_39

Siegel, S., & Castellan Jr, N. J. (1988). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.

Sims, R. (2000). An interactive conundrum: Constructs of interactivity and learning theory. *Australasian Journal of Educational Technology*, *16*(1). https://doi.org/10.14742/ajet.1821

Spence, I., & Feng, J. (2010). Video games and spatial cognition. *Review of General Psychology*, *14*(2), 92-104. https://doi.org/10.1037/a0019491

Spieler, B., & Slany, W. (2018). *Game development-based learning experience: Gender differences in game design*. arXiv preprint arXiv:1805.04457

Statista: The Statistics Portal (2018). Retrieved from https://www.statista.com/statistics/232383/gender-split-of-us-computer-and-video-gamers

Tiede, J., & Grafe, S. (2018, September). Piloting Two Educational Games in Five European Countries: Teachers' Perceptions of Student Motivation and Classroom Engagement. In *2018 10th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)* (pp. 1-8). IEEE. https://doi.org/10.1109/vs-games.2018.8493438

Tobias, S., Fletcher, J. D., & Wind, A. P. (2014). Game-based learning. In M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 485-503). New York, NY: Springer.

Yurt, E., & Tünkler, V. (2016). A Study on the Spatial Abilities of Prospective Social Studies Teachers: A Mixed Method Research. *Educational Sciences: Theory and Practice*, *16*(3), 965-986. https://doi.org/10.12738/estp.2016.3.0324

Van Eck, R. N., Shute, V. J., & Rieber, L. P. (2017). Leveling up: Game design research and practice for instructional designers. In R. Reiser & J. Dempsey (Eds.), Trends and issues in instructional design and technology (pp. 227-285). Upper Saddle River: Pearson Education. Retrieved from http://myweb.fsu.edu/vshute/pdf/levelingup.pdf

Veltri, N. F., Krasnova, H., Baumann, A., and Kalayamthanam, N. (2014). Gender differences in online gaming: a literature review. P*roceedings of the 20th Americas Conference on Information Systems.* Boston, MA: AMCIS. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.667.4530&rep=rep1&type=pdf

Wang, Y., Rajan, P., Sankar, C. S., & Raju, P. K. (2016). Let them play: the impact of mechanics and dynamics of a serious game on student perceptions of learning engagement. *IEEE*

Transactions on Learning Technologies, *10*(4), 514-525. https://doi.org/10.1109/tlt.2016.2639019

Wasserman, J. A., & Rittenour, C. E. (2019). Who wants to play? Cueing perceived sex-based stereotypes of games. *Computers in Human Behavior*, *91*, 252-262. https://doi.org/10.1016/j.chb.2018.09.003

Westera, W. (2019). Why and How Serious Games can Become Far More Effective: Accommodating Productive Learning Experiences, Learner Motivation and the Monitoring of Learning Gains. *Educational Technology & Society*, *22*(1), 59–69. Retrieved from https://www.j-ets.net/ets/journals/22_1/5.pdf

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of educational psychology*, *81*(3), 329-339. https://doi.org/10.1037//0022-0663.81.3.329

Zhonggen, Y. (2019). A Meta-Analysis of Use of Serious Games in Education over a Decade. *International Journal of Computer Games Technology*, *2019*. https://doi.org/10.1155/2019/4797032

Appendix: The questionnaire's items

Factor	Item
Immersion	I was deeply concentrated in the application If someone was talking to me, I couldn't hear him I forgot about time passing while using the application I felt detached from the outside world while using the application
Enjoyment	I think the application was fun I felt bored while using the application* I enjoyed using the application I really enjoyed studying with this application It felt good to successfully complete the tasks in this application I felt frustrated*
Perceived usefulness-knowledge improvement	I felt that this application can ease the way I learn This application was a much easier way to learn compared to the usual teaching This application made learning more interesting I felt that the application increased my knowledge I felt that I caught the basic ideas of what I was taught with this application I will definitely try to apply the knowledge I learned with this application
Perceived narratives' adequacy	I was captivated by the application's story from the beginning I enjoyed the fantasy or story provided by the application I could clearly understand the application's story I was very interested in seeing how the events in the application will unfold
Perceived realism	When interacting with the virtual objects, these interactions seemed like real ones There were times when the virtual objects seemed to be as real as the real ones The virtual objects seemed like real objects to me When I used the application, the virtual world was more real than the real world
Perceived feedback's adequacy	I received immediate feedback on my actions I was notified of new tasks immediately I received information on my success (or failure) on the intermediate goals immediately
Perceived audiovisual adequacy	I enjoyed the sound effects in the application I think the application's audio fits the mood or style of the application I felt the application's audio (e.g., sound effects, music) enhanced my (gaming) experience I enjoyed the music in the application I enjoyed the application's graphics

	I think the application was visually appealing I think the graphics of the application fit the mood or style of the application
Perceived relevance to personal interests	The content of this application was relevant to my interests I could relate the content of this application to things I have seen, done, or thought about in my own life It is clear to me how the content of the application is related to things I already know
Perceived goal's clarity	The application's goals were presented at the beginning of the application The application's goals were presented clearly The intermediate goals were presented at the beginning of each scene
Perceived ease of use	I think it was easy to learn how to use the application I found the application unnecessarily complex* I imagine that most people will learn to use this application very quickly I needed to learn a lot of things before I could get going with this application* I felt that I needed help from someone else in order to use the application because it was not easy for me to understand how to use it* It was easy for me to become skillful at using the application
Adequacy of the learning material	In some cases, there was so much information that it was hard to remember the important points* The exercises in this application were too difficult* I could not really understand quite a bit of the material in this application*
Motivation	This application did not hold my attention* When using the application, I did not have the impulse to learn more about the learning subject* The application did not motivate me to learn*

Note. * = Item for which its scoring was reversed