Journal of Information Technology Education: Research

An Official Publication of the Informing Science Institute
InformingScience.org

JITEResearch.org

Volume 18, 2019

FACTORS INFLUENCING THE SUBJECTIVE LEARNING EFFECTIVENESS OF SERIOUS GAMES

Emmanuel Fokides*	University of the Aegean, Rhodes, Greece	fokides@aegean.gr
Penelope Atsikpasi	University of the Aegean, Rhodes, Greece	pred17015@aegean.gr
Polyxeni Kaimara	Ionian University, Corfu, Greece	a16kaim@ionio.gr
Ioannis Deliyannis	Ionian University, Corfu, Greece	yiannis@ionio.gr

* Corresponding author

ABSTRACT

Aim/Purpose	This work examines which factors influence user views on the learning effectiveness of serious games. For that matter, a model was developed and tested.
Background	Although the impact of serious games on learning is their most widely examined aspect, research is spread thin across a large number of studies having little in common in terms of their settings, samples, and learning subjects. Also, there is a lack of consensus regarding which factors have an impact on their effectiveness. The most significant problem seems to be the fact that most assessment tools examined just a few factors.
Methodology	The initial model included eleven factors responsible for shaping the learning outcomes, belonging to four groups: (a) content, (b) technical features, (c) user state of mind, and (d) learning enabling features. All possible relationships between these factors and subjective learning effectiveness were examined. Data were collected using the Serious Games Evaluation Scale. The target group was 483 university students who played two serious games. The model was tested using covariance-based structural equation modeling.
Contribution	The study offers the prototype of a rather complex model, accurately explaining the intricate relationships between the substantial number of factors that were measured and their impact on user views regarding the subjective learning effectiveness of serious games.

Accepting Editor Man Fung (Kelvin) LO | Received: July 16, 2019 | Revised: September 27, 2019 | Accepted: September 30, 2019.

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

Findings	The final model fit statistics were very good, and 58.4% of the variance in subjective learning effectiveness was explained. The factor with the most significant impact was enjoyment, followed by subjective narration quality and realism. Quite interestingly, motivation did not have any effect on subjective learning effectiveness, while subjective feedback quality was not included as a construct in the final model. Moreover, the subjective ease of use and audiovisual fidelity had a minimal impact on other factors. Finally, the model proved to be invariant across genders and across the serious games that were used.
Recommendations for Practitioners	Serious game developers can use the model so as to decide on which factors to focus, depending on their needs. Educators and education policymakers can also benefit from the model's use, together with scales evaluating the quality of educational software. By assessing technical and content features and by using the model as a blueprint, they can envisage how enjoyable and motivating a serious game might be, as well as how it is going to impact user views regarding its learning effectiveness.
Recommendation for Researchers	Researchers can use the model in order to understand what shapes the learning experience of users when they play serious games. They can also use it for understanding the interactions between different the factors that come into play.
Future Research	Several alternative models have to be tested so as to develop a much simpler one which, at the same time, will have the capacity to adequately explain what users think of serious games. Several different target groups and serious games have to be examined in order to establish that the model is indeed invariant across a wide range of serious games genres and users. Finally, an interesting idea is to examine the relationship between subjective and objective learning effectiveness.

Keywords	enjoyment, learning effectiveness, serious games, structural equation modeling, subjective narration quality, subjective realism

INTRODUCTION

Abt (1970) was perhaps the first to coin the term "serious game" (SG), which pairs play and learning. He also stated that SGs' primary focus is not fun, although this does not necessarily mean that they are boring; the play activities are driven by carefully planned educational objectives. Thus, SGs can be viewed as (digital) games intended to impact cognition, skills, and behavior (Ravyse, Blignaut, Leendertz, & Woolner, 2017). Through the years, SGs have become a trend, corresponding to the zeitgeist of using playful learning together with technology. Education stakeholders increasingly recognize their value and SGs are expected to significantly affect all levels of education (Johnson et al., 2016). Indeed, their field of application includes a rather extensive array of learning domains and scenarios (e.g., Feng, González, Amor, Lovreglio, & Cabrera-Guerrero, 2018; Hersh & Leporini, 2018). Also, several studies reported encouraging learning outcomes (e.g., de Freitas, 2018; Erhel & Jamet, 2019; Lamb, Annetta, Firestone, & Etopio, 2018).

However, more than a few issues related to SGs are still not adequately addressed. How knowledge acquired while playing is transferred to real-life conditions is a characteristic one (Blumberg, Monte, Anthony, & Hashimoto, 2013). Probably the most significant issue is that of SG evaluation. It is true that their development is a complex process and experts from many fields are involved. The thorough assessment of the end-result of such an effort requires robust evaluation frameworks that take into consideration all (or at least as many as possible) salient factors. Yet, this task is almost impossi-

ble to carry out, as there are many different SG genres, covering diverse subjects, addressed to different groups of people. As a result, some concluded that the evaluation of SGs' functional components remains unclear (e.g., Alonso-Fernández, Pérez-Colado, Freire, Martínez-Ortiz, & Fernández-Manjón, 2018; Zhonggen, 2019) or that it is biased (e.g., Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012). Others declared that past studies failed to solve this problem (Shi & Shih, 2015) or supported the view that there is no established methodology for measuring their effectiveness (e.g., Hersh & Leporini, 2018; Serrano-Laguna, Manero, Freire, & Fernández-Manjón, 2018). Indeed, although there is a substantial amount of studies exploring this topic, the majority of them considered just a few factors when assessing SGs. What is more, several factors are ill-defined, causing confusion about how to measure them (Fokides, Atsikpasi, Kaimara, & Deliyannis, 2019). Nevertheless, what most studies have in common is their focus on the learning outcomes as a key determinant of SGs' success, measuring them with an assortment of techniques.

The originality of the work at hand lies in the fact that it deviated from the mainstream mentality regarding SG assessment. Although the main concern remained their learning effectiveness, the importance of user views on the matter was accentuated, given that it is vital to understand which features users consider important in rendering SGs effective learning tools. Thus, the predominant research question the current work sought to answer was "Which perceived factors are important in shaping their views for SG learning effectiveness?" Furthermore, another objective was to thoroughly examine the interactions of as many factors as possible. To achieve the above, a model was developed and tested that included a total of eleven factors theorized to affect the subjective learning effectiveness of SGs. The steps that were followed for the model's development are detailed in the coming sections.

LITERATURE REVIEW

Research on what makes SGs effective learning tools is not new. In fact, it is still the main concern of most involved in this field (All, Castellar, & Van Looy, 2015; Calderón & Ruiz, 2015; de Freitas, 2018; de Freitas & Ketelhut, 2014; Ravyse et al., 2017). There are several contributing factors explaining SG effectiveness. For example, Calderón and Ruiz (2015) listed eighteen (subjective) factors such as: aesthetics and design; social impact; interface; player efficacy, motivation, behavior, attitude, emotions, satisfaction, enjoyment, engagement, acceptance, and performance; playability, learnability, and understandability; usefulness; and educational aspects. The learning content and pedagogy together with storytelling/narrative, game mechanics, interface, and user experience draw Winn's (2009) attention. Others examined the importance of the scenario, learning-game integration, gameplay, interaction, feedback, challenge, fun, immersion, and game design (e.g., Faizan, Löffler, Heininger, Utesch, & Krcmar, 2019; Marsh, 2011).

Then again, the bulk of the research focused on specific (and few) factors. Some examined the impact of engagement and motivation (e.g., Hookham & Nesbitt, 2019; Huang, Huang, & Tschopp, 2010; Kiunsi & Ferwerda, 2019). Others assessed the role of the interface, workload, and usability (Thorpe, Nesbitt, & Eidels, 2019). For measuring game quality and effectiveness, Steiner et al. (2015) considered usability and enjoyment. Many emphasized the importance of one or more of the following factors: learning-game integration, gameplay, narration, feedback, interaction, enjoyment, scenario, immersion, and game design (e.g., Dobrovsky, Borghoff, & Hofmann, 2019; Khan & Webster, 2017; Marsh 2011; Muratet, Viallet, Torguet, & Jessel, 2009). From the above, one can easily understand that the relevant literature is rather fragmented in terms of (a) the SGs that were examined (e.g., learning content and genre), (b) the context in which they were examined (e.g., target groups and settings), and (c) the factors that were considered. What is more, research taking into consideration many salient factors is scarce. Clearly, more research is needed for establishing which features play a significant role in shaping SG learning effectiveness (Hersh & Leporini, 2018) so as to alleviate the ongoing debates about the role of certain factors (Ravyse et al., 2017).

Unfortunately, SGs assessment is not an easy task. That is because they are two-sided applications; on one hand, they are games and, on the other, they are educational applications, not perfectly fitting to either category. Thus, existing evaluation frameworks developed explicitly for games or for digital learning environments may fall short when it comes to SGs. An additional barrier to the development of universal assessment frameworks is the differences between game genres. For example, a tool developed for assessing the fantasy state in role-playing games can hardly be used for assessing puzzle games (Choi, Huang, Jeffrey, & Baek, 2013). Mitgutsch and Alvarado (2012) noted that in order SGs to be thoroughly validated, the player mindset has to be considered as well. In essence, we are in need of an interdisciplinary approach; assessment models and tools have to take into account SGs pedagogical aspects (e.g., the learning material and goals) and their leisure/gaming aspects (e.g., gameplay and fun), as both influence gamers' experiences and views for their learning effectiveness.

Such models have already been suggested and/or examined. For example, de Freitas and Oliver (2006) in their four-dimensional model focused on the context, the learners' attributes, the games' internal representation (e.g., immersion, interactivity, and fidelity), and on the learning processes. Mayer et al.'s (2014) model comprised of pre-, in-, and post-game constructs. Players' behavioral intentions, attitudes, and skills were considered in the pre-game construct. The in-game construct included factors related to player experiences (e.g., presence, flow, and immersion). The post-game construct also included factors related to player gaming experiences (e.g., engagement, fun, and satisfaction), as well as learning. A four-dimensional model was also proposed by Galván-Pérez, Ouariachi, Pozo-Llorente, and Gutiérrez-Pérez (2018): (a) the narrative dimension (storyline and character depiction), (b) the contents dimension (terminology, concepts, and definitions), (c) the gameplay dimension (game mechanics, feedback, interactivity, and reward system), and (d) the didactic dimension (competences, need for previous knowledge, interdisciplinarity, and learning curve). Ravyse et al. (2017) identified five themes that should be intertwined with the pedagogical content: (a) realism, (b) artificial intelligence/adaptivity), (c) backstory, (d) feedback, and (e) interaction. The Serious Game Design Assessment Framework proposed by Mitgutsch and Alvarado (2012) threated SGs as purpose-based game systems. As such, purpose has to be reflected in their components, namely, the fiction/narrative, framing, content, aesthetics/graphics, and game-mechanics. Finally, Annetta and Bronack (2011) offered a complex model encompassing thirteen factors: prologue, communication, interactive feedback, immersion, levels for practice/tutorial, identity, pleasurable frustration, rules, complexity, manipulation, informed learning, and pedagogical/reading effectiveness. What became evident from the literature review was that very few studies proposed models in which specific relationships between factors were suggested, tested, and quantified (e.g., Hamari et al., 2016; Yusoff, Crowder, & Gilbert, 2010).

To summarize, the literature review brought into light several research gaps and uncertainties, worth further examination. First, although SG impact on learning is their most widely examined aspect, research is spread thin across a large number of studies having little in common. Second, there is a lack of consensus regarding which factors have an impact on SG effectiveness. Third, there is a need for thorough assessment methods and models, as most take into account just a few factors. Finally, few studies tested and quantified the relationships between factors theorized to affect SG effectiveness.

FACTORS AFFECTING THE (SUBJECTIVE) LEARNING EFFECTIVENESS OF SERIOUS GAMES

Given that the study sought to develop a model for examining if and how several factors affect user views for SGs' learning effectiveness, the first task was to probe further into the literature, in order to uncover which factors are commonly used in this type of research. Indeed, eleven such factors were located, that can be grouped into four categories. As one might notice, all factors were treated as subjective ones, as they refer to user views, emotions, sensations, and feelings.

Factors Related to Content

- *Subjective feedback quality.* Feedback, as a result of user actions, is a key feature of all educational applications and research suggested that can it can improve SG effectiveness (Sušnik et al., 2018; Wilson, Broadbent, McGrath, & Prescott, 2017). Either as unobtrusive support when needed or as a pedagogical intervention, it is highly appreciated by users (Ke & Abras, 2013; Serrano-Laguna, Torrente, Moreno-Ger, & Fernández-Manjón, 2014). That is because it grasps their attention, forces them to focus on what is relevant and important (Ketamo & Kiili, 2010), allows them to reflect on what they have learned or on the outcomes of their actions (Cheng & Annetta, 2012), revise their strategies, and have the sense of progress (Cheng, Lin, & She, 2015).

- *Subjective quality of the learning material.* The learning material consists of all data, information, and facts offered and used within an SG. How users perceive its quality is quite important, especially when the learning objectives are hard to achieve or when the learning subject is technical or even boring. In such cases, how SGs present the material to users has a significant effect on learning (Michael & Chen, 2005). Moreover, it is important the learning content not to be seen as a juxtaposition of information and game layers (Mortara et al., 2014); it has to be flawlessly integrated into the SG (Khenissi, Essalmi, & Jemni, 2015).

- *Subjective learning goals clarity.* In SGs, the goals are twofold, gaming and learning. Both have to be well-crafted, clearly defined, and specific (through measurable achievements) (Bellotti, Kapralos, Lee, Moreno-Ger, & Berta, 2013). Additionally, gameplay and learning goals have to be related; otherwise, learning is removed from context (Barnes, Richter, Powell, Chaffin, & Godwin, 2007). When the above hold true, players have an enjoyable experience and are motivated (Shi & Shih, 2015). Goal clarity was also considered as a contributing factor in players' flow state (Fu, Su, & Yu, 2009). Then again, differences were noted between explicitly and non-explicitly defined goals; the latter although they affected memorization, did not have an impact on comprehension (Erhel & Jamet, 2019).

- *Subjective narration quality.* The plot/narrative/storyline is gradually uncovered to players, providing them with declarative knowledge (Kiili, 2005a). The learning content has to be stealthily/seamlessly integrated into the narrative in order players not to lose interest (Couceiro, Papastergiou, Kordaki, & Veloso, 2013; Westera, 2019) and credible learning to be achieved (Ke & Abras, 2013). However, others reported that an engaging narrative can distract users, removing them from meaningful cognitive activities (Mayer, Griffith, Naftaly, & Rothman, 2008).

Factors Related to Technical Features

- *Subjective ease of use/usability/playability.* In general, there are no established (and coherent) definitions of the terms playability and usability. Playability can be conceptualized as "the degree to which a game is fun to play and is usable, with an emphasis on the interaction style and plot-quality of the game; the quality of gameplay" (Usability-First, 2009). Playability is affected by a number of other factors such as storyline, interaction, realism, and audiovisual quality. Usability, on the other hand, describes how easily a user can understand a game and learn how to control it (Pinelle, Wong, & Stach, 2008). The features of usability include satisfaction, the interface, and ease of use (Moizer et al., 2019). Some considered usability as a subset of playability (Pinelle et al., 2008), others treated both terms as synonyms (Sánchez, Vela, Simarro, & Padilla-Zea, 2012). In any case, trying to measure both is a rather challenging task, as different definitions suggest that different factors should be considered (Sánchez et al., 2012), many of which are already included in this study. Thus, it was decided to examine the ease of use, a factor appearing in both terms. Ease of use is employed in almost all instruments assessing digital tools/applications and describes the extent users believe that

they can effortlessly use the given tool (Davis, Bagozzi, & Warshaw, 1989). Moreover, ease of use can indirectly measure the users' play literacy, a competence commonly overlooked (Mitgutsch & Alvarado, 2012).

- *Subjective audiovisual fidelity/aesthetics.* Understandably, users demand SGs with better graphics (Couceiro et al., 2013). Indeed, experienced gamers find low-fidelity SGs unengaging (de Freitas, Rebolledo-Mendez, Liarokapis, Magoulas, & Poulovassili, 2009). Advanced audiovisual features do have a place in SGs (Petridis, Dunwell, de Freitas, & Panzoli, 2010), especially to simulations, where high levels of fidelity are needed. The visual effects make SGs attractive and motivating (Huang, Johnson, & Han, 2013) and contribute to the understanding of abstract concepts (Cai, Lu, Zheng, & Li, 2006). On the other hand, Jarvis and de Freitas (2009) suggested that fidelity's level must be mapped onto the learning objectives. What is more, the role of music and sound effects is mostly unexplored, although some advocated that they are strong learning facilitators (Yoho, 2011).
- *Subjective SG realism.* While audiovisual fidelity and realism have a lot in common, the latter is not limited only to technical features but has functional and psychological dimensions as well (Ravyse et al., 2017). It also encompasses game-player interactions (Mortara et al., 2014). High levels of realism certainly contribute to SG quality, but its impact on learning is dubious. Studies suggested that realistic environments fostered users' understanding of the learning content (e.g., Byun & Loh, 2015; Chittaro & Buttussi, 2015; Verpoorten, Castaigne, Westera, & Specht, 2014). Others declared that the learning outcomes were not affected by realism (Norman, Dore, & Grierson, 2012; Vogel et al., 2006). Kiili (2005b) stated that users do not require realistic games as long as the flow state is achieved. Ke and Abras (2013) demonstrated that realism distracted students from their learning tasks. On top of that, Cook et al. (2013), in their review on instructional design for simulation-based games, rejected realism as a variable, because it did not reach acceptable reliability levels.

Factors Related to Users' State of Mind

- *Presence/immersion.* Amongst the most confusing factors (in terms of how they are defined) commonly used in research related to computer applications, SGs included, are presence and immersion (Fokides & Atsikpasi, 2018). Indeed, both are used for describing similar psychological states during playing (Brockmyer et al., 2009). In presence, users perceive the virtual objects as being real ones, in both sensory and nonsensory ways (Ivory & Kalyanaraman, 2007). On the other hand, immersion describes the feeling that players actually exist in the game. In many cases these terms were used interchangeably and, quite interestingly, they were examined using similarly (or even identically) worded items in questionnaires (Nilsson, Nordahl, & Serafin, 2016). Given the above, it was decided the term "presence" to encapsulate both states. Although it is often implied that presence/immersion has an impact on learning, the number of studies systematically investigating this relationship (in the context of SGs and/or educational games) is limited (e.g. Fassbender, Richards, Bilgin, Thompson, & Heiden, 2012), while their results were controversial. For example, some concluded that, when users were highly competent in playing games, immersion had a positive impact (e.g., Cheng, She, & Annetta, 2015). Then again, Hamari et al. (2016) did not find any effect.
- *Enjoyment.* Enjoyment is used in many SG evaluation frameworks (e.g., Hookham & Nesbitt, 2019; Steiner et al., 2015). Yet, its role is controversial and our understanding for its impact is not that well-developed. That is because some supported the view that it is totally ignored in research (Vorderer, Klimmt & Ritterfield, 2004), a number of studies reported that it had a direct impact on learning (e.g., Breuer & Bente, 2010; Connolly et al., 2012; Ke, 2011), while others found that it correlated only with motivation, concluding that instruction, the learning tasks, and support might be more decisive factors (Iten & Petko, 2016). Moreover, Prensky

(2007) advised caution; if a certain point is exceeded, fun might increase students' cognitive load, distract them, and act as a detrimental factor for the learning outcomes.

Factors Related to Learning Enabling Features

- *Subjective relevance to personal interests.* Interest is a crucial factor. It directs motivation, excites the desire to be engaged in a learning activity, and directs attention. All the above are all related to learning (Schiefele, Krapp, & Winteler, 1992). An SG, even if it is motivational and enjoyable, it does not necessarily mean that it is interesting or engaging (ImligIten & Petko, 2018). Both the entertainment it provides and the content have to be relevant to one's personal interests. In line with the above, Przybylski, Rigby, and Ryan (2010) suggested that the effects of games are based on their potential to satisfy, among other basic psychological needs, personal interests and relatedness. Moreover, other researchers, in the context of the Universal Design for Learning educational framework, suggested that SGs, building on the interests of students, allow them to connect real-world scenarios with school content, thus responding to the question "Why do I need to know this?" (Annetta, 2010; Driscoll, 2002).
- *Motivation.* Garris, Ahlers, and Driskell (2002) argued that the motivational appeal of SG is what drives users to invest time and effort in using them, because the activity, by itself, is rewarding. Thus, motivation can be viewed as one of the primary reasons for playing SGs (Kaimara & Deliyannis, 2019). Researchers have identified several SG features as having an impact on motivation, such as autonomy, competence (Przybylski et al., 2010), the content, and the game-mechanics (Habgood & Ainsworth, 2011). However, others concluded that the motivational appeal of SGs does not differ that much from other instructional methods, because users' autonomy is limited by the instructor and the curriculum (Wouters, Van Nimwegen, Van Oostendorp, & Van Der Spek, 2013).

Research Model and Hypotheses Formation

The above literature review provided a wealth of ideas and directions regarding the factors that affect subjective user views for SG learning effectiveness as well as on how these factors interplay with each other. Then again, the plethora of research gaps and uncertainties we identified (as presented in the preceding sections), led to the decision to simultaneously examine all the factors and to test all possible relationships. Consequently, the research model presented in Figure 1 was developed, and a series of research hypotheses were formulated, as presented below:

- H1a-e. Subjective feedback quality (Feed) has an impact on (a) presence, (b) enjoyment, (c) subjective relevance to personal interests, (d) motivation, and (e) on subjective learning effectiveness (SLE).
- H2a-e. Subjective learning goals clarity (Goal) has an impact on (a) presence, (b) enjoyment, (c) subjective relevance to personal interests, (d) motivation, and (e) on SLE.
- H3a-e. Subjective narration quality (Nar) has an impact on (a) presence, (b) enjoyment, (c) subjective relevance to personal interests, (d) motivation, and (e) on SLE.
- H4a-e. Subjective quality of the learning material (QLM) has an impact on (a) presence, (b) enjoyment, (c) subjective relevance to personal interests, motivation, and (d) on SLE.
- H5a-e. Subjective realism (Real) has an impact on (a) presence, (b) enjoyment, (c) subjective relevance to personal interests, (d) motivation, and (e) on SLE.
- H6a-e. Subjective ease of use (SEU) has an impact on (a) presence, (b) enjoyment, (c) subjective relevance to personal interests, (d) motivation, and (e) on SLE.

- H7a-e. Subjective audiovisual fidelity (AV) has an impact on (a) presence, (b) enjoyment, (c) subjective relevance to personal interests, (d) motivation, and (e) on SLE.
- H8a-d. Presence (Pre) has an impact on (a) enjoyment, (b) subjective relevance to personal interests, (c) motivation, and (d) on SLE.
- H9a-c. Enjoyment (Enj) has an impact on (a) subjective relevance to personal interests, (b) motivation, and (c) on SLE.
- H10a-b. Subjective relevance to personal interests (RPI) has an impact on (a) motivation and (b) on SLE.
- H11. Motivation (Mot) has an impact on SLE.

Controls are variables that do not drive a study's theory but have to be taken into account as they might have an impact on the results. Given that SGs are part educational software and part digital games, two variables that could have a confounding effect on how users perceive the learning effectiveness of SGs can be recognized, namely, their game playing and ICT competencies. As a result, the following hypothesis was tested:

- H12a-b. Participant views for the subjective learning effectiveness of SGs are affected by their (a) game playing competencies and (b) ICT competencies.

Multi-group comparisons are a form of moderation for determining if the variable relationships hypothesized in a model, differ depending on the value of a moderator/grouping variable (Kline, 2015). Gender and the SG genre/type were included as grouping variables. Although some researchers suggested that both genders regard SGs as effective and motivating (e.g., Connolly et al., 2012; Hainey, Connolly, Boyle, Wilson & Razak, 2016; Mayer, 2019), the literature examining gender differences in the context of SGs is not extensive. In addition, since in the study two SGs were planned to be used, it was considered important to examine whether this had an impact on the model. Consequently, the following research hypothesis was studied:

- H13a-d. The model relationships are moderated by (a) gender and (b) the genre/type of SGs.

SLE was treated as the dependent variable (DV), while all the other factors were the independent variables (IVs). Feed, Goal, Nar, QLM, Real, SEU, and AV were treated as exogenous variables, while Pre, Enj, Mot, and RPI were treated as endogenous ones. Given that all the above hypotheses were set for examining direct effects, it was considered interesting to examine the indirect effects the exogenous variables might have on the DV, through the endogenous variables. Thus, the following research hypothesis was adopted:

- H14a-d. The endogenous variables (a) presence, (b) enjoyment, (c) subjective relevance to personal interests, and (d) motivation, mediate the effects of the exogenous variables on the DV (subjective learning effectiveness).

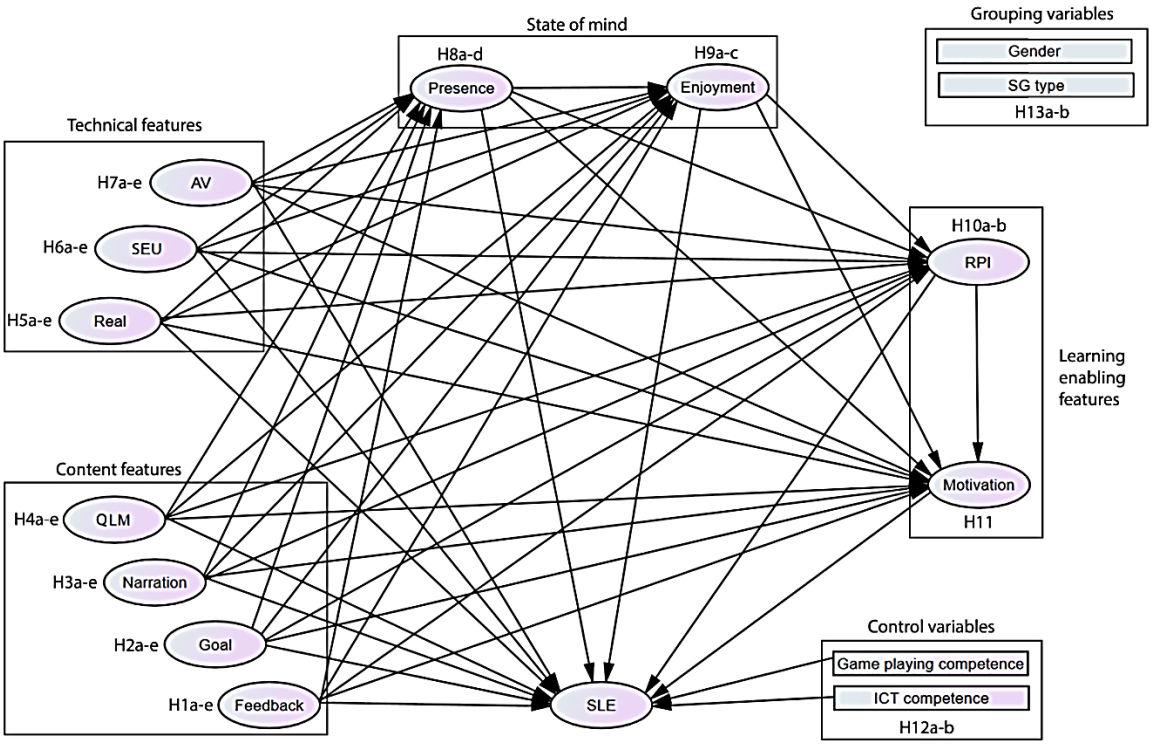

Figure 1. The initial research model

METHOD

As already stated, the objective was to develop a model able to explain the users' views regarding SG learning effectiveness. Having developed the initial research model, the next phase was to test it. For collecting the necessary data, a project was designed and implemented. It has to be noted that covariance-based structural equation modeling (SEM) was employed for examining factor relationships. SEM is a form of multivariate analysis combining factorial and multiple regression analysis. As such, it can be used for exploring, explaining, and quantifying the structural relationship between latent constructs and measured variables. It is routinely used in behavioral and social sciences for examining the causal relations between factors. Given SEM's ability to test complex models and given that in SGs many factors are involved, this research path was considered as being the most efficient in exploring SG subjective learning effectiveness. Further details for the project are presented in the sections to follow.

MATERIALS

There is a large number of SGs, having diverse learning subjects/objectives, belonging to different genres, and targeting different audiences. Thus, an issue that had to be resolved was what SGs to select as the study's material. On the other hand, the purpose was to develop a model explaining users' learning experience when playing SGs (in general) and not to evaluate a specific SG. Thus, the quality, learning subject, and genre were considered irrelevant. What it was considered important was to select SGs in which the eleven factors/features discussed in the preceding section to be present, so as to facilitate user views and thoughts. Following this line of thinking and after reviewing several commercially available SGs, two games developed by Triseum (https://triseum.com) were selected. Although they are fundamentally different, in terms of gameplay, environment story and learning domain, both can be considered as typical SGs, addressed to university/college students. Moreover, they received several awards and they are well accepted by their intended audience.

"ARTé Mecenas," is a 2D, turn-based SG, supporting Art History courses. Users assume the role of a fictitious member of the powerful Medici family during the Italian Renaissance. Through the course of the game, they have to put into practice their decision-making skills, in order to balance their relations with other merchants, states, fractions, and the Catholic Church. The objective is not just to build a financial empire but also to contribute to the creation of Renaissance's renowned artworks and monuments. Though the "game" objective is players to shape the course of art history and become the head of the Medici family, the "serious" objective is players to understand the interconnectedness of economy and art (i.e., art patronage) and to learn historical facts as well as details for artworks. "Variant: Limits" is a 3D open-world, role-playing game, supporting Math courses, specifically, ones related to calculus concepts. Players assume the role of Equa, a young girl suffering from amnesia, who tries to save her planet from a cataclysmic solar storm. Players have to solve increasingly complex puzzles which are based on calculus theory and principles. The topics covered are: (a) finite limits (e.g., one-sided limits), (b) continuity (continuity at a point definition, combined functions' continuity, and the intermediate value theorem), and (c) infinite limits (horizontal and vertical asymptotes).

One might argue that the inclusion of two SGs poses a serious threat to the study's validity and conclusions. Then again, it can be counter-argued that it was imperative to include two (or even more) SGs. First, a limitation that almost all studies acknowledged, regardless of their nature, was the generalizability of their results due to the fact that they tested one SG (of a given type/genre, learning content, etc.). Second, the objective was to develop a model that could be applied in a variety of situations. This is the main reason for including hypothesis H13a-b (Gender and the SG genre/type moderate the model's relationships). it was expected the results of multi-group comparisons to provide evidence for the model's invariability or otherwise. Thus, depending on the outcome, the model could be accepted if the differences between the two SGs were minor, modified if the differences were moderate, or even be rejected altogether if the differences were major. Given that the study at hand is part of a project in progress and given its highly exploratory nature, it was concluded that, at this stage, two SGs will suffice. The matter will be further elaborated on in the "Discussion" section.

Participants

The second major issue that had to be resolved was the target group(s). Several alternative ideas/solutions were considered. For example, one idea was to select university students studying arts and mathematics. This path was rejected, as it was possible for participants to have already studied similar (or exactly the same) subjects like the ones included in the SGs. Thus, there was a chance to consider them boring and their responses to the questionnaire to be unengaged. Another thought was to administer the SGs to students regardless of their scientific background. Although this would have allowed a larger sample size, again this idea was rejected, as participants might also be unengaged. Considering the pros and cons of each alternative, it was decided the sample to consist of students coming from the Department of Audio and Visual Arts, Ionian University and the Department of Primary Education, University of the Aegean, both in Greece. These students were considered as being the ideal target group because the curriculum of both departments includes courses: (a) related (but not the same) to the SGs' learning content and (b) related to the design of educational games, thus, students have a basic grasp of the principles of (educational) game design.

A recruiting invitation was posted to the Departments' Facebook groups, addressed to students interested to voluntarily participate in the study. The candidates were informed that they were going to play one or two SGs and complete a short questionnaire. An outline of both SGs was also included (e.g., storyline and learning content), so as to provide additional motivation to participants. It was explicitly stated that participant anonymity was guaranteed and that, by completing the questionnaire, their participation consent was deemed to have been given. In addition, participants were informed that the duration each game was played was going to be recorded. Through the above process, a total of 520 students were enrolled.

Instrument

For data collection purposes, the Serious Games Evaluation Scale (SGES) was used (Fokides et al., 2019) (see Appendix). Its reliability and factorial structure were tested and confirmed and, as its name implies, it was developed for evaluating serious games (although it can be used for evaluating other digital educational material as well). Besides demographic questions (i.e., age, gender, ICT and game-playing competences), it consists of fifty-three statements measuring a total of twelve subjective factors, the same as the ones presented in section "Factors affecting the (subjective) learning effectiveness of serious games", plus subjective learning effectiveness. Respondents indicate their strength of agreement in a five-point Likert-type scale (anchored at strongly disagree and strongly agree).

Procedure

The participating students were gathered to the Universities' computer labs. Further details regarding the SGs were provided during a short briefing. Moreover, it was clarified that participants were free to choose which SG to play depending on their interests (in this respect, they were free to play either or both SGs). On the other hand, they had to: (a) play each for a minimum of two hours and (b) to complete at least the first two levels, excluding the introductory/tutorial one. Following that, participants were split into groups of ten (ten for each lab and twenty in total for both departments) and a timetable for participation was assembled. The rationale behind this decision was labs not to be overcrowded and to provide participants with some privacy and a more relaxed environment. After checking that each participant met the requirements described above, he/she was provided with the questionnaire's link because it was available only online.

Results Analysis

The result analysis was a multi-stage process. First, the data were screened for eliminating the unengaged responses, which might affect the quality of the results. A Confirmatory Factor Analysis (CFA) was also conducted, for establishing the instrument's validity. Finally, for testing the model, SEM was conducted.

Data Screening and Descriptive Statistics

All responses were imputed into SPSS 25 for an initial set of checks and analyses. There were no missing data, as all questions were mandatory. The presence of outliers and unengaged responses was checked by examining responses' standard deviations and a total of 37 cases with values less than 0.50 were deleted. The valid questionnaires left after the case screening were 483, coming from 167 males and 316 females around 23 years old (M = 22.96, SD = 5.89). This imbalance in gender distribution was anticipated since it reflected the actual gender distribution of students studying at both Departments. Participants' skills in using computer applications were above the mid-point (M = 3.57, SD = 0.86), while their expertise in playing games was average (M = 3.08, SD = 1.11). In total, the 2D SG was played 302 times and the 3D SG was played 181 times. Scores were obtained by allocating numerical values to participant responses: "strongly agree" scored 5, "agree" scored 4; "neutral" scored 3; "disagree" scored 2 and "strongly disagree" scored 1. Descriptive statistics for the study's factors are presented in Table 1. Construct consistency, as assessed using Cronbach's alpha, was optimal since in all cases the values exceeded the .80 threshold (ranging between .871 and .948 for the constructs, while the scale's overall score was .961) (DeVellis, 2016).

Table 1. Descriptive statistics for factors

Factor	*M*	*SD*	Skewness	Kurtosis	Cronbach's *a*	Factor	*M*	*SD*	Skewness	Kurtosis	Cronbach's *a*
Pre	2.52	0.95	0.499	-0.576	.871	Goal	3.39	0.92	-0.304	-0.372	.888
Enj	3.54	0.97	-0.626	-0.253	.948	Feed	3.82	0.65	-0.385	-0.370	.879
SLE	3.19	0.96	-0.330	-0.675	.943	SEU	3.69	0.89	-0.715	-0.064	.932
Real	2.43	0.95	.0344	-0.619	.903	QLM	3.38	0.90	-0.468	-0.128	.910
Nar	3.33	0.96	-0.222	-0.523	.926	Mot	3.75	0.93	-0.907	0.560	.881
AV	3.53	0.91	-0.453	-0.568	.943	RPI	2.67	0.92	0.167	-0.502	.877

Preliminary Data Analysis

Although SGES has already been validated, it was deemed necessary to re-assess its validity and reliability by conducting CFA. The factorial structure was imputed into AMOS 25 and Maximum Likelihood was the estimation method of choice, as the data deviations from normality were moderate (skewness and kurtosis < |1|, see Table 1). For that matter, the negative effect on the parameter estimates was expected to be negligible (Brown, 2014; Matsunaga, 2010; Ryu, 2011). For checking convergent validity, the average variance extracted (AVE) was used (Table 2). Excluding the AVE of Presence, all AVEs were above the 0.70 level as suggested by Hu and Bentler (1999). As long as all the remaining indices were satisfactory, Presence's lower AVE was considered as an acceptable deviation from the recommended values. Reliability was evident, given that all critical ratios were above the 0.70 threshold (Hancock 2001). For assessing discriminant validity, the square root of AVE for any given factor was compared with the correlations between this factor and any other factor (Hu & Bentler, 1999) (Table 2). It was found that the shared variance between a factor and any other factor was less than the variance it shared with its measures. Thus, it was concluded that the discriminant validity was satisfactory.

Table 2. Convergent and discriminant validity

	CR	AVE	AV	SLE	SEU	Enj	QLM	Pre	Real	Mot	Nar	Goal	RPI	Feed
AV	0.943	0.704	**0.839**											
SLE	0.944	0.736	0.393	**0.858**										
SEU	0.932	0.703	0.252	0.432	**0.838**									
Enj	0.948	0.753	0.490	0.704	0.559	**0.868**								
QLM	0.911	0.718	0.099	0.330	0.533	0.425	**0.848**							
Pre	0.875	0.639	0.281	0.361	0.115	0.399	0.020	**0.799**						
Real	0.904	0.704	0.391	0.451	0.069	0.405	0.027	0.529	**0.839**					
Mot	0.883	0.716	0.237	0.463	0.406	0.647	0.464	0.175	0.195	**0.846**				
Nar	0.927	0.759	0.484	0.630	0.505	0.762	0.326	0.424	0.382	0.525	**0.871**			
Goal	0.889	0.729	0.297	0.492	0.474	0.496	0.327	0.203	0.255	0.263	0.502	**0.854**		
RPI	0.877	0.704	0.245	0.562	0.361	0.581	0.352	0.331	0.369	0.399	0.527	0.374	**0.839**	
Feed	0.880	0.709	0.364	0.383	0.546	0.476	0.214	0.116	0.130	0.290	0.481	0.493	0.332	**0.842**

Notes. AVE: Average Variance Extracted; CR: Critical ratio; diagonal: square root of AVE extracted from observed variables; off-diagonal: correlations between constructs

Given that participants' perceptions/views were measured at a single point in time, there was some concerned about the presence of Common Method Variance (CMV) (or Common Method Bias), that is the variance caused by the measurement method itself (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). A Common Latent Factor was added and a test of equal specific bias was run, as suggested by Gaskin and Lim (2017). The zero-constrained chi-square difference test resulted in a non-significant result; thus, it was concluded that CMV was not an issue.

Because SEM was going to be employed for testing the study's hypotheses, it was examined whether the data were fit for this type of testing. For conducting SEM, Kline (2015) suggested a minimum of 200 cases and, preferably, 5 to 10 cases per variable. Other researchers recommended a minimum of 450 cases for complex mediation models, as was our model (Wolf, Harrington, Clark, & Miller, 2013). The study's 483 cases and 53 variables satisfied Wolf's et al.'s (2013) guidelines and were very close to satisfying Kline's stricter rule for a ratio of 10 cases per variable (9.83 cases per variable). The multivariate assumptions were satisfied because (a) there were no influential cases and outliers (Cook's distance in all cases < 0.07, while the maximum recommended value is 1.00), and (b) multicollinearity was not a concern (the highest Variable Inflation Factor value that was observed was 2.94, while the recommended maximum is 3.00 and the lowest Tolerance value that was observed was .34, while the recommended minimum is .10) (O'Brien, 2007). Finally, for conducting SEM, it is recommended to use control variables. As already mentioned in the "Research model and hypotheses formation" section, participants' ICT and gaming competencies acted as such.

For the model's initial fit assessment, the following indices were used: (a) Comparative Fit Index (CFI), (b) Root Mean Square Error of Approximation (RMSEA), (c) Standardized Root Mean Square Residual (SRMR), and (d) chi-square/df ratio. CFI values exceeding .950, RMSEA values less than .060, and SRMR values less than .080 indicate excellent fit (Hu & Bentler, 1999; McDonald & Ho, 2002). It has to be noted that the chi-square test statistic for model assessment was not used, as it tends to indicate statistically significant differences for sample sizes exceeding 200 cases (Hu & Bentler, 1999). Instead, the recommended chi-square/df ratio was used, with acceptable values ranging between 1 and 3 (Hu & Bentler, 1999). The results revealed that all the fit indices were excellent (χ^2/df = 1.672, SRMR = .045, RMSEA = .037, CFI = .961).

Structural Equation Modeling

For selecting the optimal model, all the direct paths of the initial model were made optional. The resulting hierarchy of models was then examined using AMOS's Specification Search Facility. The model with the smallest BCC_0 value (BCC_0 = 0.00) was selected as the final model (Burnham & Anderson, 1998). The fit statistics for this model remained excellent (χ^2/df = 1.649, SRMR = .035, RMSEA = .037, CFI = .960). Table 3 and Figure 2 present the direct effects of the final model and the resulting confirmation or rejection of hypotheses H1 to H12. The IVs explained 58.4% of the variance in the study's DV (subjective learning effectiveness) (R^2 = .584). Coming to the endogenous variables, the corresponding R^2s were .679 for enjoyment, .349 for presence, .405 for subjective relevance to personal interests, and .488 for motivation.

Table 3. Direct effects and hypotheses testing results

Hypothesis	Path	Path coefficient (β)	*t*-value	*p*	Result
H1a	Feed → Enj	.067	1.636	.102	rejected
H1b	Feed → Pre	-.078	-1.550	.121	rejected
H1c	Feed → Mot	Deleted during the specification search			rejected
H1d	Feed → RPI	.082	1.594	.111	rejected
H1e	Feed → SLE	Deleted during the specification search			rejected
H2a	Goal → Enj	Deleted during the specification search			rejected
H2b	Goal → Pre	Deleted during the specification search			rejected
H2c	Goal → Mot	.131	2.719	.007	supported
H2d	Goal → RPI	Deleted during the specification search			rejected
H2e	Goal → SLE	.154	3.690	< .001	supported
H3a	Nar → Enj	.464	9.812	< .001	supported
H3b	Nar → Pre	.330	5.877	< .001	supported
H3c	Nar → Mot	.155	2.239	.025	supported
H3d	Nar → RPI	.177	2.481	.013	supported
H3e	Nar → SLE	.115	1.976	.048	supported
H4a	QLM → Enj	.160	4.250	< .001	supported
H4b	QLM → Pre	-.081	-1.799	.072	rejected
H4c	QLM → Mot	.224	4.835	< .001	supported
H4d	QLM → RPI	.144	3.030	.002	supported
H4e	QLM → SLE	Deleted during the specification search			rejected
H5a	Real → Enj	.117	3.015	.003	supported
H5b	Real → Pre	.416	8.765	< .001	supported
H5c	Real → Mot	Deleted during specification search			rejected
H5d	Real → RPI	.207	4.227	< .001	supported
H5e	Real → SLE	.144	3.797	< .001	supported
H6a	SEU → Enj	.157	3.450	< .001	supported

Hypothesis	Path	Path coefficient (β)	t-value	p	Result
H6b	SEU → Pre	Deleted during the specification search			rejected
H6c	SEU → Mot	Deleted during the specification search			rejected
H6d	SEU → RPI	Deleted during the specification search			rejected
H6e	SEU → SLE	Deleted during the specification search			rejected
H7a	AV → Enj	.118	3.250	.001	supported
H7b	AV → Pre	Deleted during the specification search			rejected
H7c	AV → Mot	-.077	-1.689	.091	rejected
H7d	AV → RPI	.124	2.456	.014	supported
H7e	AV → SLE	Deleted during the specification search			rejected
H8a	Pre → Enj	.078	2.022	.043	supported
H8b	Pre → Mot	-.073	-1.616	.106	rejected
H8c	Pre → RPI	Deleted during the specification search			rejected
H8d	Pre → SLE	Deleted during the specification search			rejected
H9a	Enj → Mot	.565	7.664	< .001	supported
H9b	Enj → RPI	.323	4.226	< .001	supported
H9c	Enj → SLE	.386	6.424	< .001	supported
H10a	RPI → Mot	Deleted during the specification search			rejected
H10b	RPI → SLE	.183	4.038	< .001	supported
H11	Mot → SLE	Deleted during the specification search			rejected
H12a	Games comp → SLE	-.065	-1.984	.047	supported
H12b	ICT comp → SLE	Deleted during the specification search			rejected

Note. The highlighted rows indicate statistically significant paths

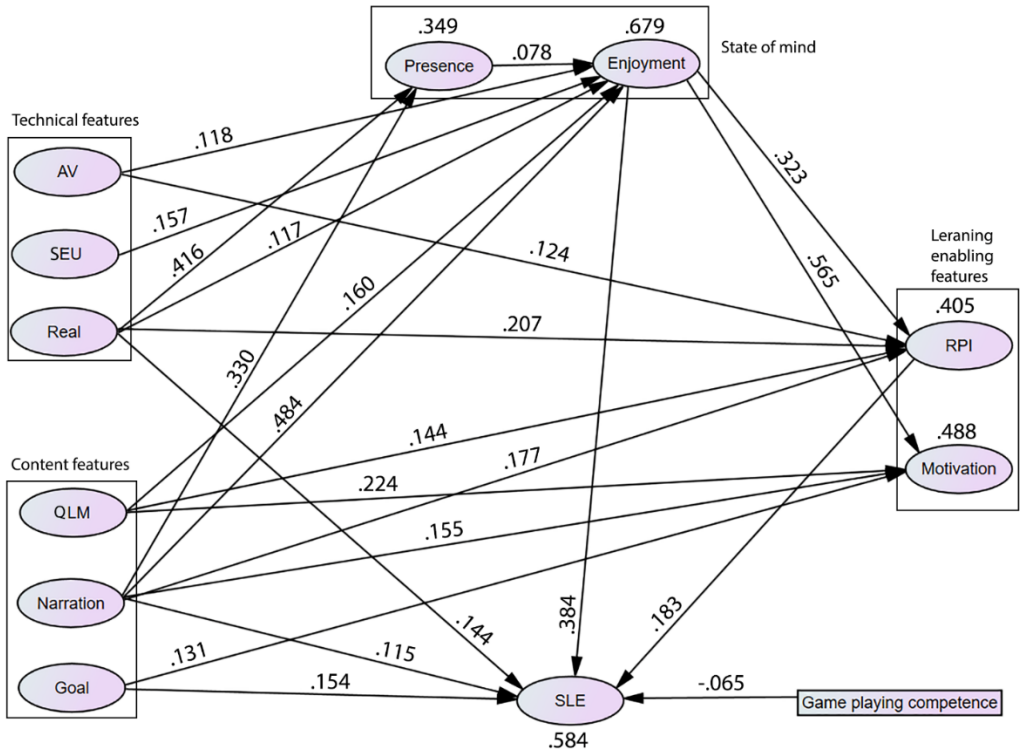

Figure 2. The final model (non-significant factors and paths were omitted for clearance of presentation)

Multi-group Testing

For examining H13a-b, multi-group comparisons were conducted, following the chi-square difference approach (Byrne, 2016; Kline, 2015). Gender, as well as the SG type, were used as moderators/grouping variables. It was found that the model did not differ in any of the above (gender: CMIN = 86.763, p = .085; SG type: CMIN= 82.834, p = .095). As a result, this hypothesis was rejected; the model's relationships remained unchanged among genders and among the two SGs that were used.

Mediation

For the examination of the mediation effects of the endogenous variables (H14a-d), Hayes's (2017) bootstrapping technique was used. It was found that: (a) enjoyment mediated the indirect effects of subjective narration quality, subjective quality of the learning material, subjective realism, subjective ease of use, subjective audiovisual fidelity, and presence and (b) subjective relevance to personal interests mediated the indirect effects of subjective narration quality, subjective quality of the learning material, subjective realism, subjective audiovisual fidelity, and enjoyment (Table 4).

Table 4. Mediation results

Hypothesis	Path	Estimate	p	Mediation effect	Direct effect
H14a	Pre did not have a direct effect on SLE; therefore, it could not act as a mediator				
H14b	Nar → Enj → SLE	.175	.001	yes	yes
	QLM → Enj → SLE	.064	.001	yes	no
	Real → Enj → SLE	.043	.002	yes	yes
	SEU → Enj → SLE	.065	.005	yes	no
	AV → Enj → SLE	.044	.001	yes	no
	Pre → Enj → SLE	.028	.058	no	no
H14c	Nar → RPI → SLE	.032	.010	yes	yes
	QLM → RPI → SLE	.027	.004	yes	no
	Real → RPI → SLE	.036	< .001	yes	yes
	AV → RPI → SLE	.022	.011	yes	no
	Enj → RPI → SLE	.060	< .001	yes	yes
H14d	Mot did not have a direct effect on SLE; therefore, it could not act as a mediator				

Post-hoc Statistical Power

A post-hoc power analysis was conducted using Soper's method (2016). For the DV's observed R^2 of .584, eleven predictors, a probability level of .05, and 483 participants, the model's statistical power was 1.00. Thus, it was concluded that: (a) the final model demonstrated an outstanding power to detect the significant effects and (b) the non-significant effects were definitely not significant.

Discussion

The study embarked on the task of developing a model unfolding the interconnections of a number of factors responsible for shaping the user views for SGs learning effectiveness, as well as their impact on this subjective construct. The data analysis delivered a series of interesting findings worthy of further discussion.

Results Summary

The relevant literature suggested that a model's predictive power is meaningful when it demonstrates high R^2s as well as structural paths above .30 (or at least close to .20) (Chin, 1988). Yet, interaction terms below the .20 threshold can also be taken into account, as long as they are statistically significant (Chin, Marcolin, & Newsted, 2003). Given the above, it can be supported that the model's predictive power is more than satisfactory. That is because a substantial 58.4% (R^2 = .584) of subjective learning effectiveness (the main DV) was explained, having as significant determinants enjoyment (β = .384), subjective relevance to personal interests (β = .183), subjective learning goal clarity (β = .154), subjective realism (β = .144), subjective narration quality (β = .115), and user game-playing competence (β = -.065). The fact that six predictors were able to interpret almost 60% of the variance of SG subjective learning effectiveness, is rather notable. As for the endogenous variables, an impressive 67.9% (R^2 = .679) of enjoyment was explained by subjective narration quality (β = .484), subjective quality of the learning material (β = .160), subjective ease of use (β = .157), subjective audiovisual fidelity (β = .118), subjective realism (β = .117), and presence (β = .078). Motivation was also adequately explained (R^2 = .488) by subjective quality of the learning material (β = .224), subjective narration quality (β = .155), and subjective learning goal clarity (β = .131). Subjective relevance to personal interests (R^2 = .405) was affected by enjoyment (β = .323), subjective realism (β = .207), subjective narration quality (β = .177), subjective quality of the learning material (β = .144), and subjective audiovisual fidelity (β = .124). The least explained construct was presence (R^2 = .349), which was significantly influenced by subjective realism (β = .416) and subjective narration quality (β = .330).

Main Findings

Associating the findings with that of previous research is a difficult task, given that the model examined multiple factors, in contrast to other studies in which a few factors were considered (e.g., Dobrovsky et al., 2019; Hookham & Nesbitt, 2019; Khan & Webster, 2017; Kiunsi & Ferwerda, 2019; Steiner et al., 2015; Thorpe et al., 2019). Even so, the most influential factor in shaping user views for SG learning effectiveness was enjoyment. In fact, besides having a rather strong impact on subjective learning effectiveness (β = .384), it had a striking impact on motivation (β = .565) and a notable one on subjective relevance to personal interests (β = .323). In general, it is accepted that when a learning activity (SGs included) is enjoyable, there is a positive impact on competencies and performance (Csikszentmihalyi, Rathunde, & Whalen, 1993). Logically enough, the findings regarding enjoyment give further support to previous studies which reported that this factor had a significant contribution to the effectiveness of SGs (e.g., Connolly et al., 2012; Hookham & Nesbitt, 2019; Steiner et al., 2015) and in their motivational appeal (e.g., Connolly et al., 2012; Iten & Petko, 2016).

Findings related to factors in the content features category

Coming to the variables included in the content features category, the results indicated that subjective narration quality directly affected subjective learning effectiveness (β = .115). Additionally, it had indirect effects through enjoyment and subjective relevance to personal interests (β = .174 and β = .032 respectively). It also had a remarkable impact on enjoyment (β = .484), an equally significant impact on presence (β = .330), while its impact on subjective relevance to personal interests and motivation was less impressive but still noteworthy (β = .177 and β = .155 respectively). On the basis of the above, narration quality proved to the construct that influenced the larger number of other factors. These results provide further support to the literature that noted the impact of this construct on learning (e.g., Calderón & Ruiz, 2015; Galván-Pérez et al., 2018; Mitgutsch & Alvarado, 2012; Ravyse et al., 2017). Additionally, narration's considerable impact on presence confirms Van Elk (2006) who proposed that narration facilitates situated cognition by immersing players into a setting which frames knowledge. What is more, the present study highlights even further the importance of this construct by linking it to the enjoyment one feels when playing SGs and by indicating that it can elevate users' personal interest for the learning content. In sum, the multiplicity of its effects affirms Prensky (2007) who concluded that the educational games' narrative is one of their key elements.

Subjective goal clarity had an impact on subjective learning effectiveness (β = .154) as well as motivation (β = .131). In psychology, the role of goals in cognition and motivation is a well-explored subject. It is accepted that when they are well-crafted, difficult but achievable, individuals are motivated (Fishbein & Ajzen, 1975). Moreover, when there is no conflict with other goals, there is a (positive) linear relation with performance (Locke & Latham, 2002). In the context of SGs, Shi and Shih (2015) theorized that there is a relation between goals and motivation as well as between goals and pleasurable experiences. While the results support the former, they cannot confirm the latter.

Subjective quality of the learning material affected -rather weakly- subjective learning effectiveness only indirectly through enjoyment (β = .064) and subjective relevance to personal interests (β = .027). It also had a substantial impact on motivation (β = .224) and lesser ones on enjoyment (β = .160) and subjective relevance to personal interests (β = .144). A number of researches recognized the predominant role of the learning content in SGs either by indicating that this feature and the gaming experience are correlated (Khenissi et al., 2015) or by advising its seamless integration into the game (e.g., Breuer & Bente, 2010; Mortara et al., 2014). The link that was found between the subjective quality of the learning content and enjoyment provides support for these views. Indeed, it is possible that when there is a good balance between gaming and learning characteristics, the latter does not obstruct the former. On the other hand, no studies were located indicating that the quality of the learning material has an impact on motivation. Given that, it can be supported that the study provides further evidence for the importance of this factor. Nevertheless, this link is a logical one; the closer the content is related to an individual's interests, the more he/she is motivated to learn.

Findings related to factors in the technical features category

As far as the variables in the technical features category are concerned, only subjective realism had a direct effect on subjective learning effectiveness (β = .144) as well as indirect ones through enjoyment and subjective relevance to personal interests (β = .064 and β = .036 respectively). The effects of subjective audiovisual fidelity were indirect ones, through enjoyment (β = .044) and subjective relevance to personal interests (β = .022). The case of subjective ease of use will be discussed in a coming paragraph.

Together with subjective narration quality, subjective realism proved to be the most important factors. Besides the positive effect on subjective learning effectiveness, it had a substantial positive impact on presence (β = .416), a considerable one on subjective relevance to personal interests (β = .207), and -a lesser one- on enjoyment (β = .117). As for audiovisual fidelity, its effects were limited on subjective relevance to personal interests (β = .124) and enjoyment (β = .118). There is a thin line separating realism and audiovisual fidelity, as already discussed in the "Factors affecting the (subjective) learning effectiveness of serious games" section. Then again, it was decided to examine them as two distinct factors because realism has functional and psychological dimensions. Given that these closely related constructs affected -mostly- different factors, is a good justification for this decision.

Realism's role in learning has fueled several debates; some supported that it positively affects learning (e.g., Byun & Loh, 2015; Chittaro & Buttussi, 2015), others viewed it as a distractor (e.g., Ke & Abras, 2013), while some suggested that it does not really matter how realistic SGs are (Norman et al., 2012; Vogel et al., 2006). On the basis of the findings, it can be supported that, at least from the participants' point of view, there is indeed an impact on learning. The positive impact of realism on presence has been discussed by Ivory and Kalyanaraman (2007) but in the context of video games in general. The results suggest that this link exists in SGs as well. Both subjective realism and audiovisual fidelity had an impact on enjoyment. This finding was anticipated, although it was expected their impact to be significantly stronger, as the findings of several studies indicated (e.g., Couceiro et al, 2013; Huang et al., 2013; Hunicke, Leblanc, & Zubek, 2004).

Interestingly enough, no paths linking both factors to motivation were found. This result contrasts with previous research supporting the existence of such connections (e.g., Huang et al., 2013; de Freitas et al., 2009). On the other hand, several have noted that gaming and educational features need

to be well-balanced because, otherwise, problems related to learning might arise (Ravyse et al., 2017; Squire & Jenkins, 2003; Wouters et al., 2011). In a way, the findings demonstrate the essentiality of the above directive. They might have revealed that the gaming abstraction followed by both SGs was not directly connected (or well-aligned) with the learning tasks, given that both factors had an impact on enjoyment, none had an impact on motivation, and only subjective realism had a weak impact on subjective learning effectiveness. Then again, further investigation is certainly needed.

Findings related to other factors

Subjective relevance to personal interests, presence, and subjective ease of use impacted just one factor each. Subjective relevance to personal interests had a moderate impact on subjective learning effectiveness (β = .183) but also mediated the effects of other factors on the latter (see Table 4). Iten and Petko (2016) argued that an SG has to be wholesomely engaging (i.e., cognitively, emotionally, and behaviorally). On the basis of the findings, it can be added that it is hard to achieve the above if users cannot relate what they learn through an SG with their own interests or if they do not find it personally meaningful.

Presence, besides being the least explained construct, it -weakly- affected only enjoyment (β = .078) and did not have a direct or an indirect effect on subjective learning effectiveness (see Table 4), despite the fact that others supported that this link does exist (e.g., Cheng et al., 2015). It seems that presence not only suffers from definitional issues but also its effects are negligible or even non-existent as noted by Hamari et al. (2016). Two issues, related to this study, can be identified as negatively affecting this psychological state. First, the study was conducted in computer labs. Although certain measures were taken in order to make participants feel more "at home" (e.g., by having a limited number of them at each lab), distractions were unavoidable. Second, participants' game-playing skills were average (M = 3.08, SD = 1.11); Cheng et al. (2015), on the other hand, argued that presence's effects emerge when users are highly competent in playing games. Given that SGs are played not only in private but also during lessons and considering its minimal effect, it can be excluded from future research in which the settings do not accommodate its manifestation.

Subjective ease of use, which, in this study, included aspects of playability and usability, had a direct effect only on enjoyment (β = .157) and an indirect one on subjective learning effectiveness through this factor (β = .065, Table 4). This is one of the study's surprising findings as it is in contrast with the findings of previous research. Indeed, it is routinely used for assessing SG quality and effectiveness and most studies underlined its impact on user experience and learning outcomes (e.g., All et al., 2015; Calderón & Ruiz, 2015; Steiner et al., 2015). Contrary to that, our findings suggest that users consider easy to use SGs as more fun to play and nothing more. The inclusion of ease of use in research can be traced back to the early days of computers when they constituted a novelty and many people did not know how to use them. Quite logically, models that tried to predict the behavioral intention to use computers and their applications, included ease of use as a construct (e.g., Technology Acceptance Model, Davis et al., 1989). On the basis of the findings, it can be suggested that the notion of ease of use is not useful anymore and that we have to move forward by considering the inclusion of factors that specifically examine playability and/or usability. Then again, the definitional problems of these terms have to be resolved (Mitgutsch & Alvarado, 2012; Sánchez et al., 2012).

Finally, it was found that users' game-playing competencies had a weak negative impact on subjective learning effectiveness (β = -.065), meaning that the higher the game playing skills, the less one considered SGs as being effective in terms of learning. A plausible explanation is that gamers viewed SGs more as games (focusing on gaming features) and less as instructional material. This finding stress, even more, the need to balance pedagogy and enjoyment (Ravyse et al., 2017; Squire & Jenkins, 2003; Wouters et al., 2011).

Missing paths and factors

While the paths that are present in the model can lead to interesting conclusions, equally interesting ones can be drawn from the paths that are missing or from the fact that some variables proved to be non-significant. For example, subjective feedback quality was eliminated as a factor; all the relevant hypotheses were rejected either during the specification search or because they were statistically non-significant (all *p* values > .100, see Table 3). This is somehow surprising, given that many noted its role in learning (e.g., Galván-Pérez et al., 2018; Ravyse et al., 2017; Sušnik et al., 2018; Wilson et al., 2017) or even enjoyment (e.g., Klimmt, Hartmann, & Frey, 2007) and learning goals (Prensky, 2007). This finding is puzzling and hard to decipher. Considering that a total of eleven constructs predicting subjective learning effectiveness were included in the initial model, a plausible explanation is that other factors (and their paths) outshined its importance, rendering it irrelevant altogether.

It was also found that gender did not act as a moderator. The relevant literature examining gender differences in SGs suggested that they are considered effective and motivating regardless of ones gender (Connolly et al., 2012; Hainey et al., 2016; Mayer, 2019). Our findings give further support to these views. The SG type/genre was also not a moderator. This finding is important, as it will be further elaborated in the "Implications for research and practice" section. Participants' ICT competence, that was used as a control variable, is also absent. This result was, more or less, expected, as participants' skills in using computers were above the mid-point ($M = 3.57$, $SD = 0.86$). This finding gives further support to the argument that subjective ease of use can stop being used in future research.

Probably the most controversial finding is that of the missing path from motivation to subjective learning effectiveness, given that there are numerous studies emphasizing its role in learning (e.g., Calderón & Ruiz, 2015; Hookham & Nesbitt, 2019; Huang et al., 2010; Kiunsi & Ferwerda, 2019; Yang, Lin, & Liu, 2017; Zhou & Mayer, 2017). A number of reasons might have led to this outcome. For example, Ryan et al. (2006) theorized that SGs limit user autonomy and control, as they are, basically, instructional applications; thus, their motivation is negatively affected and, in turn, the learning outcomes. Others supported that problematic motivation is an indicator of incompatibilities between entertainment and instructional design (Wouters et al., 2011). The same authors also noted that affective states, such as motivation, cannot be accurately measured after an SG has been played and advised in-game methods. However, as the above do not offer a concrete explanation for the missing link and as this finding is rather unusual, further examination of this issue is definitely needed.

Implications for Research and Practice

It is true that there is evidence about SG learning effectiveness (e.g., de Freitas 2018; Erhel & Jamet, 2019) and about their applicability in a wide range of learning subjects/scenarios (Feng et al., 2018; Hersh & Leporini, 2018). Then again, all involved in the field of SGs, struggle to fully understand what makes them "work." In essence, there is a need of tools that would allow all involved in the field of SGs to separate the "good" from the "bad" ones and to make informed decisions on the basis of certain characteristics and/or quality dimensions. The core reasons for this problematic situation is the multiplicity of the factors involved and that the researchers focus on a limited number of them, depending on their preferences (e.g., Faizan et al., 2019; Hookham & Nesbitt, 2019; Steiner et al, 2015). Definitional problems of some factors and the lack of instruments holistically evaluating SGs (Alonso et al., 2019; Hersh & Leporini, 2018) render the whole situation even harder to tackle. In this -full of uncertainties- environment, a model for explaining how users view SG learning effectiveness was proposed and tested. Although this can be considered as just a first step toward a better understanding of SGs and despite the limitations of such an effort, as discussed in a coming section, there are some noteworthy implications for research and practice.

The model's fit indices as well as its capacity to adequately explain subjective learning effectiveness (and most of the endogenous variables) are indicators of its trustworthiness. Thus, in terms of the study's contribution to research, it can be argued that it offers the prototype of a rather complex model, accurately explaining the intricate relationships between the substantial number of factors

that were measured. In this respect, researchers can use it in order to understand what shapes users' learning experiences when playing SGs. An additional argument in favor of the model is that it proved to be invariant across the two -rather different- SGs we used. As already presented in the "Materials" section, the inclusion in the study of more than one SG was a concern. The fact that differences were not observed, it can be proclaimed that there are some "universal" factors impacting subjective learning effectiveness, regardless of SG type. On the other hand, it cannot be ruled out that their impact fluctuates within a -yet undefined-range. In this respect, the model can be viewed as a first step in overcoming a problem highlighted by previous research, that of the exact role of certain factors (e.g., Hersh & Leporini, 2018; Ravyse et al., 2017).

The practical applications of the model also derive from the above. SG developers can use the model so as to decide on which factors to focus, depending on their needs. For example, by knowing beforehand that enjoyment has a significant impact on motivation, they can decide that there is no need to make a game more fun to play if they consider that it is already motivating enough. Others, by knowing that realism has a rather weak effect on enjoyment, they might decide to make a game more realistic, anticipating that this will increase enjoyment. Educators and education policymakers can also benefit from the use of the model together with scales evaluating the quality of educational software. By assessing, for example, technical and content features and by using the model as a blueprint, they can envisage how enjoyable and motivating an SG might be, as well as how it is going to impact user views regarding its learning effectiveness.

Limitations and Future Work

Several limitations to the present research can be acknowledged. Then again, these limitations might provide interesting ideas for future research paths. To start with, the sample size could have been larger, allowing for more confidence in the robustness of the statistical analysis and results. As already indicated in the "Method" section, the target group(s) could have been more diverse or more well-aligned to the SGs learning subjects. The use of a questionnaire for data collection purposes is always a concern; the trustworthiness of participant responses is not guaranteed. The study was conducted under controlled conditions that might have affected their attitudes (e.g., they might have thought that the study was a part of a course). Although two hours were considered enough time for participants to form an idea about the SGs, one can still argue that more time was needed. Finally, it is possible that certain significant factors were not examined. For example, the interface quality could have been included as a parameter, as some supported that, if it is effectively designed, users become even more engaged in the game world (Thorpe et al., 2019). On the other hand, there were some strong reservations regarding the number of constructs to include in the model, as its complexity exponentially increases each time a factor is added (see Figure 1).

On the basis of the above, it would be useful, in the future, to test several alternative models so as to develop a much simpler one (with just the most significant factors) which, at the same time, will have the capacity to adequately explain what users think of SGs. Several different target groups and SGs have to be examined in order to establish that the model is indeed invariant across a wide range of SGs genres and users. Finally, an interesting idea is to collect data regarding actual learning outcomes, include a corresponding variable in the model, and examine the relationship between subjective and objective learning effectiveness.

Conclusion

Within the theoretical framework and findings of previous research, a model for exploring users' views on SG learning effectiveness was developed and tested. It was found that enjoyment, relevance to personal interests, subjective goal clarity, subjective SG realism, and subjective narration adequacy, have a significant impact on SG subjective learning effectiveness. Furthermore, the above factors (with the exception of subjective goal clarity) affect several other factors as well. On the other hand, presence, subjective ease of use, and motivation had a minimal impact. Interestingly enough, subjec-

tive feedback quality was removed from the model. Given the above and despite the study's limitations, it is quite justified to support that it contributes to the relevant literature, as it illustrates the complex relationships between the factors that were considered as well as their impact on subjective learning effectiveness. Therefore, the study's outcomes might prove useful to the industry for developing effective SGs, to educators and education policymakers for exploiting their instructive potential, and to researchers for understanding the interactions between different factors that come into play when users play SGs.

REFERENCES

Abt, C. A. 1970. *Serious games.* New York: Viking.

All, A., Castellar, E. P. N., & Van Looy, J. (2015). Towards a conceptual framework for assessing the effectiveness of digital game-based learning. *Computers & Education, 88*, 29-37. https://doi.org/10.1016/j.compedu.2015.04.012

Alonso-Fernández, C., Pérez-Colado, I., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2018). Improving serious games analyzing learning analytics data: Lessons learned. *Proceedings of the International Conference on Games and Learning Alliance* (pp. 287-296). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-030-11548-7_27

Annetta, L. A. (2010). The "I's" have it: A framework for serious educational game design. *Review of General Psychology, 14*(2), 105-113. https://doi.org/10.1037/a0018985

Annetta, L. A., & Bronack, S. (Eds.). (2011). *Serious educational game assessment: Practical methods and models for educational games, simulations and virtual worlds.* Berlin, Germany: Springer Science & Business Media. https://doi.org/10.1007/978-94-6091-329-7

Barnes, T., Richter, H., Powell, E., Chaffin, A., & Godwin, A. (2007). Game2Learn: Building CS1 learning games for retention. *ACM SIGCSE Bulletin, 39*(3), 121-125. https://doi.org/10.1145/1269900.1268821

Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P., & Berta, R. (2013). Assessment in and of serious games: An overview. *Advances in Human-Computer Interaction, 2013*, 1. https://doi.org/10.1155/2013/136864

Blumberg, F. C., Almonte, D. E., Anthony, J. S., & Hashimoto, N. (2013). Serious games: What are they? What do they do? Why should we play them? In K. E. Dill (Ed.), *The Oxford handbook of media psychology* (pp. 334-351). Oxford, UK: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195398809.013.0019

Breuer, J., & Bente, G. (2010). Why so serious? On the relation of serious games and learning. *Journal for Computer Game Culture, 4*, 7-24.

Brockmyer, J. H., Fox, C. M., Curtiss, K. A., McBroom, E., Burkhart, K. M., & Pidruzny, J. N. (2009). The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology, 45*(4), 624-634. https://doi.org/10.1016/j.jesp.2009.02.016

Brown, T. A. (2014). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: Guilford Press.

Burnham, K. P., & Anderson, D. R. (1998). *Model selection and inference: A practical information-theoretic approach.* New York: Springer-Verlag. https://doi.org/10.1007/978-1-4757-2917-7

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming.* Abingdon, UK: Routledge.

Byun, J., & Loh, C. S. (2015). Audial engagement: Effects of game sound on learner engagement in digital game-based learning environments. *Computers in Human Behavior, 46*, 129-138. https://doi.org/10.1016/j.chb.2014.12.052

Cai, Y., Lu, B., Zheng, J., & Li, L. (2006). Immersive protein gaming for bio edutainment. *Simulation & Gaming, 37*(4), 466-475. https://doi.org/10.1177/1046878106293677

Calderón, A., & Ruiz, M. (2015). A systematic literature review on serious games evaluation: An application to software project management. *Computers & Education, 87*, 396-422. https://doi.org/10.1016/j.compedu.2015.07.011

Cheng, M. T., & Annetta, L. (2012). Students' learning outcomes and learning experiences through playing a Serious Educational Game. *Journal of Biological Education, 46*(4), 203-213. https://doi.org/10.1080/00219266.2012.688848

Cheng, M. T., Lin, Y. W., & She, H. C. (2015). Learning through playing Virtual Age: Exploring the interactions among student concept learning, gaming performance, in-game behaviors, and the use of in-game characters. *Computers & Education, 86*, 18-29. https://doi.org/10.1016/j.compedu.2015.03.007

Cheng, M.-T., She, H.-C., & Annetta, L. A. (2015). Game immersion experience: Its hierarchical structure and impact on game-based science learning. *Journal of Computer Assisted Learning, 31*(3), 232-253. https://doi.org/10.1111/jcal.12066

Chin, W. W. (1988). Issues and opinion on structural equation modeling. *MIS Quarterly, 22*, vii-xvi.

Chin, W. W., Marcolin, B. L., & Newsted, P. R. (2003). A partial least squares latent variable modeling approach for measuring interaction effects: Results from a Monte Carlo simulation study and an electronic-mail emotion/adoption study. *Information Systems Research, 14*(2), 189-217. https://doi.org/10.1287/isre.14.2.189.16018

Chittaro, L., & Buttussi, F. (2015). Assessing knowledge retention of an immersive serious game vs. a traditional education method in aviation safety. *IEEE Transactions on Visualization and Computer Graphics, 21*(4), 529-538. https://doi.org/10.1109/TVCG.2015.2391853

Choi, B., Huang, J., Jeffrey, A., & Baek, Y. (2013). Development of a scale for fantasy state in digital games. *Computers in Human Behavior, 29*(5), 1980-1986. https://doi.org/10.1016/j.chb.2013.04.007

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

Cook, D. A., Hamstra, S. J., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., ... & Hatala, R. (2013). Comparative effectiveness of instructional design features in simulation-based education: systematic review and meta-analysis. *Medical Teacher, 35*(1), e867-e898. https://doi.org/10.3109/0142159X.2012.714886

Couceiro, R. M., Papastergiou, M., Kordaki, M., & Veloso, A. I. (2013). Design and evaluation of a computer game for the learning of Information and Communication Technologies (ICT) concepts by physical education and sport science students. *Education and Information Technologies, 18*(3), 531-554. https://doi.org/10.1007/s10639-011-9179-3

Csikszentmihalyi, M., Rathunde, K., & Whalen, S. (1993). *Talented teenagers: The roots of success and failure*. New York: Cambridge University Press.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*(8), 982-1003. https://doi.org/10.1287/mnsc.35.8.982

de Freitas, S. (2018). Are games effective learning tools? A review of educational games. *Journal of Educational Technology & Society, 21*(2), 74-84.

de Freitas, S., & Ketelhut, D. J. (2014). Preface: Introduction for the Journal of Information Sciences special issue on serious games. *Information Sciences: An International Journal, 264*, 1-3. https://doi.org/10.1016/j.ins.2014.01.036

de Freitas, S., & Oliver, M. (2006). How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? *Computers & Education, 46*(3), 249-264. https://doi.org/10.1016/j.compedu.2005.11.007

de Freitas, S., Rebolledo-Mendez, G., Liarokapis, F., Magoulas, G., & Poulovassilis, A. (2009). Developing an evaluation methodology for immersive learning experiences in a virtual world. *Proceedings of the Games and Virtual Worlds for Serious Applications Conference* (pp. 43-50). Piscataway, NJ: IEEE. https://doi.org/10.1109/VS-GAMES.2009.41

DeVellis, R. F. (2016). *Scale development: Theory and applications* (Vol. 26). Thousand Oaks, CA: Sage Publications.

Dobrovsky, A., Borghoff, U. M., & Hofmann, M. (2019). Improving adaptive gameplay in serious games through interactive deep reinforcement learning. In *Cognitive infocommunications, theory and applications* (pp. 411-432). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-95996-2_19

Driscoll, M. P. (2002). How people learn (and what technology might have to do with it). *ERIC Digest*, 144.

Erhel, S., & Jamet, E. (2019). Improving instructions in educational computer games: Exploring the relations between goal specificity, flow experience and learning outcomes. *Computers in Human Behavior, 91*, 106-114. https://doi.org/10.1016/j.chb.2018.09.020

Faizan, N., Löffler, A., Heininger, R., Utesch, M., & Krcmar, H. (2019). Classification of evaluation methods for the effective assessment of simulation games: Results from a literature review. *International Journal of Engineering Pedagogy, 9*(1), 19-33. https://doi.org/10.3991/ijep.v9i1.9948

Fassbender, E., Richards, D., Bilgin, A., Thompson, W. F., & Heiden, W. (2012). VirSchool: The effect of background music and immersive display systems on memory for facts learned in an educational virtual environment. *Computers & Education, 58*(1), 490-500. https://doi.org/10.1016/j.compedu.2011.09.002

Feng, Z., González, V. A., Amor, R., Lovreglio, R., & Cabrera-Guerrero, G. (2018). Immersive virtual reality serious games for evacuation training and research: A systematic literature review. *Computers & Education, 127*, 252-266. https://doi.org/10.1016/j.compedu.2018.09.002

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Fokides, E., & Atsikpasi, P. (2018). Factors affecting primary school students' learning outcomes when using MUVEs. Development and validation of a scale. In J. Y. Qian (Ed.), *Integrating multi-user virtual environments in modern classrooms* (pp. 185-206). Hershey, PA: IGI Global. https://doi.org/10.4018/978-1-5225-3719-9.ch009

Fokides, E., Atsikpasi, P., Kaimara, P., & Deliyannis, I. (2019). 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

Fu, F. L., Su, R. C., & Yu, S. C. (2009). EGameFlow: A scale to measure learners' enjoyment of e-learning games. *Computers & Education, 52*(1), 101-112. https://doi.org/10.1016/j.compedu.2008.07.004

Galván-Pérez, L., Ouariachi, T., Pozo-Llorente, M., & Gutiérrez-Pérez, J. (2018). Outstanding videogames on water: a quality assessment review based on evidence of narrative, gameplay and educational criteria. *Water, 10*(10), 1404. https://doi.org/10.3390/w10101404

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33*(4), 441-467. https://doi.org/10.1177/1046878102238607

Gaskin, J. & Lim, J. (2017). *"CFA Tool", AMOS Plugin*. Retrieved from http://statwiki.kolobkreations.com

Habgood, M. J., & Ainsworth, S. E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *The Journal of the Learning Sciences, 20*(2), 169-206. https://doi.org/10.1080/10508406.2010.508029

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

Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior, 54*, 170-179. https://doi.org/10.1016/j.chb.2015.07.045

Hancock, G. R. (2001). Rethinking construct reliability within latent variable systems. In *Structural equation modeling: Present and future* (pp. 195-216). Scientific Software International.

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.

Hersh, M., & Leporini, B. (2018). Serious games, education and inclusion for disabled people. *British Journal of Educational Technology, 49*(4), 587-595. https://doi.org/10.1111/bjet.12650

Hookham, G., & Nesbitt, K. (2019, January). A systematic review of the definition and measurement of engagement in serious games. *Proceedings of the Australasian Computer Science Week Multiconference* (p. 42). New York, NY: ACM. https://doi.org/10.1145/3290688.3290747

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Huang, W. H., Huang, W. Y., & Tschopp, J. (2010). Sustaining iterative game playing processes in DGBL: The relationship between motivational processing and outcome processing. *Computers & Education, 55*(2), 789-797. https://doi.org/10.1016/j.compedu.2010.03.011

Huang, W. D., Johnson, T. E., & Han, S. H. C. (2013). Impact of online instructional game features on college students' perceived motivational support and cognitive investment: A structural equation modeling study. *The Internet and Higher Education, 17*, 58-68. https://doi.org/10.1016/j.iheduc.2012.11.004

Hunicke, R., Leblanc, M., & Zubek, R. (2004). A formal approach to game design and game research. *Proceedings of the AAAI Workshop on Challenges in Game AI* (Vol. 4).

Imlig-Iten, N., & Petko, D. (2018). Comparing serious games and educational simulations: Effects on enjoyment, deep thinking, interest and cognitive learning gains. *Simulation & Gaming, 49*(4), 401-422. https://doi.org/10.1177/1046878118779088

Iten, N., & Petko, D. (2016). Learning with serious games: Is fun playing the game a predictor of learning success? *British Journal of Educational Technology, 47*(1), 151-163. https://doi.org/10.1111/bjet.12226

Ivory, J. D., & Kalyanaraman, S. (2007). The effects of technological advancement and violent content in video games on players' feelings of presence, involvement, physiological arousal, and aggression. *Journal of Communication, 57*(3), 532-555. https://doi.org/10.1111/j.1460-2466.2007.00356.x

Jarvis, S., & de Freitas, S. (2009,). Evaluation of an immersive learning programme to support triage training. *Proceedings of the Games and Virtual Worlds for Serious Applications Conference* (pp. 117-122). Piscataway, NJ: IEEE. https://doi.org/10.1109/VS-GAMES.2009.31

Johnson, L., Becker, S. A., Cummins, M., Estrada, V., Freeman, A., & Hall, C. (2016). *NMC horizon report: 2016 higher education edition.* Austin, TX: The New Media Consortium.

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. Smart pedagogy for technology enhanced learning* (pp. 113-117). Hershey, PA: IGI Global. https://doi.org/10.1007/978-3-030-01551-0_6

Ke, F. (2011). A qualitative meta-analysis of computer games as learning tools. In *Gaming and simulations: Concepts, methodologies, tools and applications* (pp. 1619-1665). Hershey, PA: IGI Global. https://doi.org/10.4018/978-1-60960-195-9.ch701

Ke, F., & Abras, T. (2013). Games for engaged learning of middle school children with special learning needs. *British Journal of Educational Technology, 44*(2), 225-242. https://doi.org/10.1111/j.1467-8535.2012.01326.x

Ketamo, H., & Kiili, K. (2010). Conceptual change takes time: Game based learning cannot be only supplementary amusement. *Journal of Educational Multimedia and Hypermedia, 19*(4), 399-419.

Khan, A., & Webster, J. (2017). Digital game narrative quality: Developing a measure. *Proceedings of the Thirty-eighth International Conference on Information Systems*, *Seoul, Korea.*

Khenissi, M. A., Essalmi, F., & Jemni, M. (2015). Comparison between serious games and learning version of existing games. *Procedia-Social and Behavioral Sciences, 191*, 487-494. https://doi.org/10.1016/j.sbspro.2015.04.380

Kiili, K. (2005a). Digital game-based learning: Towards an experiential gaming model. *The Internet and Higher Education, 8*(1), 13-24. https://doi.org/10.1016/j.iheduc.2004.12.001

Kiili, K. (2005b). Content creation challenges and flow experience in educational games: The IT-Emperor case. *The Internet and Higher Education, 8*(3), 183-198. https://doi.org/10.1016/j.iheduc.2005.06.001

Kiunsi, D., & Ferwerda, B. (2019). Using a serious game to teach user-centered design. *Companion Proceedings of the 24th International on Intelligent User Interfaces: 3rd Workshop on Theory-Informed User Modeling for Tailoring and Personalizing Interfaces* (HUMANIZE).

Klimmt, C., Hartmann, T., & Frey, A. (2007). Effectance and control as determinants of video game enjoyment. *Cyberpsychology & Behavior, 10*(6), 845-848. https://doi.org/10.1089/cpb.2007.9942

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Lamb, R. L., Annetta, L., Firestone, J., & Etopio, E. (2018). A meta-analysis with examination of moderators of student cognition, affect, and learning outcomes while using serious educational games, serious games, and simulations. *Computers in Human Behavior, 80*, 158-167. https://doi.org/10.1016/j.chb.2017.10.040

Locke, E. A., & Latham, G. P. (2006). New directions in goal-setting theory. *Current Directions in Psychological Science, 15*(5), 265-268. https://doi.org/10.1111/j.1467-8721.2006.00449.x

Marsh, T. (2011). Serious games continuum: Between games for purpose and experiential environments for purpose. *Entertainment Computing, 2*(2), 61-68. https://doi.org/10.1016/j.entcom.2010.12.004

Matsunaga, M. (2010). How to factor-analyze your data right: do's, don'ts, and how-to's. *International Journal of Psychological Research, 3*(1), 97-110. https://doi.org/10.21500/20112084.854

Mayer, I., Bekebrede, G., Harteveld, C., Warmelink, H., Zhou, Q., van Ruijven, T., ... & 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

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, R. E., Griffith, E., Jurkowitz, I. T., & Rothman, D. (2008). Increased interestingness of extraneous details in a multimedia science presentation leads to decreased learning. *Journal of Experimental Psychology: Applied, 14*(4), 329. https://doi.org/10.1037/a0013835

McDonald, R. P., & Ho, M. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods, 7*(1), 64-82. https://doi.org/10.1037/1082-989X.7.1.64

Michael, D. R., & Chen, S. L. (2005). *Serious games: Games that educate, train, and inform*. Muska & Lipman/Premier-Trade.

Mitgutsch, K., & Alvarado, N. (2012). Purposeful by design? A serious game design assessment framework. *Proceedings of the International Conference on the Foundations of Digital Games*, 121-128. New York, NY: ACM. https://doi.org/10.1145/2282338.2282364

Moizer, J., Lean, J., Dell'Aquila, E., Walsh, P., Keary, A. A., O'Byrne, D., ... & Sica, L. S. (2019). An approach to evaluating the user experience of serious games. *Computers & Education, 136*, 141-151. https://doi.org/10.1016/j.compedu.2019.04.006

Mortara, M., Catalano, C. E., Bellotti, F., Fiucci, G., Houry-Panchetti, M., & Petridis, P. (2014). Learning cultural heritage by serious games. *Journal of Cultural Heritage, 15*(3), 318-325. https://doi.org/10.1016/j.culher.2013.04.004

Muratet, M., Viallet, F., Torguet, P., & Jessel, J. (2009). Une ingénierie pour jeux sérieux [Engineering for serious Games]. *Proceedings of Conférence EIAH-Atelier Jeux Sérieux*, *Le Mans, France,* 53-63.

Nilsson, N. C., Nordahl, R., & Serafin, S. (2016). Immersion revisited: A review of existing definitions of immersion and their relation to different theories of presence. *Human Technology, 12*. https://doi.org/10.17011/ht/urn.201611174652

Norman, G., Dore, K., & Grierson, L. (2012). The minimal relationship between simulation fidelity and transfer of learning. *Medical Education, 46*(7), 636-647. https://doi.org/10.1111/j.1365-2923.2012.04243.x

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity, 41*(5), 673-690. https://doi.org/10.1007/s11135-006-9018-6

Petridis, P., Dunwell, I., de Freitas, S., & Panzoli, D. (2010). An engine selection methodology for high fidelity serious games. *Proceedings of the Games and Virtual Worlds for Serious Applications Second International Conference* (pp. 27-34). Piscataway, NY: IEEE. https://doi.org/10.1109/VS-GAMES.2010.26

Pinelle, D., Wong, N., & Stach, T. (2008). Heuristic evaluation for games: usability principles for video game design. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1453-1462). New York, NY: ACM. https://doi.org/10.1145/1357054.1357282

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879. https://doi.org/10.1037/0021-9010.88.5.879

Prensky, M. (2007). *Digital game-based learning*. St. Paul, MN: Paragon House.

Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. *Review of General Psychology, 14*(2), 154. https://doi.org/10.1037/a0019440

Ravyse, W. S., Blignaut, A. S., Leendertz, V., & Woolner, A. (2017). Success factors for serious games to enhance learning: a systematic review. *Virtual Reality, 21*(1), 31-58. https://doi.org/10.1007/s10055-016-0298-4

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion, 30*(4), 344-360. https://doi.org/10.1007/s11031-006-9051-8

Ryu, E. (2011). Effects of skewness and kurtosis on normal-theory based maximum likelihood test statistic in multilevel structural equation modeling. *Behavior Research Methods, 43*(4), 1066-1074. https://doi.org/10.3758/s13428-011-0115-7

Sánchez, J. L. G., Vela, F. L. G., Simarro, F. M., & Padilla-Zea, N. (2012). Playability: Analysing user experience in video games. *Behaviour & Information Technology, 31*(10), 1033-1054. https://doi.org/10.1080/0144929X.2012.710648

Schiefele, U., Krapp, A., & Winteler, A. (1992). Interest as a predictor of academic achievement: a meta-analysis of research. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 183-212). Hillsdale, N.J: Erlbaum.

Serrano-Laguna, Á., Manero, B., Freire, M., & Fernández-Manjón, B. (2018). A methodology for assessing the effectiveness of serious games and for inferring player learning outcomes. *Multimedia Tools and Applications, 77*(2), 2849-2871. https://doi.org/10.1007/s11042-017-4467-6

Serrano-Laguna, Á., Torrente, J., Moreno-Ger, P., & Fernández-Manjón, B. (2014). Application of learning analytics in educational videogames. *Entertainment Computing, 5*(4), 313-322. https://doi.org/10.1016/j.entcom.2014.02.003

Shi, Y. R., & Shih, J. L. (2015). Game factors and game-based learning design model. *International Journal of Computer Games Technology, 2015*, 11. https://doi.org/10.1155/2015/549684

Soper, D. S. (2016). *Post-hoc statistical power calculator for multiple regression* [Software]. Retrieved from https://www.danielsoper.com/statcalc/calculator.aspx?id=9

Squire, K., & Jenkins, H. (2003). Harnessing the power of games in education. *Insight, 3*(1), 5-33. https://doi.org/10.1145/950566.950591

Steiner, C., Hollins, P., Kluijfhout, E., Dascalu, M., Nussbaumer, A., Albert, D., & Westera, W. (2015). *Evaluation of serious games: A holistic approach*. Retrieved from http://dspace.ou.nl/bitstream/1820/6139/1/RAGE_ICERI2015_final.pdf

Sušnik, J., Chew, C., Domingo, X., Mereu, S., Trabucco, A., Evans, B., ... & Brouwer, F. (2018). Multi-stakeholder development of a serious game to explore the water-energy-food-land-climate nexus: The SIM4NEXUS approach. *Water, 10*(2), 139. https://doi.org/10.3390/w10020139

Thorpe, A., Nesbitt, K., & Eidels, A. (2019). Assessing game interface workload and usability: A cognitive science perspective. *Proceedings of the Australasian Computer Science Week Multiconference* (p. 44). New York, NY: ACM. https://doi.org/10.1145/3290688.3290749

Usability-First, 2009. *Playability definition*. Retrieved from http://www.usabilityfirst.com/glossary/playability/

Van Eck, R. (2006). Digital game-based learning: It's not just the digital natives who are restless. *Educause Review, 41*(2), 16.

Verpoorten, D., Castaigne, J. L., Westera, W., & Specht, M. (2014). A quest for meta-learning gains in a physics serious game. *Education and Information Technologies, 19*(2), 361-374. https://doi.org/10.1007/s10639-012-9219-7

Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research, 34*(3), 229-243. https://doi.org/10.2190/FLHV-K4WA-WPVQ-H0YM

Vorderer, P., Klimmt, C., & Ritterfield, U. (2004). Enjoyment: at the heart of media entertainment. *Communication Theory, 14*(4), 388 -408. https://doi.org/10.1111/j.1468-2885.2004.tb00321.x

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. *Journal of Educational Technology & Society, 22*(1), 59-69.

Wilson, A. S., Broadbent, C., McGrath, B., & Prescott, J. (2017). Factors associated with player satisfaction and educational value of serious games. In *Serious games and edutainment applications* (pp. 513-535). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-51645-5_23

Winn, B. M. (2009). The design, play, and experience framework. In *Handbook of research on effective electronic gaming in education* (pp. 1010-1024). Hershey, PA: IGI Global. https://doi.org/10.4018/978-1-59904-808-6.ch058

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models an evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 73*(6), 913-934. https://doi.org/10.1177/0013164413495237

Wouters, P., Van Nimwegen, C., Van Oostendorp, H., & Van Der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of educational psychology, 105*(2), 249. https://doi.org/10.1037/a0031311

Yang, J. C., Lin, Y. L., & Liu, Y. C. (2017). Effects of locus of control on behavioral intention and learning performance of energy knowledge in game-based learning. *Environmental Education Research, 23*(6), 886-899. https://doi.org/10.1080/13504622.2016.1214865

Yoho, C. (2011). Using music to increase math skill retention. *Journal of the American Academy of Special Education Professionals, 146*, 151.

Yusoff, A., Crowder, R., & Gilbert, L. (2010). Validation of serious games attributes using the technology acceptance model. *Proceedings of the 2010 Second International Conference on Games and Virtual Worlds for Serious Applications* (pp. 45-51). Piscataway, NJ: IEEE. https://doi.org/10.1109/VS-GAMES.2010.7

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

Zhou, Q., & Mayer, I. S. (2017). Models, simulations and games for water management: A comparative q-method study in The Netherlands and China. *Water, 10*(1), 10. https://doi.org/10.3390/w10010010

APPENDIX – THE SGES

Factor	Item
Presence	I was deeply concentrated in the game
	If someone was talking to me, I couldn't hear him
	I forgot about time passing while using the game
	I felt detached from the outside world while using the game
Enjoyment	I think the game was fun
	I felt bored while using the game*
	I enjoyed using the game
	I really enjoyed studying with this game
	It felt good to successfully complete the tasks in this game
	I felt frustrated*
Subjective learning effectiveness	I felt that this game can ease the way I learn
	This game was a much easier way to learn compared to the usual teaching
	This game made learning more interesting
	I felt that the game increased my knowledge
	I felt that I caught the basics of what I was taught with this game
	I will definitely try to apply the knowledge I learned with this game
Subjective narration adequacy	I was captivated by the game's story from the beginning
	I enjoyed the fantasy or story provided by the game
	I could clearly understand the game's story
	I was very interested in seeing how the events in the game will unfold
Subjective 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 game, the virtual world was more real than the real world
Subjective feedback quality	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
Subjective audiovisual fidelity	I enjoyed the sound effects in the game
	I think the game's audio fits the mood or style of the game
	I felt the game's audio (e.g., sound effects, music) enhanced my (gaming) experience
	I enjoyed the music in the game
	I enjoyed the game's graphics
	I think the game was visually appealing
	I think the game's graphics fit the mood or style of the game
Subjective relevance to personal interests	The content of this game was relevant to my interests
	I could relate the content of this game to things I have seen, done, or thought about in my own life
	It is clear to me how the content of the game is related to things I already know
Subjective learning goal clarity	The game's goals were presented at the beginning of the game
	The game's goals were presented clearly
	The intermediate goals were presented at the beginning of each scene
Subjective ease of use	I think it was easy to learn how to use the game
	I found the game unnecessarily complex*
	I imagine that most people will learn to use this game very quickly
	I needed to learn a lot of things before I could get going with this game*

Factor	Item
	I felt that I needed help from someone else in order to use the game because It was not easy for me to understand how to control the game*
	It was easy for me to become skillful at using the game
Subjective 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 game were too difficult*
	I could not really understand quite a bit of the material in this game*
	The good organization of the content helped me to be confident that I would learn this material
Motivation	This game did not hold my attention*
	When using the game, I did not have the impulse to learn more about the learning subject*
	The game did not motivate me to learn*

Note. * = Item for which its scoring was reversed

BIOGRAPHIES

Dr. Emmanuel Fokides is an Assistant Professor in the Department of Primary School Education, University of the Aegean, Greece. His courses focus on the educational uses of emerging technologies, virtual reality, digital storytelling, augmented reality, and serious games. Since 1994, he has been involved in a number of research projects regarding distance and lifelong learning and the educational uses of virtual and augmented reality. He is also the founder of the Emerging Technologies in Education initiative (ETiE). His work is published in several conference proceedings, chapters in edited books, and journals. He is also the co-author of two books.

Pinelopi Atsikpasi holds a Master's Degree in Educational Sciences and is also a PhD student in the Department of Primary Education of the University of the Aegean, Greece. Her interests include the educational uses of VR, AR, and tablets, as well as the development of tools for assessing the learning outcomes of the above. She has been a core member of the Emerging Technologies in Education initiative (ETiE) since its beginning. Her work is published in a number of national and international journals, chapters in edited books, and conference proceedings.

Polyxeni Kaimara, MSc Educational Psychologist, is a PhD student in the Department of Audiovisual Arts of the Ionian University in Corfu, Greece. She also holds a Master's Degree in Public Health and she is a certified Counselor and Adult Educator. She currently works as a psychologist at an educational center for people with disabilities. She is the co-author of several papers and chapters. Her research interests focus on the design and evaluation of multimedia educational systems for inclusive education.

Ioannis Deliyannis is an Assistant Professor at the Department of Audiovisual Arts of the Ionian University in Corfu, Greece, in the scientific field of Analog and Digital Technologies of Interactive Multimedia. His research interests include the application of Interactive and Multi-media Technologies under innovative case studies within the fields of Interaction and Learning. He is the author of a series of journal and conference publications in the above fields, followed by a series of books targeting the experimental and creative aspects of the technologies involved. He is involved in the design of user-centered software products and services, focusing on the use of mobile sensory systems to create intelligent interactive systems, entertainment education systems, educational applications for people with disabilities, multimedia adapters, holograms, interactive navigation narrative applications, augmented and virtual reality system.