

# Adapting the Complexity Level of a Serious Game to the Proficiency of Players

Herre van Oostendorp<sup>1,\*</sup>, Erik D. van der Spek<sup>2</sup> and Jeroen Linssen<sup>3</sup>

<sup>1</sup>Department of Information and Computing Sciences, Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands

<sup>2</sup>Department of Industrial Design, Eindhoven University of Technology, Den Dolech 2, 5612 AZ Eindhoven, The Netherlands

<sup>3</sup>Department of Human Media Interaction, University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands

## Abstract

As games are continuously assessing the player, this assessment can be used to adapt the complexity of a game to the proficiency of the player in real time. We performed an experiment to examine the role of dynamic adaptation. In one condition, participants played a version of our serious game for triage training that automatically adapted the complexity level of the presented cases to how well the participant scored previously. Participants in the control condition played a version of the game with no adaptation. The adapted version was significantly more efficient and resulted in higher learning gains per instructional case, but did not lead to a difference in engagement. Adapting games to the proficiency of the player could make serious games more efficient learning tools.

**Keywords:** dynamic adaptation, engagement, learning efficiency, proficiency, serious game

Received on 29 December 2013, accepted on 05 March 2014, published on 22 May 2014

Copyright © 2014 H. van Oostendorp *et al.*, licensed to ICST. This is an open access article distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/3.0/>), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/sg.1.2.e5

## 1. Introduction

Serious games can be used to engender learning in a player, and two recent meta-analyses have shown that the usage of serious games may even lead to superior learning compared with traditional (but passive) instructional methods [1, 2]. However a serious game is found to be primarily efficacious if a person is allowed to play the game multiple times [1], a result that Wouters and Van Oostendorp [3] argue underlines the notion that games are complex environments in which the player first has to learn how to control the game and the way in which it conveys the instructional material, before this material itself can be learned. Games in turn are products that have to be made beforehand and have a preset pace, and often do not take into account the individual learning rate.

People learn at different speeds, which may lead to a number of problems. Firstly, the rich multimodal information of a game may *overload* the limited working memory capacity of a player, leading to incorrect learning

[4], and some learners will therefore benefit from a slower pace in the presentation of instructional material in order to correctly organize all the new information that is coming in. Conversely, efficient learning may also be hindered by cognitive *underload*, where the learner is stimulated too little, for instance when a quick learner plays a game that has a slow pace in order to accommodate slow learners. Cognitive underload can lead to (passive) fatigue, which has been shown to result in disengagement from the task and higher distractibility and can subsequently degrade performance [5, 6]. If a game were to actively prevent the player from becoming cognitively overloaded or underloaded, it could therefore be more efficient [7].

Secondly and closely related to this, Csikszentmihalyi [8] posited that one can experience the feeling of flow, which is a feeling where someone is completely engaged in an activity to the point of losing self-consciousness and the activity becomes rewarding in its own right, and that this leads to the individual functioning at his or her fullest capacity [9]. This is achieved when the provided challenge is optimally suited to the skills of the user; and as videogames are often stated to be engaging, with players

\*Corresponding author : h.vanoostendorp@uu.nl

reporting an experience of being completely absorbed in the game, they seem to be ideally suited to produce flow [10, 11]. Flow has been shown to be positively correlated to learning [12]; therefore, keeping players in a sense of flow by adjusting the challenge to their skills could improve learning [13].

Summarizing, if quick learners were able to progress in the game at a faster pace, for instance because the game recognizes their proficiency and *adapts* the game accordingly, engagement in performing the task could be enhanced which in turn results in a higher efficiency of the game. Similarly, a slower pace for slower learners would also improve engagement and efficiency for them. In this paper we will examine in an experimental study whether adapting a serious game to the proficiency of players improves learning and engagement. But first we will, in the next section, discuss different aspects of adaptivity in general, how we monitored or assessed proficiency of players and how we implemented adaptivity in a dynamic way in the serious game *Code Red Triage*.

## 2. Adaptivity

### 2.1. Aspects of adaptivity

In line with Lopes and Bidarra [14], we can distinguish several components of adaptation. 1) The game world and its objects can be varied, e.g. the layout of the game world can be made simpler for underachieving players [15]. 2) The game play mechanics, how game elements work, including actions like running or shooting, e.g. adjusting shooting difficulty by providing player aim assistance, according to individual skills [16]. 3) Adapting the attributes of the non-player characters in the game, e.g. increasing the abilities of the non-playing character when the player performs well. Domain knowledge is here automatically gathered by the game based on Artificial Intelligence-techniques, in order to offer more challenging behavior of the non-player characters [17]. 4) Game narratives, e.g. adapting the sequence of events to the pace or behavior of the player [18], and 5) game scenarios - more or less similar to the previous one: adapting the flow of events and actions within a game, that is, adapting the progression within a game level to the learning goals of the player. For instance, monitoring the players actions and based on that certain points in the plot are included in the game (or not) [19].

A next issue in creating adaptive games is to decide on the method of generating the content. Lopes and Bidarra distinguish two general methods. First, offline adaptivity (or customized content generation); adjustments are made considering player-dependent data, but prior to initiating the gameplay. Secondly, online adaptivity, i.e. adjusting the game to its players, in real time, as they play.

A further discussion on the way adaptation can be implemented in games and the associated challenges can, for instance, be found in Lopes and Bidarra [14]. Though in the (game) industry and academia now many different adaptive

(serious) games are developed, and progress has been made, empirical research to effects of adaptivity in terms of learning and engagement are still scarce [see also 20]. In this paper we will remedy this and present results of an empirical study on the learning and affective effects of a game with dynamic adaptivity. That is, a game where the challenges of, or difficulties caused by, the game are increasing, and at a rate dependent on the proficiency of the player (online adaptation). We will mainly be concerned with varying the attributes of the non-player characters.

### 2.2. Assessing the proficiency of players

For the principle of fitting the instruction to the learner's proficiency level to be implemented in serious games effectively, it is important first that the proficiency should be assessed and secondly that the challenge should be adapted to the player automatically in a non-obtrusive way. Automatically assessing and adapting the challenge or difficulty of a game to the proficiency of a player is slowly becoming commonplace in entertainment games. For instance in *Rocksmith* [21], a musical instrument simulation game, the player needs to hit the correct notes of a song with good timing. The game adds more notes and places a greater emphasis on timing when the player performs well, or vice versa when the player performs badly. Racing games like *Mario Kart* [22] and *Need for Speed* [23], implement a simple adaptation known as 'rubber banding': when the player lags behind the other racing contestants, they will slow down in order to let the player catch up with them – when the player is up front, his opponents will become faster and try to keep up with him.

Here, we will elaborate on two modes of assessing that are most relevant to our research. Firstly, one interesting avenue in which a game can be adapted to the player was undertaken by Yun *et al.* [24], who used an infrared camera that was mounted on a TV displaying the game. This camera (overtly) recorded the faces of the participants while they were playing a game that revolved around shooting enemy robots. Looking at the heat signatures from the supra-orbital region of the face, they were able to derive how much apparent stress the game exerted on the player during game play. At the same time, the player reported at set intervals whether they found the game too easy, just right or too difficult, and whether they were enjoying the game or would like to quit. This research is relevant to our own for two reasons. One, they discovered that people who found the game too difficult and wanted to quit actually had lower stress levels than when the game was moderately difficult. They argued that this is due to the player becoming disengaged with the game, thereby corroborating the previously made assertion that too high a challenge leads to cognitive overload and is detrimental to the engagement or flow experience. Two, a version of the game where the game automatically assesses and adapts to the stress level of the player was shown to lead to higher engagement and better in-game performance (in terms of how many robots

were defeated) than in conditions with preset difficulty levels, even for the easy difficulty level.

Another interesting example of how to adapt the game to the player is the entertainment game *The Elder Scrolls 4: Oblivion* [25]. Here, the player roleplays a character in a large and open medieval fantasy world. As the player encounters new locales, performs quests and defeats monsters, his or her character will gradually become stronger and gain better weapons and items (see further Shute *et al.*, [26]). Because the game features an open world for the player to explore freely, this traditionally leads to problems where the player may encounter monsters that are far too strong for his or her avatar to defeat at that point in time. To counter this and provide the optimal experience for everyone, the player's adversaries in the game also progress in power at the same rate as the skill level of the player. Contrarily to what would be expected, many gamers criticized this feature, as it made them feel that their actions were largely inconsequential [27]; they were not getting stronger than their enemies and therefore they didn't feel like they were mastering the game.

Above we mentioned two different techniques of assessing the player proficiency within the game. The first was a more *overt* technique, where in real life settings the player would have to install an infrared camera for it to work; the second example featured so-called 'stealth' assessment [7, 26], that is, a more *covert* assessment that is coupled to the naturally occurring moves of the player in the game. In essence, all games are an assessment device, in that progressing past an obstacle is contingent on acquiring the needed knowledge of how to do so. As digital games are played on computers, which require that every game rule and in-game problem encountered is computable, determining whether the player succeeded is often easily quantifiable.

### 2.3. Dynamic adaptivity in the serious game Code Red Triage

As indicated we want to study whether the online adaptation of the challenge or difficulty of a learning experience to the proficiency of players, improves learning and enhances engagement. Following [28] we use the term *dynamic adaptivity* to designate online adaptation of game experiences in terms of complexity and matching that to the proficiency of players. In order to test this hypothesis we used the serious game *Code Red Triage*, a total conversion mod of *Half-Life 2* [29-31]. The game is designed to teach the triage procedure, a procedure for medical first responders to prioritize the victims of a mass casualty event according to how urgently the victim needs medical

attention. The mobility (sieve) triage taught here is a relatively simple procedure, where it takes the first responder between one and five steps to determine the severity of the victim's injuries. When the game starts, the player finds himself in an empty train station with signs of recent panic. Here, he learns that he is a medical first responder who has received a call that a bomb has gone off on a subway platform. The player is then told to find the subway platform and perform the triage procedure on the victims. Upon reaching the subway platform (see Figure 1), a visible timer starts counting down from seventeen minutes. When the timer reaches zero, the game ends. This timer was added to instill a sense of immediacy and stress; in practice almost every participant is able to triage all victims comfortably within this time. At the subway platform, the player can then walk up to a victim and press a button to enter the triage menu, which consists of eight buttons for triage actions, and four buttons for the four different triage categories (see Figure 2). Pressing a triage button will give a few lines of general information on what the action entails and approximately at what stage in the procedure it should be used, and a line with specific information on how the action affected the victim the player's looking at. After choosing a few triage actions the player should be able to have an idea how heavily injured the victim is and assign a triage category.



Figure 1. Subway platform in the game *Code Red Triage*



**Figure 2.** Triage menu in the game



**Figure 3.** Feedback after categorizing a victim in a triage category

Once this is done, the victim changes color to depict the chosen category and the player receives a score showing how well he did, as well as a few lines telling him whether or not a) he forgot to take procedure steps, b) took steps in the wrong order, c) took unnecessary steps and d) whether it was done within the allotted time (between 10 and 55 seconds), see Figure 3 for a screenshot. The in-game score that can be obtained per victim ranges from 0 to 100 and is based on the previous four criteria.

In the case of *Code Red Triage*, we already have a measure to *assess* how well the player is performing in the game, namely the in-game score, which provides us with an objective measure of whether the player is able to correctly apply the procedure to a given victim case. The player's performance can therefore be seen as an indication of their proficiency level [7]. We can thus use the above mentioned covert method to assess the proficiency of players here.

We used this in-game score to *adapt* the difficulty of the game to the proficiency of the player. In *Code Red Triage*, there are a total of six paths with an increasing number of steps in the triage procedure that are taught with the game, but there are multiple victims for any given path. As the victims are encountered in increasing order of complexity (i.e. the number of steps needed to come to a correct categorization), these groups of victims are called 'victim tiers'. In the set of victims 6 tiers or levels of complexity were distinguished. In other words, the attributes of the non-player characters were varied in complexity. If a player scores above a preset threshold, he or she has proven to have a certain level of proficiency and can move on to a more complex victim tier. In the adaptive condition of *Code Red Triage* this was

operationalized as the game deleting all remaining victim cases within the same tier, if the player scored higher than a threshold value for that victim. The threshold was determined with the data from a pilot experiment, by rounding up the average score per victim tier. A player who was unable to triage a victim case and scored below the threshold, received one or more of the remaining cases of that tier before going to the next level of complexity. In other words, more successful players could attain the most complex case in less cases, and consequently learn to perform the triage more efficiently. In the control version of the game all (19) cases were presented in a gradually increasing complexity.

We hypothesize that players feel more engaged by the dynamic adaptive version, because the game always remains challenging (compared to a control version), and secondly we expect in the dynamic adaptive version of the game that players are able to learn more efficiently, because redundant learning experiences (triage cases) can be skipped.

## 3. Method

### 3.1. Participants

In total 28 individuals of university-level education, 19 male and 9 female, participated in the experiment, and were randomly assigned to the adaptive game condition (n=14), and the control condition (n=14). Average age was 22.86 with a standard deviation of 5.68.



### 3.2. Materials

To measure the learning of players, three types of instruments were used. The *in-game score* (see above) formed the first measure: an indication of the progression of the player in the game. In several studies done with the same game and the same in-game score we found that the in-game score significantly correlated with a knowledge test presented after the game [31, 32], which gives plausibility to the notion that the in-game score, conceived as analytical learning tool [33], is a valid measure of learning. Statistics from the game that were logged furthermore included triaged victims, number of triaged victims, tier of victim, time per victim, total time, score per victim and total score. Second and third, we measured how much a participant learned in the game with two measures: a pen-and-paper *knowledge test* and a *structural knowledge assessment*. The knowledge test was in the form of eight verbal and eight pictorial multiple choice questions where the player had to answer questions related to the triage procedure by choosing one of four alternatives (total score range 0-16).

Whereas the knowledge test measured how well the participant could reproduce declarative knowledge, the structural knowledge assessment determined how the information was organized on a deeper, more structural level. Here, a computer program called *PCKNOT* [34] was used, that let participants rate the degree of relatedness of pairs of concepts from the triage procedure. These ratings could subsequently be used to elicit a participant's knowledge structure with the Pathfinder metric [35] and compared to the knowledge structure of experts; resulting in a similarity measure that indicated how well the participant had organized the information of the triage procedure structurally [36]. The score range varies from -1 through 0 to +1. Pathfinder has been successfully applied by [37] to measure learning from a complex videogame. They found that it was also predictive of skill retention and skill transfer. For further information see Wouters, Van der Spek and Van Oostendorp [38]. In our case we focused on 8 important concepts from the triage procedure and consequently 28 pairs were presented for the related judgments. The created networks were compared with the referent structure that was derived by averaging the elicited knowledge structures of the current researchers.

The *engagement* of players was measured by using the subscale of the ITC Sense of Presence Inventory (ITC-SOPI), which indicates the participant's feelings of engagement with a twelve item five-point Likert scale [39]. If the challenge of the game is better adjusted to the abilities of the player, one would expect the player to be drawn into the game more, which we hoped to see expressed in the scores on this subscale. The reliability of the ITC-SOPI Engagement questionnaire appeared to be relatively low, Cronbach's coefficient  $\alpha = 0.59$ .

### 3.3. Apparatus and procedure

The game was played on a 17" laptop at a resolution of 1920 x 1200 with circum-aural headphones in a room with the lights turned off. The graphics settings were set at their maximum and the game ran at a constant 60 frames per second. The participants were asked to perform the structural knowledge assessment with the *PCKNOT* software. Then, the knowledge test was administered. Before playing the game, the participants were given instructions about *Code Red Triage* and were informed about its goal. Nothing was revealed to them about the condition they took part in. Playing the game from start to finish took each participant at most 25 minutes: a few minutes for the entry level, a few more for the hallway part and a maximum of 17 minutes was allowed for the metro platform part, in which the triages took place. The scores participants reached in the game gave information about their performance (see also section 2.3). Directly after the participants finished playing the game, they were asked to fill out the engagement questionnaire. They were then asked to do the structural knowledge assessment and knowledge test as before, but with the questions in a different order. Finally, the participants were thanked for their cooperation and they received a coupon for their work. An overview of the procedure can be seen in Figure 4.



Figure 4. Procedure of the experiment

## 4. Results

### Engagement

The mean scores and standard deviations of the engagement questionnaire are mentioned in Table 1. An ANOVA showed no significant effect of condition on the ITC-SOPI engagement questionnaire,  $F(1,26) < 1$ .

### Learning Efficiency

There are several ways to determine whether learning was more efficient in the adaptive condition. A reliable measure for efficiency is to divide the posttest scores of the participants by the number of victim cases triaged, giving us an indication of how much the participant has learned per unit of instruction, and whether this would be higher in a game that adapts the information presentation to the player's proficiency. Another way would be to divide learning performance by total time spent playing the game. However some players navigate more efficiently than others towards the platforms etc, which

blurs what we want to measure. We therefore decided to use learning performance divided by the number of cases triaged, as a purer measure of learning efficiency.

**Table 1:** Mean engagement and efficiency scores on knowledge test and structural knowledge assessment (sd).

	<i>Control Condition</i>	<i>Adaptive Condition</i>
<i>Engagement (1-5)</i>	3.63 (.33)	3.66 (.45)
<i>Knowledge test</i>	.57 (.19)	1.02 (.30)
<i>Structural knowledge assessment</i>	.015 (.004)	.028 (.019)

An ANCOVA with the pretest as covariate, condition as fixed factor and posttest score divided by the total number of victims triaged as dependent variable showed that condition had a significant effect on both the knowledge test ( $F(1,25) = 21.98, p < .001, d = 1.81$ ) and the structural knowledge assessment ( $F(1,25) = 5.05, p < .05, d = .89$ ). The means on these relative measures and standard deviations of these tests are listed in Table 1.

*In-game score*

The total in-game score was significantly higher for the control condition ( $M = 777.7, SD = 321.2$ ) than for the adaptive condition ( $M = 316.4, SD = 107.8, F(1,26) = 25.95, p < .001$ ), however this more or less follows from the result that participants triaged significantly less victims in the adaptive condition.

**5. Conclusion and discussion**

We hypothesized that a serious game that dynamically adapts its challenge, or complexity presentation, to quick learners could make a serious game more engaging and more efficient. The first part of the hypothesis was not confirmed, while the second part was confirmed; participants in the adaptive game version learned significantly more per victim case than in the control condition, and were therefore more efficient.

We found no difference in the engagement ratings. If the improved learning per unit of instruction was due to less disengagement from the task, one would expect this to appear from the results of the engagement questionnaire. We propose four explanations why we did not find a difference in engagement.

Firstly, when participants had to appraise their engagement just after playing the game, they lacked knowledge of the other condition and thereby a reference

point. The intervention itself may be too small next to all the other determinants of engagement, such as the game’s setting, world, expectations, control interface, et cetera, to show up as a difference on the rating scale, but the adaptive version may still be preferred when the conditions were placed side by side. A second explanation could be related to the fact that we only asked participants to appraise their engagement *after* the game. It is unclear whether a continuous measurement of a participant’s engagement, for instance with an infrared camera as in the research by [24], as we mentioned in the introduction, would have resulted in higher ratings throughout the game in the adaptive version. Thirdly, people may play games for different reasons; a higher challenge could lead to higher engagement in some players, whereas it has the opposite effect on others. Lastly, and perhaps as a result of the previous explanation, we found that the homogeneity of the engagement questionnaire (Cronbach’s alpha) was low. Perhaps this measurement problem contributed to the fact that we did not find an effect of engagement.

We saw that participants learned more per victim case in the adaptive condition compared to the control condition. It could be that the moment a participant grasps the procedure to resolve a victim case pertaining to a certain tier, the information presented in the following victims in that tier is redundant, at least to a point that it does not improve learning of the procedure anymore, making the adaptive version more efficient.

In order to determine whether the adaptive condition not only made learning the instructional material more efficient, but also leads to deeper learning [40], other experiments should be set up such as e.g. a study where learning is also measured after a longer delay or with transfer tasks. However, some corroboration may be found in the structural knowledge assessments. They point to deeper learning in the adaptive condition.

One last observation concerns the relation between engagement and learning; the results found indicate that an increase in engagement does not seem necessary to enhance learning efficiency. Also the correlation between engagement and learning efficiency appeared to be low and not significant ( $p > .05$ ) for both groups of participants. However, for this finding too, the same remarks as before should be made concerning the measured engagement of players.

All in all, a rather simple alteration of a serious game where it dynamically adapts the presentation of complexity to the player’s performance and thereby its challenge has been shown to markedly improve the efficiency thereof. This is a promising result for serious games developers that worry about the comparative efficiency of their game, as well as for researchers interested in improving games with the aid of more sophisticated adaptation engines. It can also be a useful result for entertainment game developers, as many games need to incorporate tutorial levels that are necessary for players to understand the game, but are not a lot of fun to play, especially upon repeated playthroughs. A dynamic

adaptive version that adapts to the player's proficiency could greatly speed up these mandatory instructional sequences and (possibly) make them more challenging.

#### Future research

Above we already mentioned two limitations to our study, viz. that it is impossible to conclusively state whether dynamically adapting to the player's performance only resulted in more efficient instruction, or also in deeper learning, and that it is unclear whether participants differed in engagement during gameplay. In addition, another limitation of our experimental setup that warrants future research is that we did not measure retention over longer time periods. Participants in the adaptive game version received less practice and consequently less opportunity to internalize the information. Therefore there is a real possibility - or even danger - that the participants in the dynamic adaptive condition remember less of the instruction after several weeks. Regarding dynamic adaptation itself, in this study we did make some specific choices during the design and implementation process. We focused on the nature of the non-player characters and let them vary in number of steps needed to perform a correct triage. Several alternatives are open for continued research to the role of dynamic adaptation. For instance, the set of buttons for executing the triage actions could be adapted, that is, starting simple and increasing over time, depending on performance. Or the feedback given to players could be adapted, e.g. stating more or less explicitly what went right and what went wrong while performing the triage [41], or explanatory feedback could be included, particularly in early learning phases [42].

Finally, another option for making training procedures adapt themselves to participants is the notion of *adaptability*. In this form of offline adaptation participants indicate themselves what direction they want to practice and what part of the procedure they want to repeat. These are questions that still need to be examined in the future.

## Acknowledgements

This research has been supported by the GATE project (<http://gate.gameresearch.nl/>), funded by the Netherlands Organization for Scientific Research (NWO).

## References

- [1] Sitzmann, T. (2011) A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology*: 489-528.
- [2] Wouters, P.J.M., van Nimwegen, C., van Oostendorp, H. and van der Spek, E.D. (2013) A Meta-analysis of the Cognitive and Motivational Effects of Serious Games. *Journal of Educational Psychology*, **105**: 249-265.
- [3] Wouters, P.J.M., and van Oostendorp, H. (2013) A meta-analytic review of the role of instructional support in game-based learning. *Computers & Education*, **60**: 412-425.
- [4] Moreno, R. and Mayer, R. (2007) Interactive multimodal learning environments. *Educational Psychology Review*, **19**(3): 309-326.
- [5] Saxby, D. J., Matthews, G., Hitchcock, E. and Warm, J. S. (2007) Development of active and passive fatigue manipulations using a driving simulator. In *Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting* (Santa Monica, CA: Human Factors and Ergonomics Society), 1237-1241.
- [6] Paas, F., Renkl, A. and Sweller, J. (2004) Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science*, **32**(1): 1-8.
- [7] Shute, V. J. (2011) Stealth assessment in computer-based games to support learning. In S. Tobias & J.D. Fletcher (eds.), *Computer games and instruction* (Charlotte, NC: Information Age), 503-524.
- [8] Csikszentmihalyi, M. (1988) The flow experience and its significance for human psychology. In M. Csikszentmihalyi & I.S. Csikszentmihalyi (eds.), *Optimal experience: Psychological studies of flow in consciousness* (Cambridge, MA: Cambridge University Press), 15-35.
- [9] Shernoff, D. J., Csikszentmihalyi, M., Shneider, B. and Shernoff, E. S. (2003) Student engagement in high school classrooms from the perspective of flow theory. *School Psychology Quarterly*, **18** (2): 158-176.
- [10] Sweetser P. and Wyeth, P. (2005) GameFlow: a model for evaluating player enjoyment in games. *Computers in Entertainment*, **3**(3): 1-24.
- [11] Cowley, B., Charles, D., Black, M. and Hickey, R. (2008) Toward an understanding of flow in video games. *Computers in Entertainment* **6**(2): 1-27.
- [12] Webster, J., Trevino, L. K. and Ryan, L. (1993) The dimensionality and correlates of flow in human-computer interactions. *Computers in Human Behavior*, **9**(4): 411-426.
- [13] Liu, M., Horton, L., Olmanson, J. and Toprac, P. (2011) A study of learning and motivation in a new media enriched environment for middle school science. *Education Technology Research Development*, **59**: 249-265.
- [14] Lopes, R. and Bidarra, R. (2011) Adaptivity challenges in games and simulations: a survey. *IEEE Transactions on Computational Intelligence and AI in Games* **3**(2): 85-99.
- [15] Walker, J. (2009) Left 4 Dead 2: Exclusive RPS Hand on Preview. *Rock Paper Shotgun, PCGaming since 1873*. Online available: <http://www.rockpapershotgun.com/2009/06/01left-4-dead-2-exclusive-rps-preview>.
- [16] Brathwaite, B. and Schreiber, I. (2009) *Challenges for game designers*, (Boston, MA: Charles River Media).
- [17] Bakkes, S., Spronck, P. and van den Herik, J. (2009) Rapid and reliable adaptation of videogame AI. *IEEE Transactions Computational Intelligence and AI in Games*, **1**(2): 93-104.
- [18] Roberts, D.L. and Isbell, C.L. (2008) A survey and qualitative analysis of recent advances in drama management. *Int. Trans. Sci. Appl.*, **3**(2): 61-75.

- [19] Niehaus, J. and Riedl, M.O. (2009) Scenario adaptation: an approach to customizing computer-based training games and simulation. In *Proceedings AIED Workshop Intell Edu. Game* (Brighton, UK), 89-98.
- [20] Lavie, T., & Meyer, J. (2010) Benefits and costs of adaptive user interfaces. *International Journal of Human-Computer Studies*, **68**, 508-524.
- [21] Rocksmith [Computer software]. San Francisco, CA: Ubisoft.
- [22] Mario Kart [Computer software]. Kyoto, Japan: Nintendo.
- [23] Need for Speed [Computer software]. Redwood City, CA: Electronic Arts.
- [24] Yun, C., Shastri, D., Pavlidis, I. and Deng, Z. (2009) O' game, can you feel my frustration?: Improving user's gaming experience via stresscam. In *Proceedings of the 27th international conference on Human factors in computing systems (CHI '09)* (New York: ACM), 2195-2204.
- [25] The Elder Scrolls 4: Oblivion [Computer software]. Rockville, MD: Bethesda.
- [26] Shute, V. J., Ventura, M., Bauer, M. I. and Zapata-Rivera, D. (2009) Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In U. Ritterfeld, M. Cody, & P. Vorderer (eds.), *Serious games: Mechanisms and effects* (New York, NY: Routledge), 295-321.
- [27] Bostan, B. and Ogut, S. (2009) Game challenges and difficulty levels: lessons learned From RPGs. In *International Simulation and Gaming Association Conference*. Retrieved from [http://www.silentblade.com/presentations/Bostan\\_Ogut\\_Full\\_Paper.pdf](http://www.silentblade.com/presentations/Bostan_Ogut_Full_Paper.pdf)
- [28] Bailey, C. and Katchabaw, M. (2005) An experimental testbed to enable auto-dynamical difficulty in modern video games. In *Proceedings of the 2005 GameOn North America Conference*, 18-22.
- [29] Half-Life 2 [Computer software]. Bellevue, WA: Valve Corporation.
- [30] Van der Spek, E. D., Wouters, P. and van Oostendorp, H. (2011) Code Red: Triage Or COgnition-based DEsign Rules Enhancing Decisionmaking TRaining In A Game Environment. *British Journal of Educational Technology*, **42**(3): 441-455.
- [31] Van der Spek, E.D., van Oostendorp, H. and Meyer, J.-J.Ch. (2013) Introducing surprising events can stimulate deep learning in a serious game. *British Journal of Educational Technology*, **44**(1): 156-169.
- [32] Van der Spek, E. D. (2011) Experiments in serious game design: a cognitive approach. SIKS Dissertation Series, Utrecht University, The Netherlands.
- [33] Ferguson, R. (2012) Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, **4**(5/6): 304-317.
- [34] PCKNOT [Computer Software]. Las Cruces, NM: Interlink Inc.
- [35] Schvaneveldt, R. W., Durso, F. T. and Dearholt, D. W. (1985) Pathfinder: Scaling with network structures (Memorandum in Computer and Cognitive Science, MCCS-85-9, Computing Research Laboratory). Las Cruces: New Mexico State University.
- [36] Gomez, R. L., Hadfield, O. D. and Housner, L. D. (1996) Conceptual maps and simulated teaching episodes as indicators of competence in teaching elementary mathematics. *Journal of Educational Psychology*, **88**(3): 572-585.
- [37] Day, E. A., Arthur Jr., W. and Gettman, D. (2001) Structural knowledge assessment: Comparison of referent structures. *Journal of Educational Psychology*, **86**(5): 1022-1033.
- [38] Wouters, P.J.M., van der Spek, E.D. and van Oostendorp, H. (2011) Measuring learning in serious games: a case study with structural assessment. *Educational technology research and development*, **59**(6): 741-763.
- [39] Lessiter, J., Freeman, J., Keogh, E. and Davidoff, J. (2001) A cross-media presence questionnaire: The ITC-sense of presence inventory. *Presence: Teleoperators and Virtual Environments*, **10** (3): 282-297.
- [40] Graesser, A.C., Chipman, P., Leeming, F. and Biedenbach, S. (2009) Deep learning and emotion in serious games. In U. Ritterfeld, M. Cody & P. Vorderer (eds.), *Serious Games: Mechanisms and Effects* (New York: Routledge), 81-100.
- [41] Erhel, S. and Jamet, E. (2013) Digital game-based learning: impact of instructions and feedback on motivation and learning. *Computers & Education*, **67**: 156-167.
- [42] Butler, A.C., Godbole, N. and Marsh, E.J. (2013) Explanation Feedback is Better Than Correct Answer Feedback for Promoting Transfer of Learning. *Journal of Educational Psychology*, **105**(2): 290-298.