Eye Tracking in Game-based Learning Research and Game Design

[1]Kristian Kiili, [2]Harri Ketamo, [3]Michael D. Kickmeier-Rust
[*1]Tampere University of Technology, Pori, Finland, kristian.kiili@tut.fi
[2]*Satakunta University of Applied Sciences, Pori, Finland, harri.ketamo@samk.fi*
[3]*Graz University of Technology, Graz, Austria, michael.kickmeier-rust@tugraz.at*

Abstract

The challenge of educational game design is to develop solutions that please as many players as possible, but are still educationally effective. Educational game designers need to understand how users interact with different types of user interfaces and how this interaction affects users' educational experiences and effectiveness of learning. In this research we utilized eye tracking method in order to explore the game-based learning process and the perception of user interfaces of four educational games. Based on perceptual data we evaluated the playing behavior of 43 Finnish and Austrian children aged from 7 to 13. The results indicated that players' perception patterns varied a lot and some players even missed relevant information during playing. The results showed that extraneous elements should be eliminated from the game world in order to avoid incidental processing in crucial moments. Animated content easily grasps player's attention, which may disturb learning activities. Especially low performers and inattentive players have difficulties in distinguishing important and irrelevant content and tend to stick to salient elements no matter of their importance for a task. However, it is not reasonable to exclude all extraneous elements because it decreases engagement and immersion. Thus, balancing of extraneous and crucial elements is essential. Overall, the results showed that eye tracking can provide important information from game based learning process and game designs. However, we have to be careful when interpreting the perceptual data, because we cannot be sure if the player understands everything that he or she is paying attention to. Thus, eye tracking should be complemented with offline methods like retrospective interview that was successfully used in this research.

Keywords: *Serious Games, Human-Computer Interaction, Play Testing; Eye Tracking*

1. Introduction

Why there is no big success, like Angry Birds or Clash of Clans, in educational games? Because educational games too often suck as games! In academic world we focus on proving our hypothesis and we tend to forget the user experience. This is fine as long as we play with university prototypes, but when going to consumer markets, we should focus on users instead of only proving our excellence. In consumer markets the users are used to get games that are easy to use, engaging and entertaining. As long as game-based learning community claims to have excellent products but users don't understand them fast enough, there will be no big success stories in educational games market.

The aim of game designers is to create appealing experiences to players. Thus, games can be seen only as artefacts that arouse experiences [1]. Dewey [2] has stated that the experience is a result of interplay between the present situation and person's prior experiences. Consequently, players do not have identical playing experiences, but each player's experience is totally unique. The challenge of educational game design is to develop solutions that please as many players as possible, but are still educationally effective. In particular, educational game designers need to understand how users interact with different types of user interfaces and how this interaction affects users' educational experiences and effectiveness of learning.

In research reported in this paper we utilized eye tracking method in order to explore the game-based learning process as well as the perception of game user interfaces of four educational games. According to [3] eye tracking studies can be either top-down or bottom-up. Top-down studies are based on cognitive theories whereas bottom-up approaches analyze the data without any theories. This is a top-down study in which a problem-based gaming model [4] is used as a theoretical framework.

Kiili [4] has proposed a problem-based gaming model that aims to describe learning mechanisms in educational games at an abstract level. According to the model game based learning process is an iterative process in which a player tries to overcome the challenges that the game provides and to adopt the use of the game. Generally, the model considers a game as a big problem that is composed of smaller causally linked challenges [5]. Furthermore, the model emphasizes the meaning of feedback and reflective thinking. According to [6], reflection is a human activity in which people recapture their experience, think about it, mull it over and evaluate it. During playing users actively construct their mental models based on the feedback that the game world provides (consequences of players' actions). In ideal situation this leads to mental growth that helps player to manage with more challenging situations in the game as well as apply the learned knowledge outside the game.

The quality and the form of the feedback influence the learning outcomes. Thus, [7] have emphasized the meaning of cognitive feedback in educational games. The aim of cognitive feedback is to grasp player's attention and focus it on essential learning content. In other words, cognitive feedback aims to stimulate player to reflect on his experiences and tested solutions in order to further develop mental models, validation of hypothesis and formation of new playing strategies. The model distinguishes also a gulf of evaluation (derived from [8]) that refers to players' problems with perceiving the consequences of their actions in the game. Inadequate perception of consequences usually leads to failure of reflective thinking and adoption of the user interface and may even lead to misconceptions.

The overall aim of this research is to study the meaning of feedback in educational games and consider the usefulness of eye tracking method in game based learning research and educational game design. To be more precise, based on perceptual data we study the playing behavior, adoption of games' user interfaces and effectiveness of cognitive feedback in four different educational games. In the first three studies a player teaches a virtual pet, a teachable agent, which can reason, based on how it is taught. Previous studies on games involving teachable agents have provided clear evidence of learning gains [7][9]. The current research is designed to take a closer look at the learning process and adoption of the user interfaces focusing on players' perceptual processes during playing. Through the fourth game experiment the usefulness of eye tracking measures in game based learning research are evaluated by investigating the relationships of learning performance, gender and eye movements.

First, we shortly present the eye tracking method and previous work about eye tracking in learning material research. After that we present the results of four eye tracking studies focusing on game based learning. Finally, the conclusions about eye tracking in game based learning research and educational game design are presented.

2. Eye Tracking in Game-based Learning Research

Observing users' eye movements has a long tradition in usability field as well as in psychology. In recent years, the adoption of eye tracking in various research fields has increased. Eye tracking is based on identifying fixations (processing of attended information with stationary eyes) and saccades (quick eye movements occurring between fixations without information processing). Fixations usually last approximately 200-500 milliseconds depending on the task. Thus, when a person interacts with a visual environment, he or she makes a sequence of fixations separated with saccades. In eye tracking method fixations and saccades are used to index mental processes that are on-going when person interact with a visual environment. Research relies on an assumption according to which a person attends to and process information that he or she is currently looking at. [10] reminds that this assumption holds only if the visual information is relevant to the task at hand. Furthermore, although eye tracking can reveal what a person perceives, it does not tell whether or not the person comprehends the information that he or she was looking at. Thus, eye tracking should be complemented with offline measures such as retrospective comprehension test or retrospective think aloud methods.

According to [10] eye tracking has quite a long history in reading research and recently it is applied also on multimedia learning (e.g. [11][12]). For multimedia learning research eye tracking has provided more detailed information about found multimedia principles and the ways how different people process certain materials. So far, the use of eye tracking method in game based learning research and educational game design has been minor [7]. However, for example [13] have shown that eye tracking can be successfully applied to measure the quality of serious games. Based on their eye tracking results [10] have argued that the layout of the game plays bigger role than the content in capturing user attention. In general, for game based learning research eye tracking can provide new knowledge about how learning happens in games, what game elements can be used to enhance learning, how to focus player's attention to important game elements, how to avoid evaluation gulfs etc. [7]. Such knowledge can help educational game designers to develop higher quality educational games.
One important analysis tool in eye tracking is Areas of Interests (AOI). AOIs are areas of a display or visual environment that is of interest to the researcher and thus predefined by them. AOI analysis is used to quantify gazed data within a defined region of the visual stimulus. The number of fixations on such particular display element indicates the importance of that element. Consequently more important display elements will be fixed more frequently and longer. Regarding the evaluation of learning games, such information is crucially important since it provides very clear indications of which elements on the screen are attended (sufficiently) and, which elements may be missed during playing.

3. AnimalClass Studies

In this section we report the results of two eye tracking studies, in which two different *AnimalClass* games were used as test-beds. Both eye tracking and retrospective interview methods were used to study how the cognitive feedback affects the game based learning process.

3.1. AnimalClass Games and Cognitive Feedback

All AnimalClass games rely on a learning-by-teaching approach. In AnimalClass games a player teaches a virtual pet, a teachable agent, which can reason based on how it is taught. The task of the player is to teach his or her agent the subject of the game, for example in this study mathematics and geography of Europe. At the beginning of the game, the agent does not know anything. Its mind is an empty set of concepts and relations. The player has complete freedom to teach the agent what he or she wants, even wrongly. In AnimalClass games teaching is always based on statements constructed by the player. In mathematics game a player forms equations and in the geography game a player forms "Which does not belong to the group?" type of statements related to European map (Figure 1) for his or her agent. The agent answers the statements according to its previous knowledge. If there is no previous knowledge, it will guess. The player then tells the agent if the answer was correct or not, and based on this, the agent forms relations between concepts. When the agent achieves a concept structure of a certain size, it can start to conclude.

Figure 1. Teaching in AnimalClass games

The previous research has shown that teaching others is a powerful way to learn (e.g. [9][15]). According to [16], the aspects that make teaching beneficial are the structuring of knowledge, taking responsibility, and reflecting. Furthermore, teaching is motivating because the player is not responsible only for his own learning, but also his agent's learning. Reflection on one's own teaching and the agent's performance aids structuring knowledge and perceiving progress toward goals. Teaching in AnimalClass games, supports learning in two ways. Firstly, constructing a question requires knowledge about the subject. If the player does not have enough knowledge, player is encouraged to discuss the problem for example with his friends or search information from Internet. Secondly, evaluating the answer of the agent supports reflective thinking: "What have I taught to the agent? Why did the agent answer in this way? What should I do next? What happened when I evaluated my agent's answer?"

To facilitate reflection and learning AnimalClass games provide cognitive feedback for the player in several ways. First of all, the agent's gestures illustrate the certainty of its knowledge. Three levels of certainty are included as illustrated in Figure 2. Based on the agent's gestures, a player can figure out what his agent knows and what to teach next. It is noteworthy that certainty is based on the beliefs of the character and it is not determined based on facts. Secondly, a brain icon (see Figure 1) describes the quality of the agent's conceptual structure compared to formal goals. The brain gets bigger if the quality of the conceptual structure increases and smaller if the quality decreases. Beside the brains there is also a number that indicates how many percent of the learning content the agent knows. If the overall conceptual structure is totally wrong, the brain is replaced with a cactus icon in order to show the player that he/she is doing something completely wrong. Thirdly, a player can send his/her agent to a competition (Figure 2, right). In the competition, the agent competes in a quiz against someone else's agent that has been taught by a real person (possibly a friend or a classmate). The competition is completely based on previous teaching. The role of the player is to observe the successes and failures of his/her agent in order to grasp the agent's current skills and misconceptions. Finally, in geography game player can visualize his agent's conceptual structure in the form of European map. Only the elements, for example countries, that agent knows are shown in the map.

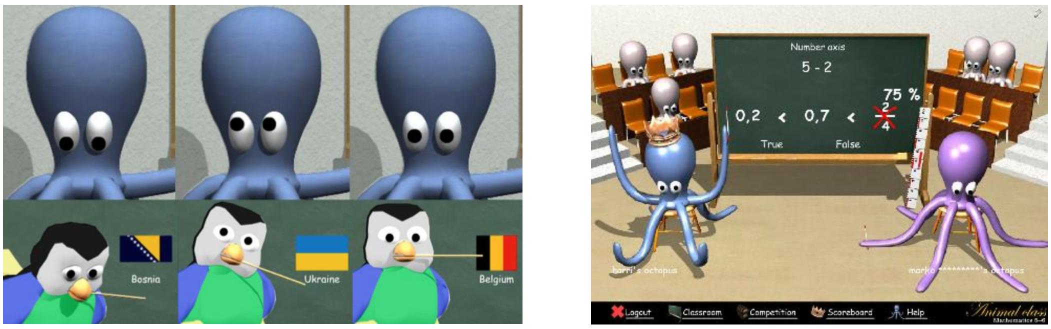

Figure 2. Left: Representation of certainty in mathematic (eyes) and geography (body) games (Left: guessing, Middle: reasoning, Right: knows)
Right: Competition in progress

3.2. The Eye Tracking Device

A Tobii T60 eye tracker with 17 inches display was used to record players' eye movements (Figure 4). Large freedom of head movement allows players to behave naturally during the playing session. Furthermore, Tobii T60 has no visible or moving "tracking devices" that might affect the subject. The Tobii software was used to record the eye movements, operate the calibration process, and replay the recordings of participants' eye movements.

3.3. Study 1 - Mathematics

Participants were ten to eleven years old Finnish primary school pupils (N=14). Participants were randomly selected from one class of 27 pupils. The gender distribution was even. Participants had studied the content of the game, fractions, approximately one year ago, but they had never played AnimalClass games before.

The participants were tested one by one. First, the idea, the story of the game and instructions how to use the user interface of the game were told to a participant. Second, the participant answered for four background questions. Third, the eye tracker was calibrated and the participant started playing - participant was asked only to teach the octopus not to compete. Participant played the game approximately 5 to 8 minutes depending on his or her playing speed. After the playing phase retrospective interview phase followed. In practice the researcher and the participant watched a replay of the gaming session with gaze plots. The meaning of the gaze plots was told to the participant. The researcher stopped the recording in crucial places and asked questions from the participant. For example, did you notice the eye movements of your octopus; do you know what the octopus's eye movements mean?

3.3.1. Results

Players made approximately 20 equations to their octopus (M = 20.83, SD = 4.98). In mathematics game we were interested in how often and when a player focused attention on his or her octopus's brains, eyes and classroom's binders that reflect the amount of teaching. Table 1 shows the time to first fixation to brains, eyes and binders. As we can see the times vary a lot between players. Generally, brains (M = 36.67, SD = 64.34) and eyes (M = 28.13, SD = 17.75) are noticed quite quickly. However, binders catch players' attention much slower (M = 92.77, SD = 57.02). This is not a problem because binders' meaning is not important from learning point of view. Furthermore, the time to first fixation to brains did not correlate with the success in the game (r = .28, p = .38). Player's success in the game was determined based on the quality of pet's brains (conceptual structure).

Table 1. Time to first fixation in seconds

	N	Min	Max	M	SD
Brains	12	3.07	224.74	36.67	64.34
Eyes	12	2.66	69.05	28.13	17.75
Binders	12	38.06	252.30	92.77	57.02

Table 2 shows the fixation counts to brains, eyes, binders and brains after evaluation of octopus's answers. Again standard deviations are quite big that indicates that players' strategy to use brains in teaching varied a lot. Interesting is that only approximately half of the fixations made to the brains was after evaluation of octopus's answers (M = 6.75, SD = 4.13). After the evaluation of an answer player could have used brains to evaluate whether or not he or she had taught his pet correctly. This may have affected the finding that fixations to brains did not correlate with the game performance. Most fixations were made on octopus's eyes (M = 11.67, SD = 11.08) and least on binders (M = 3.75, SD = 3.38).

Table 2. Fixation counts

	N	Min	Max	M	SD
Brains altogether	12	5	28	14.33	8.11
Brains after task	12	2	13	6.75	4.13
Eyes	12	3	42	11.67	11.08
Binders	12	1	12	3.75	3.388

Although fixations were made more on the eyes, the fixations to brains were longer (M = 6.35, SD = 3.26) than on eyes (M = 3.96, SD = 3.84) as the table 3 shows. Based on the basic assumption of eye tracking theory this indicates that the meaning of the brains was hard to understand. However, the retrospective interview indicated differently. Most of the players did not understand or remember what the octopus's different eye movements meant and they could not utilize them strategically. On the other hand, most of the players knew what the changes in brains meant and the length of fixations on brains can be explained with processing of brains' state. The eyes were passed quickly because only few players knew how to interpret them and they started to ignore them. This indicates that the idea to use eyes as a feedback channel did not work well (gulf of evaluation) and players missed information that was important from learning point of view.

Table 3. Fixation lengths in seconds

	N	Min	Max	M	SD
Brains	12	2.38	12.58	6.35	3.26
Eyes	12	0.83	14.24	3.96	3.84
Binders	12	0.28	6.16	1.62	1.63

3.4. Study 2 - Geography

Participants were 11 to 12 years old Finnish primary school pupils (N=16). Participants were from one class. The gender distribution was almost even. Almost all participants (14/16) had played the geography game before and thus they knew how the game works. The participants were tested one by one. First, the idea, the story of the game and instructions how to use the user interface of the game were told to a participant. Second, the participant answered for four background questions. Third, the eye tracker was calibrated and the participant started playing - participant was asked only to teach the octopus not to launch a competition. Participant played the game approximately 5 minutes. After the playing phase retrospective interview phase followed. In practice the researcher and the participant watched a replay of the recorded gaming session with gaze plots. The meaning of the gaze plots was told to the participant. The researcher stopped the recording in crucial places and answered questions from the participant. For example, did you notice the gestures of your bird when it answered to questions, do you know what the bird's gestures mean?

3.4.1. Results

Players made approximately 17 questions to their bird (M = 17.00, SD = 6.53). In geography game we were interested in how often and when player focus attention on his or her bird's brains, classroom's globe and classroom's binders that reflect the amount of teaching. The fixations to the spinning globe were tracked in order to study how much the extraneous animations catch players' attention. The certainty gestures of the bird could not be tracked, because the bird answers to questions by moving below the flag that it answer. Table 4 shows the time to first fixation to brains, globe and binders. As we can see the times vary a lot between players. Brains (M = 46.01, SD = 38.18) and globe (M = 44.54, SD = 65.25) are noticed quicker than binders (M = 77.93, SD = 75.42). It was surprising that it took quite long from several players to pay attention on brains, although they have played the game before and they knew the meaning of the brains. In fact, the time to first fixation to brains correlated with the success in the game (r = .46, p = .049). Player's success in the game was determined based on the quality of player's pet's brains (conceptual structure). The relation of noticing the brains early and success in the game is much stronger than in the first study, because there was more variance in the performance of geography game players (M = 9.86, SD = 13.09) than in mathematics game players (M = 6.83, SD = .88).

Table 4. Time to first fixation in seconds

	N	Min	Max	M	SD
Brains	14	4.38	109.54	46.01	38.18
Globe	14	1.39	218.16	44.54	65.25
Binders	14	36.58	262.18	77.93	75.42

Table 5 shows the fixation counts to brains, globe, binders and brains after evaluation of birds' answers. Again standard deviations are quite big that indicates that players' strategy to use brains in teaching varied a lot. Less than half of the fixations made to the brains were made after evaluation of octopus's answers (M = 4.71, SD = 3.71). Most fixations were made on brains (M = 11.67, SD = 11.08) and least on binders (M = 3.75, SD = 3.38). The high amount of fixations on the globe is surprising. Although players knew that the globe does not have any function in the game, but is only decoration, they paid a lot attention to it. The replays of players gaze plots revealed that the globe caught players' attention also in crucial times – moments after tasks, when players should have paid attention on brains. Same is true also for binders. The game designers' needs to decide whether the extraneous materials engage players so much that it is reasonable to include them into the game or not. Every element in the game world has it cognitive price. In fact, research has shown that multimedia presentations are more effective when irrelevant material is excluded [17].

The fixations to brains were longer (M = 5.12, SD = 5.71) than on globe (M = 1.63, SD = 1.47) and on binders (M = 1.27, SD = 1.64) as the table 6 shows. Most of the players knew what the changes in brains meant and the length of fixations on brains can be explained with processing of brain's state. The retrospective interview also revealed that most of the players comprehend the 'knows' and 'guessing' states of the representations of certainty. However, the reasoning state was unclear for almost all players. It seems that the graphical presentation of these states were more clearly animated than in mathematics game in which only eyes of the octopus were used as an indicator.

Table 5. Fixation counts

	N	Min	Max	M	SD
Brains altogether	14	0	38	11.71	11.34
Brains after task	14	0	11	4.71	3.71
Globe	14	1	20	5.29	4.76
Binders	14	0	17	3.57	4.48

Table 6. Fixation lengths in seconds

	N	Min	Max	Mean	SD
Brains	14	0	19.89	5.12	5.71
Globe	14	0.23	5.80	1.63	1.47
Binders	14	0	6.15	1.27	1.64

4. Math Elements Study

From educational outcome point of view, AnimalClass games were a success. Furthermore, the game mechanics and AI was awarded in several educational and games industry contests. However, the game never made a worldwide breakthrough in everyday classroom use, because the gameplay was too hard to understand fast enough. In this context we have to understand the realities in a) classroom management and b) casual gaming that are totally different to academic research in laboratory settings: First of all, in a classroom teacher expects he/she can start a lesson in minutes. If it takes too long to start, teacher chooses a different pedagogical approach. Secondly if the game requires specific 3rd party plugins we can be sure that it will always cause problems in average classroom settings, which will increase teachers' rejection. From casual gaming point of view, kids are used to games that they understand in seconds. Even couple of minutes of learning curve is too long and even a good and an effective educational game can fail in the market.
The Math Elements game is based on similar AI solutions but the game mechanics and the storyline are developed to overcome the main weaknesses of the AnimalClass games. Especially the learning curve has been radically shortened. Also the technical complexity has been minimized in order to avoid teacher's frustration with non-stabile technology in classroom.
The background story of the game is that in the near future mice can get cheese only by getting through mathematics labyrinths faster than cats can. Player's task is to teach necessary skills for their pets, mice. When pets have enough skills player can send it to labyrinth to survive on its own and hunt for cheese. When starting teaching, the pet goes to classroom (Figure 3, left). In the classroom teacher (owl) asks questions from the mouse. Player can help

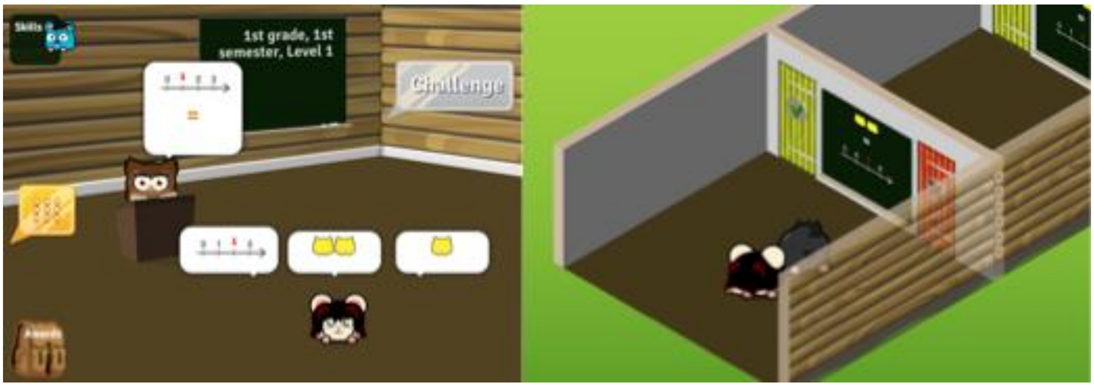

Figure 3. The classroom and the labyrinth race

his or her pet by pointing answers to the pet. The pet learns exactly according the teaching. If a player teaches the pet correctly, pet learns correctly and vice versa. After the player has taught enough conceptual relations for his or her pet, a challenge icon appears on the screen. By clicking the icon, player sends his or her pet into labyrinth to compete against the cat. In labyrinth the pet is on its own and player's task is to observe how it manages. In the labyrinth (Figure 3, right), both characters pic doors according to their taught knowledge (green door refers to right answer to presented task and red wrong answer). During the labyrinth player can observe what to teach more. If mouse wins, the level is completed and next level becomes playable.
Technically the Math Elements version that was used in this study is an online game with client-server architecture. Math Elements game clients are optimized for tablets and smart phones. Technically they are HTML5 clients and apps for iOS and Android. Game mechanics run in server side, built in Google Apps Engine. The mechanics enables that game characters can compete against any other character any time, no matter if the opponent is really online, because all behavior is always available in online. In the test version of the game, there were no sounds at all that naturally affects how the game was experienced.

4.1. Method

The Finnish experiment was done in May 2012 involving 23 first class pupils. The research procedure was following: The class were divided into 2-4 pupils groups and each group received 1-

2 iPads. Pupils started to play the game with minimum instructions. No identities were written down, nor any other notes that makes recognizing the person possible. Playing time was approximately 90 minutes. Researchers were allowed to assist pupils during the game play only in verbal way. After the gameplay, pupils received URL for the game to continue the gameplay at home. Finally eight pupils (n = 8) were randomly selected to eye tracking part. The main aim of the study was to deepen the results gathered with observation and interviews. The primary focus of the eye tracking study was on visual implementation of the game. The same Tobii eye tracking device and software that were used in AnimalClass studies were used.

Eye tracking was used to explore two research questions. 1) Does the player notice the gestures of the game characters (gestures are used to give feedback of player's performance). In order to study the gestures two areas of interest (AOIs) were formed: The gestures of the owl (teacher in the classroom) and the gestures of the mouse (player's character) were monitored (see Figure 4). In practice, AOI analysis was used to quantify the eye movement data related to owl and the mouse. In this study we made the AOI analysis based on the fixation counts and fixation lengths. In order to deepen the analysis the gaze sequences were qualitatively analyzed. 2) Does the passive viewing of labyrinth race activate cognitive processes? In other words, does the player notice and try to solve the math tasks when watching his or her mouse racing against the cat in the labyrinth. The gaze data of the race was analyzed on qualitatively basis because the dynamic nature of the race did not enable AOI analysis.

Figure 4. Recording participant's gaze plots with Tobii T60 eye tracker

The participants (6 boys and 2 girls) were tested one by one and they were already familiar with the game. First, the eye tracker was introduced to participants. Second, the eye tracker was calibrated. Third, the participant answered some background questions and after that they started playing the game. Participant played the game approximately 5 minutes and finally they watched a race against the cat in the labyrinth. After the playing phase retrospective interview phase followed. In practice the researcher and the participant watched a replay of the recorded racing session with gaze plots. The meaning of the gaze plots was told to the participant. The researcher stopped the recording in crucial places and asked questions from the participant. For example, did you notice the tasks on the black boards or what does the colors of the doors mean?

4.2. Results

During the eye tracking session players taught approximately 15 questions to their mouse. In the game's classroom players could concentrate on presented tasks and their game characters as the hot spot map shows (Figure 5). Table 7 shows that players paid much more attention to their mouse character than to the owl (teacher). As we can see the fixation count varies a lot between players. Furthermore, the fixations on the mouse were longer (M = 0.41) than on the owl (M = 0.26). However, when interpreting the hot spot maps and fixation counts on AOIs, we have to remember they give only superficial results about the game's user interface. As discussed in AnimalClass section in the investigation of fixation patterns is crucial in dynamic games. The gaze replays indicated that most of the players' fixations were made when the owl or the mouse provided cognitive feedback for the player. On the other hand, the retrospective interview revealed that although some of the players fixated on the owl they did not notice its gestures. Probably the timing of the gestures have affected this – owl's and character's gestures are shown at the same time and it is natural that player is more interested in his or her own virtual character than other game characters. Thus, we suggest that designers should prefer player's own character as a feedback channel rather than other characters. All the players were aware of mouse's gestures and knew what they mean. Because the meaning of mouse's and owl's gestures are parallel, there is no need to strengthen the gestures of the owl in this game.

Table 7. Fixation counts on the owl and the mouse

	N	Min	Max	M	SD
Mouse	8	1	17	11.5	3.25
Owl	8	1	5	3	1.0

Figure 5. Left: Areas of interest that was monitored (owl and mouse character); Right: Hot spot map from participants' fixations in the classroom scene

The analysis of labyrinth races revealed that players tend to pay attention either to their mouse character or to the tasks that are presented on labyrinth's black boards. Half of the tested players concentrated on the tasks and tried to solve the tasks during the race. After solving tasks in their mind these players tend to fixate on the doors that they thought to be the right answers as seen in Figure 6. For these players the race worked as a rehearsal tool and a reflection tool that helped them to figure out what their mouse knows and what it does not know. Thus, in case of these players the race facilitated the learning of the content. On the other hand, half of the tested players followed only the race between their mouse and the cat and totally ignored the math tasks. Such behavior is not effective from learning point of view and better ways to engage all players in processing learning content needs to be developed. Nevertheless, the retrospective interviews revealed that two of the eight players did not at all understand that the classroom activities affect the race (i.e. how their previous teaching activities affect to their mouse's performance in a labyrinth race). They were very disappointed when their mice lose the race, but they did not understand the possibility to correct the situation by teaching more mathematics to their mouse.

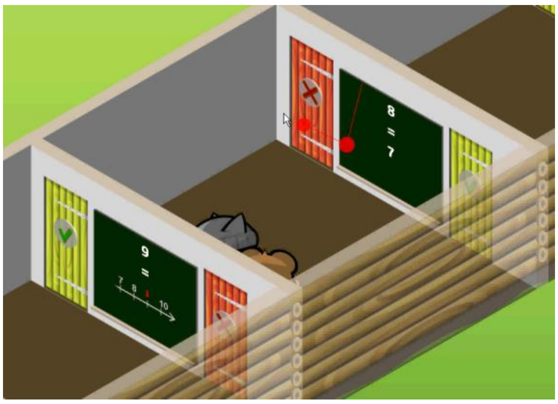

Figure 6. Fixation on the door that a player thinks to be the right answer in the labyrinth race (red dots are fixations and red lines saccades)

These findings have been important when designing the official release of Math Elements in spring 2013 and its successor SmartKid Maths (2014). The technological readiness was good when running these experiments, but the storytelling, icons and the guiding effects in UI was redesigned. So far SmartKid Maths game has been relatively successful in Appstore and Windows market: it has been in top 5 in education- and kids category monthly in more than 15 countries. Without extremely critical approach on user experience and learning curve in piloting stage, the game would not have been ready for consumer markets when it was launched. However, service type of game is never ready. In long term SmartKid Maths will be a gamified virtual school covering most of the STEM topics from global curriculums. This requires continuous research and development in pedagogy, gameplay and user experience.

5. Feon's Quest Study

The investigated game prototype was developed in the context of the European 80Days project (www.eightydays.eu). The game is a typical action adventure, designed for teaching geography for an age group of 13 to 14 years according to European. The game's story is simple: The player takes the role of a kid in whose backyard a spaceship lands one night. There is a friendly alien named Feon, an intergalactic scout with the task of collecting information about foreign planets and life forms. Together they fly with the spaceship from landmark to landmark (e.g., European capitals) and collect information by accomplishing various missions and solving various riddles. For the following eye tracking study we concentrated on two scenes from the game, a typical flying game situation and a graphical simulation task.

In the flying situation (Figure 7, top right image) the player maneuvers the spaceship over the surface of Europe. The main challenge is to find certain landmarks, for example, mountains, rivers, or cities. From an educational perspective, this task includes the competence to navigate correctly, to understand compass directions, degrees of longitude and latitude, the topography of Europe and, lately, the ability to control the spaceship. A Head-up display (HUD) is shown on the screen, with a compass in the middle, a communication window with Feon on the lower right side, a section of the map on the upper right side, and a computer text window on the lower side of the display (Figure 7, higher pictures).

The terraforming simulation scene focuses on experimenting and simulating the consequences of human interventions on the severity and the damages caused by floods. Interventions range from deforesting or the sealing of soil to positive interventions such as restoring original river courses. A 'terraforming area' on the upper left side is shown on the screen, a 'statistics area' on the upper left half, a 'flood cause and risk level area' on the upper right side, and a 'working desk' in the middle of the screen (Figure 7, lower pictures).

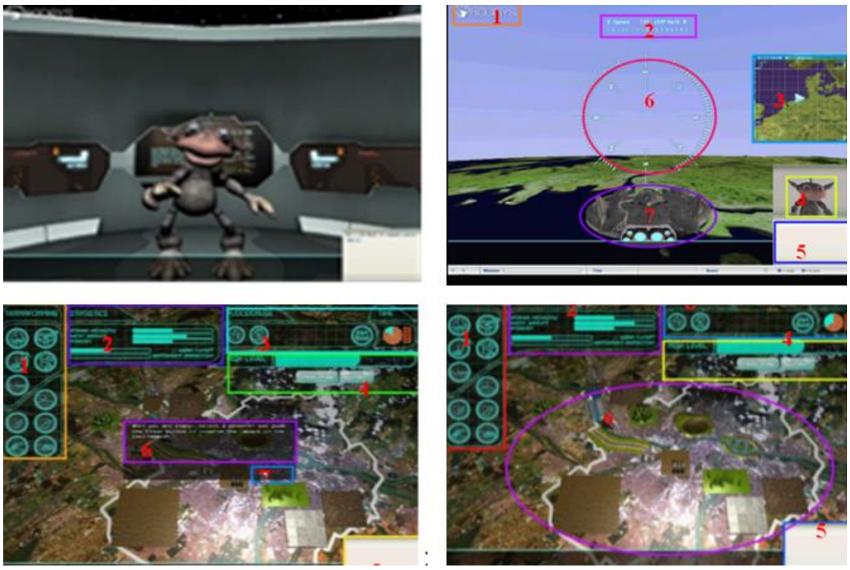

Figure 7. Screenshots from Feon's Quest.

5.1. Method

The study presented in this paper is only one of a long sequence of experiments in several European countries, with the major objective to evaluate and assess the usability as well as the educational efficacy of digital games. Due to the vast complexity of this research battery, in this paper we only present a rather concise snapshot of this work only. Nine Austrian children, 4 girls and 5 boys, took part in this study. The participants' age ranged between 11 and 16 years with the average of 13 years (SD=1.61).
The Tobii 1750 eye tracker was used to collect eye tracking data for this study. In the study, the eye movements were typically analyzed in terms of total duration as well as the duration of the situations, the relative fixation numbers, and the saccade lengths. This analysis allows getting information on how much time the participants spend in the three different situations while playing the game and on which parts their eyes are fixed (see Figure 8). We utilized a knowledge test in paper format for the pre and post assessments of knowledge. Furthermore motivational, usability-related, and attention-related tests were used. For the analysis, three scenes of the game were used: Flying to Budapest (flying situation), instructive cockpit scene in Budapest in order to introduce and explain the subsequent simulation (instruction situation), and the terraforming simulation (simulation situation).

Figure 8. The left image shows the eye tracking set up, the right a screen shot of the game's terraforming simulation. The colored recatangle indicate predefined areas of interest (AOI) for gaze data analyses.

5.2. Results

To evaluate the learning efficacy and to investigate if children actually learn by playing the game, the results of a knowledge test before and after the gaming session were compared. The average score of the pre-test was 32.33 (SD=9.45) and that of the post-test was 39.00 (SD=10.22). The difference is statistically significant (T=-3.841, df=8, p=0.005). For girls the average score was 26.25 (SD=11.09) and 31.25 (SD=10.63) respectively, which is no significant difference between these two test scores (T=-1.344, df=3, p=0.271). For boys the average of the pretest was 37.20 (SD=4.44) and that of the posttest was 45.20 (SD=4.02). There is a significant difference regarding these two test scores (T=-6.136, df=4, p=0.004).
With respect to the total duration of playing, comparisons of males and females imply that females spent more time on playing (1018.86 sec., SD=102.58) in contrast to male's average playing time of 868.80 seconds (SD=280.72), which is a remarkable results. Especially on the simulation situation females' total playing time is about 20% longer than those of the males (913.20, SD=142.28 vs. 726.41, SD=278.41). Only in the flying situation, the playing time for males was higher (M=94.30, SD=63.69) than that of girls (M=48.24, SD=24.06). With respect to the instruction situation there are nearly the same results for males (M=48.09, SD=21.87) and females (M=57.41, SD=26.94).
In the focus of this study, however, was the hypothesis that players who perform very well in terms of learning and who yield high satisfaction with the game exhibit distinct gaze patterns, significantly different from the opposite extreme group, the low performers. The results of this study provide some evidence for such assumption. In general, players who learned more spent about 17% more time on playing (cf. Table 8). This result is constant throughout the three scenes that were investigated in this study. Remarkably, despite the longer involvement in the game, the high performers yielded short fixation numbers (cf. Table 8). Vice versa, the high performers yielded an about 30% higher saccade length (73.44, SD=29.6 vs. 53.05, SD=27.13) than participants with a lower learning performance.

Table 8. Average playing time and number of fixations per second for high and low performers

	High performers	Low performers
Playing duration (in sec.)		
Total	940.10 (SD=23.52)	781.79 (SD=344.46)
Flying	100.09 (SD=79.33)	59.71 (SD=31.75)
Instruction	70.97 (SD=29.22)	45.53 (SD=20.28)
Simulation	769.04 (SD=77.21)	676.55 (SD=362.57)
Fixations per second		
Total	0.72 (SD=0.04)	0.77 (SD=0.12)
Flying	0.45 (SD=0.10)	0.40 (SD=0.04)
Instruction	0.57 (SD=0.37)	0.42 (SD=0.06)
Simulation	0.45 (M=0.06)	0.35 (SD=0.21)

We also looked into attention scores (IMMS subtest attention); participants with higher attention spent more time on playing the game (1038.53 sec., SD=114.78) than participants who yielded a low attention score (M=929.34 sec., SD=20.27). Regarding the overall fixation rate participants with lower attention yielded about 25% more fixations per second than participants with lower attention (cf. Table 9). In general, saccade length was also short for the high attention group. Specifically in the situation where the participants needed to fly the spaceship a significant difference was found, the highly attentive players exhibited an about 40% lower saccade length.

Table 9. Average number of fixations per second and average saccade length for high and low attention groups

	High attentions	Low attention
Fixations per second		
Total	0.52 (SD=0.14)	0.40 (SD=0.07)
Flying	0.48 (SD=0.07	0.39 (SD=0.04)
Instruction	0.61 (SD=0.33)	0.36 (SD=0.08)
Simulation	0.47 (SD=0.01)	0.44 (SD=0.08)
Saccade length		
Total	27.42 (SD=16.28)	46.51 (SD=6.52)
Flying	44.43 (SD=9.30).	90.07(SD=9.90)
Instruction	20.78 (SD=12.03)	22.41 (SD=8.84)
Simulation	17.06 (SD=27.05).	27.05 (SD=0.82)

Throughout the three scenes, we conducted an analysis of specific areas of interests (AOI) also (see Figures 7 and 8). An AOI is a pre-defined area on the screen that has a specific meaning or function. By the example of scene 1, flying the spaceship, we defined seven distinct areas; most importantly the head up display in the middle of the screen that gave course indications and the compass directions, the spaceship at the lower center part of the screen, and a 2D map area at the right side. Figure 9 presents the prototypical results for a player with high learning performance and high attention values (left image) in comparison to a player with low learning performance and attention (right image). These results indicate that the high performer distributed the gaze evenly over the screen without over-attending specific areas of little relevance. While the map area, that was attended equally by both players, is important for navigating and flying to Budapest, the spaceship as well as the head up display areas, which were only attending frequently by the low performer) are not crucial for the task. Such differences were found for all participants. In general, highly visible and moving areas attract the visual attention of players. Specifically inattentive players and low performers tend to stick to such areas even if they are unimportant for a task.

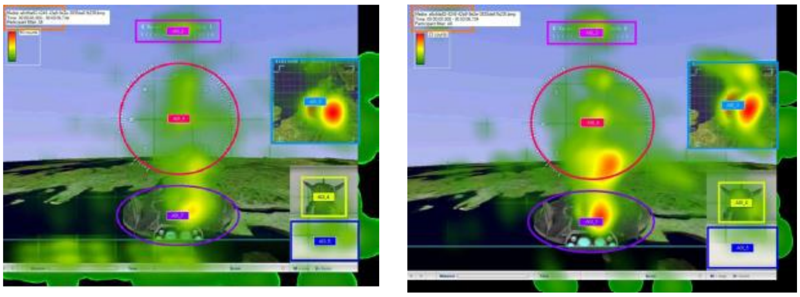

Figure 9. The left image shows prototypical gaze patterns for high performs, the right those of low performers.

6. Discussion and Conclusions

In this paper we used eye tracking method to study how players perceive game worlds, how they adopt the idea of the game and how they utilize the feedback that the game provides. The goal was to highlight the importance of critical analysis of results received from eye-tracking: The idea was not to prove our design's excellence but critically find and point out the factors that should be redesigned in order to radically improve the user experience. The results revealed that the different kinds of players perceive the game world differently and elements that grasp players' attention can vary a lot as was assumed. For example, the results showed that there were a distinct gender differences in the interaction style and perceptual paths as well as distinct differences between high and low performers. Furthermore, the results indicated that some players even missed relevant information during playing, which may disturb learning. Such information is very useful when trying to understand why some game solutions work in certain context and why some games fail to fulfill learning objectives. In academic research, we do have time for experimenting and the subjects are ready to deal with unfinished products. Furthermore, in research settings we support users in adopting the games, which is not a case in consumer products. In consumer markets users' expectations are totally different: Learning curve is expected to be as short as possible and majority of the players will reject the game if there is anything confusing.

This research focused especially on cognitive feedback that learning games use to get players to reflectively process essential learning content. It seems that what sooner the player notices the cognitive feedback and grasps it meaning that better he or she can play the game. Designers should ensure that players perceive all the crucial elements, so that players can develop effective playing strategies. The signaling method [12] should be used strongly enough to highlight all the necessary elements that should be processed. For example, in tested AnimalClass games the changes in brain size should be emphasized more and the brains could be located nearer the pet's head, because players tend to focus on their pets' gestures. The use of signaling effect is very important especially for low performers and inattentive players. Furthermore, the results showed that extraneous elements should be eliminated from the game world in order to avoid incidental processing in crucial moments. Especially animated and changing content grasp player's attention easily – In rich game environments incidental processing may overload player's mind and disturb learning activities. As the Feon's Quest study revealed, specifically low performers and inattentive players have difficulties in distinguishing important and irrelevant areas on the screen and tend to stick to salient elements no matter of their importance for a task.

Overall, it seems that eye tracking can provide important information from the visual design of games, the usefulness of provided feedback as well as the whole game based learning process. Although the basic eye tracking measures seems to provide new and important information about the learning process, it needs to be complemented with other methods. If the analysis relays only for example on fixation counts or on hot spot maps, there is a great risk to interpret the results wrongly. Based on only fixation counts and fixation lengths we cannot determine whether the user has understood the game elements that he has fixated on or not. For example without the complementary methods used in this research, we would have not realized that although the players of AnimalClass games paid attention to their game characters' gestures they did not always understand the meaning of them. Furthermore, the timing of fixations is also very important in dynamic games and timing should be considered when interpreting the eye tracking data. For example, the fixation count or length does not tell whether the player has seen the game character in the certain time when the character has provided important information to the player. In this study, we used retrospective interview and gaze replays as complementary methods that both turned out to be very useful and provided much deeper and useful information about player's behavior and understanding than the basic quantitative eye tracking measures. The downside of retrospective interview is that it is very time consuming.

Finally, we want to emphasize that traditional play testing and eye tracking provides complimentary information for developers: when traditional studies can provide information how to develop storytelling, dialogue and mechanics, eye tracking can provide very deep and objective information about interaction design and layout.

Acknowledgements

This work has been co-funded by the EU under the 7th Framework Programme , in the Games and Learning Alliance (GALA) Network of Excellence, Grant Agreement nr. 258169.

References

[1] Schell J., "The art of game design: a book of lenses". Morgan Kaufmann Publishers, Burlington, 2006.
[2] Dewey J., "Experience and Education". Simon and Schuster, New York, 1938/1997.
[3] Alkan, S., Cagiltay, K., "Studying computer game learning experience through eye tracking". British Journal of Educational Technology, 38(5), 538-542, 2007. http://dx.doi.org/10.1111/j.1467-8535.2007.00721.x
[4] Kiili, K., "Foundation for Problem-Based Gaming". British Journal of Educational Technology – Special issue on Game-Based Learning, 38(3), 394-404, 2007.
[5] Kiili, K, "Digital Game-based Learning: Towards an Experiential Gaming Model". The Internet and Higher Education, 8(1), 13-2, 2005. http://dx.doi.org/10.1016/j.iheduc.2004.12.001
[6] Boud, D., Keogh, R., Walker, R., "Reflection: Turning experience into learning". London: RoutledgeFalmer, 1985.
[7] Ketamo, H., Kiili, K., "Conceptual change takes time: Game based learning cannot be only supplementary amusement". Journal of Educational Multimedia and Hypermedia, 19(1), 39-57, 2010.
[8] Norman, D., "The psychology of everyday things". Basic Books, New York, 1998.
[9] Ketamo, H., Suominen, M., "Learning-by-Teaching in Educational Games". In proceedings of Ed-Media 2008. 30.6.–4.7.2008, Vienna, Austria, pp. 2954-2963, 2008.
[10] Hyönä, J., "The use of eye movements in the study of multimedia learning". Learning and Instruction, 20, 172-176, 2010. http://dx.doi.org/10.1016/j.learninstruc.2009.02.013
[11] Mayer, R. E., "Unique contributions of eye-tracking research to the study of learning with graphics". Learning and Instruction, 20, 167-171, 2010. http://dx.doi.org/10.1016/j.learninstruc.2009.02.012
[12] Boucheix, J.-M, Lowe, R. K., "An eye-tracking comparison of external pointing cues and internal continuous cues in learning with complex animations". Learning and Instruction, 20, 123135, 2010. http://dx.doi.org/10.1016/j.learninstruc.2009.02.015
[13] Kickmeier-Rust, M. D., Hillemann, E., Albert, D., "Tracking the UFO's paths: Using Eye-Tracking for the Evaluation of Serious Games.", In R. Shumaker (Ed.), Virtual and Mixed Reality - New Trends (pp. 315-324). Lecture Notes in Computer Science, 6773. Berlin: Springer, 2011. http://dx.doi.org/10.1007/978-3-642-22021-0_35
[14] Law, E., Mattheiss, E., Kickmeier-Rust, M. D., Alt, D., "Vicarious learning with a digital educational game: Eye-tracking and survey-based evaluation approaches". In G. Leitner, M. Hitz & A. Holzinger (Eds.), HCI in Work and Learning, Life and Leisure. Lecture Notes in Computer Science, 6389, (pp. 471-488). Berlin: Springer, 2010. http://dx.doi.org/10.1007/978-3-642-16607-5_33
[15] Chi M., Siler, S., Jeong, H., Yamauchi, T., Hausmann, R. "Learning from human tutoring." Cognitive Science, 25, 471-533, 2001. http://dx.doi.org/10.1207/s15516709cog2504_1
[16] Biswas, G., Schwartz, D., Leelawong, K., Vye, N. Tag, V., "Learning by teaching: A new agent paradigm for educational software". Applied Artificial Intelligence, 19, 363-392, 2001. http://dx.doi.org/10.1080/08839510590910200
[17] Mayer, R. E., Moreno, R., "Nine ways to reduce cognitive load in multimedia learning." Educational Psychologist, 38, 43–52, 2003. http://dx.doi.org/10.1207/S15326985EP3801_6